Publishable summary

For the period between 01-06-2009 and 31-12-2011, the following goalposts were planned to be achieved (amongst others) in this project:

1. *to genotype ~300,000 SNPs in the blood samples of the discovery database*

The decision was taken to use the Illumina QUAD 610 array: these chips allow determination of ~630,000 SNPs evenly spread over the human genome. For the Dutch population in the discovery database ~630,000 SNPs adequately covers the genome. 521,805 SNPs passed the extensive quality control (QC).

2. *the definition of relevant top SNPs in the discovery database*

The subjects, in the discovery database, were phenotyped based on the following rules for the presence /absence of obstruction:

a. control: a $FEV_1/FVC > 0.70$ and $FEV_1 > 90\%$ of the predicted value

b. case: a $FEV_1/FVC < 0.70$

For emphysema no consensus on the definition of cases or controls exists and it was decided to carry out an analysis using the continuous data.

The final dataset comprised 1030 airway obstruction cases and 1799 controls (to increase the statistical power of the analysis, blood bank controls were included). The genomic inflation factor turned out to be 1.01, indicating a lack of population stratification. A p-value of 10^{-4} was selected as threshold: 312 SNPs were selected. For the same samples as described above, the emphysema phenotype was available. Linear regression, adjusting for age and pack-years, was used. The genomic inflation factor turned out to be 1.05. A p-value of $5*10^{-4}$ was selected as threshold: 71 SNPs were selected.

3. *to define the ~30 most significant SNPs for obstruction and emphysema in the replication databases*

Samples were genotyped with an Illumina Golden Gate custom array. Association tests were performed separately for each of the cohorts and the outcome were subjected to a meta-analysis. The data from the discovery cohorts were also included in the meta-analysis. The meta-analysis replicated 10 SNPs: these showed below 5% p-values, although none showed genome-wide significance ($p < 5*10^{-8}$).

chromosome	*SNP*	*p-value from GWAS*	*p-value from replication*	*p-value from meta-analysis*	*odds-ratio*	*gene*
10	rs2601751	0.000248	0.00260	$6.623*10^{-6}$	1.15	FAM107B

chromosome	*SNP*	*p-value from GWAS*	*p-value from replication*	*p-value from meta-analysis*	*odds-ratio*	*gene*
10	rs7921286	0.000357	0.00353	$1.167*10^{-5}$	1.14	FAM107B
20	rs13041320	0.000402	0.00332	$1.311*10^{-5}$	0.86	C20orf186
15	rs2280033	0.000372	0.00462	$1.832*10^{-5}$	1.16	
4	rs6838261	0.000558	0.00384	$1.872*10^{-5}$	0.87	SORCS2
21	rs7279886	0.000351	0.01216	$5.585*10^{-5}$	1.14	PDE9A
12	rs7956804	0.000225	0.01590	$6.297*10^{-5}$	0.86	CD4
12	rs1641716	0.000475	0.01486	$7.641*10^{-5}$	1.14	BCL2L14
2	rs10189511	0.000276	0.10890	$9.046*10^{-5}$	0.71	
19	rs2115299	0.000130	0.02827	$9.165*10^{-5}$	1.12	CABP5

Table 1 Results of obstruction replication analysis, showing the chromosome, the SNP identifier, the p-values from resp the GWAS, the replication and meta-analysis and the resulting odds-ratio and gene identifier.

The same approach, as with obstruction, was used to replicate emphysema SNPs only for a continuous analysis was used. The meta-analysis replicated 7 SNPs: these showed below 5% p-values, although none showed genome-wide significance ($p< 5*10^{-8}$).

chromosome	*SNP*	*p-value from GWAS*	*p-value from replication*	*p-value from meta-analysis*	*odds-ratio*	*gene*
6	rs1224526	$2.16*10^{-5}$	0.03465	$5.30*10^{-6}$	-2.79	AC002485.1
8	rs333048	$1.33*10^{-5}$	0.01381	$2.57*10^{-6}$	3.52	RP11-618M23.1
9	rs2479028	$6.82*10^{-5}$	0.09609	$1.74*10^{-5}$	1.72	AL158151.1
13	rs1486949	$2.72*10^{-5}$	0.2939	$2.92*10^{-5}$	1.88	AL356241.1
9	rs1167763	$3.28*10^{-5}$	0.116	$3.64*10^{-5}$	3.03	GABBR2
6	rs1378301	$7.96*10^{-6}$	0.4003	$1.47*10^{-5}$	-1.83	RNU7-66P
4	rs4697618	$7.91*10^{-5}$	0.3225	0.0002946	-1.65	SEL1L3

Table 2 Results of emphysema replication analysis, showing the chromosome, the SNP identifier, the p-values from resp. the GWAS, the replication and meta-analysis and the resulting odds-ratio and gene identifier.

These results were discussed in depth and at length during the COPACETIC consortium meeting on Dec 16th 2010. The general opinion was that the outcome in terms of significant SNPs was lower than expected or hoped for. There is a general consensus within geneticists that a threshold of $1*10^{-8}$ has to be passed. That threshold is based on the need for a Bonferroni correction due to multiple testing. As the consortium agreed amongst each other that the SNPs reported made sense and a need for more statistical power was widely acknowledged, a decision was made for additional cohorts to be incorporated before final conclusions can be drawn. The data above are therefore to be considered as preliminary, as on-going. Additional cohorts are e.g Doetinchem, Rucphen and Glucold, as well as the Rotterdam cohort. These cohorts are added to the analysis in silico and no extra laboratory work is needed.

4. *to delineate the molecular mechanisms and pathways underlying COPD by investigating the changes in gene expression in peripheral blood.*

The outcome of the GWAS / replication analysis, as discussed above was deemed not to be decisive and it was decided that further work on the objectives of this work package can only proceed when more decisive results are available. This will not impede future activities as the work is mostly in silico.

5. *a set of predicting SNPs for COPD diagnosis* .

The outcome of the GWAS / replication analysis, as discussed above was deemed not to be decisive and it was decided that further work on the objectives of this work package can only proceed when more decisive results are available. This will not impede future activities as the work is mostly in silico.