

COPACETIC FINAL REPORT

Grant Agreement number: 201379

Project acronym: COPACETIC

Project title: COPD Pathology: Addressing Critical gaps, Early Treatment & diagnosis and Innovative Concepts

Funding Scheme:

Date of latest version of Annex I against which the assessment will be made:

Periodic report: final ☒

Period covered: from 01-01-2008 to 31-12-2010

Name, title and organisation of the scientific representative of the project's coordinator:

P. Zanen MD, PhD
UMCU
Division of Heart & Lungs
Heidelberglaan 100
3584 CX Utrecht
The Netherlands

Tel: +31 88 7556151

Fax: +31 88 7555421

E-mail: p.zanen@umcutrecht.nl

Project website address: <http://www.copacetic-study.eu/>

4.1 Final publishable summary report

4.1.1. executive summary

COPACETIC aims to unravel the genetic determinants of the susceptibility to develop Chronic Obstructive Pulmonary Disease (COPD) in (ex-)smokers at high risk. The Dutch Nelson cohort of ≥ 20 pack years (ex-)smokers delivered to the COPACETIC project pulmonary function tests / emphysema estimates, next to DNA / RNA samples. The genetic determinants found in this discovery cohort were replicated in nine cohorts: 1] *Poland*: the BOLD cohort; 2] *Denmark*: the Copenhagen City Heart Study (CCHS) and the Danish Lung Cancer Screening Trial (DLCST); 3] *the Netherlands*: the Vlagtwedde / Vlaardingen, the Doetichem, Rucphen and the Glucold studies; 4] *Germany*: the LUISI study and 5] *Sweden*: the EUROSCOP study. Subjects in the discovery / replication cohorts were dichotomized on obstruction as follows: a] case: $FEV_1/FVC < 0.70$; b] control: $FEV_1/FVC > 0.70$ and $FEV_1 > 90\%$ of predicted. Emphysema was estimated with the 15% HU-level method as a continuous trait and not dichotomized.

After genetic analysis via PLINK in the discovery / replication cohorts and subsequent meta-analysis, the outcome below is available.

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
10	rs2601751	0.000248	0.00260	6.623×10^{-6}	1.15	FAM107B
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

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
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 1a^b Results of obstruction / emphysema analysis, showing the chromosome number, the SNP identifier, the p-values from resp. the discovery and the replication cohorts, the meta-analysis p-value with the resulting odds-ratio and lastly the gene identifier.

4.1.2. *summary description of project*

COPACETIC aims to generate knowledge on the pathogenesis of Chronic Obstructive Pulmonary Disease (COPD) by using large datasets containing prospective clinical characteristics (pulmonary function tests and CT scans) and biological data (DNA and RNA). The data were collected in groups of individuals at high risk to develop COPD. Genome wide DNA and RNA analyses with replication of the results identify novel genetic markers for COPD susceptibility and molecular targets, with the intention to develop new diagnostic approaches and treatment strategies.

Starting at 01-01-2008 and ending at 31-12-2011, the following goalposts were planned to be achieved (amongst others) in this project:

1. to build a discovery database containing CT scans, pulmonary function tests and blood samples of ~4000 subjects
2. to collect clinical and biological data in the replication cohorts to be able to replicate genetic markers found in the discovery cohorts
3. to genotype ~300000 SNPs in the blood samples of the discovery database and to identify gene markers associated with COPD susceptibility
4. the definition of relevant top SNPs in the discovery database for both obstruction and emphysema
5. to define the ~30 most significant SNPs for obstruction and emphysema in the replication databases and to cross validate these in all available databases
6. to delineate the molecular mechanisms and pathways underlying COPD by investigating the changes in gene expression in peripheral blood.
7. to define a set of predicting SNPs for COPD diagnosis

ad 1] *building the discovery database*: the COPACETIC project is based on the Nelson study (a population based randomized multi-centre lung cancer screening trial, studying mostly male heavy current and ex-smokers). The UCMU / UMCG Pulmonary and Radiology departments collected data on pulmonary function and emphysema severity. The Nelson study started data collection in 2004 and the last entry was foreseen in 2008/2009. COPACETIC started 01/01/2008, and so data collection could be closed according to scheme, as it was up and running prior to the COPACETIC start.

ad 2] *building the replication databases:*

Poland: the Jagiellonian University School of Medicine in Krakow performed a COPD screening, following the Burden of Obstructive Lung Disease (BOLD) initiative. As result, a cohort of 487 subjects with pre- and postbronchodilator spirometry with blood samples for DNA was available

Denmark: the Copenhagen City Heart Study (CCHS) sampled approx 8750 subjects of which 1220 were diagnosed as COPD-patients. Blood samples of these subjects were already obtained prior to the start of COPATIC and DNA was extracted

the Netherlands: the Vlagtwedde/Vlaardingen study sampled approx 2500 subjects of which 633 were diagnosed as COPD-patients; the Doetinchem, Rucphen and the Glucold cohorts were added in silico at the end of the project

Sweden: the AstraZeneca EUROSCOP study was a 3-year long multicenter study on patients with mild COPD, all of which were active smokers. The COPD-patients were fully characterised via pulmonary function testing. DNA samples are available from 700 COPD patients from 9 different countries and 500 controls

Denmark: the Danish Lung Cancer Screening Trial (DLCST) followed a study design very similar to the Nelson study and included both spirometry and CT-scans. ~1300 subjects could be included. Data collection was closed ~1 year before COPACETIC closure.

Germany: the Heidelberg University built a cohort of ~2000 smoking subjects sampled from the general population, who underwent spirometry and CT-scans (LUISI study). The inclusion criteria of this study are similar to the NELSON and the DLCST

ad 3] *to genotype ~300,000 tag SNPs in the blood samples of the discovery database*

The decision was taken to use Illumina QUAD 610 arrays: 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. After extensive quality control (QC) 521,805 SNPs passed.

ad 4 *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:

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

ad] 5 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 ($p < 5 \times 10^{-8}$).

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
10	rs2601751	0.000248	0.00260	6.623×10^{-6}	1.15	FAM107B
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 2 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 with the difference that a continuous analysis was used.

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
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

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
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 3 represents the number of cases and controls from each cohort as well as number of SNPs that were included in final analysis. The meta-analysis replicated 7 SNPs: these showed below 5% p-values, although none showed genome-wide significance ($p < 5 \times 10^{-8}$).

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
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 3 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 the 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 in GWAS analysis. 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. As the current results are still preliminary, the decision was also taken to postpone work on WP5 and WP 6 till final data are available.

ad 6] *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.

ad 7] *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.

4.1.3. description of the main results

In the sections below we review the outcome of each work package in this project.

work package 1

The main objective of this work package is to ensure that the project's main objectives are realised on schedule and within the budgetary limits. Secondly, this work package is geared towards effectively disseminating the knowledge resulting from the project. The review of the progress in this work package is broken down by item in an alphabetical order.

consortium agreement. The process of drafting the agreement was started using the Desca simplified FP7 consortium agreement as a template. Aided by the legal departments of the partner institutions and the FP7 IPR helpdesk, a final text was agreed upon. The presence of an industrial partner and the possible use of foreground lead to some discussion as points of view varied. In the end it was decided that a joint ownership, as suggested by the Desca template, would work best. The final text was agreed upon on the consortium meeting of Nov 25, 2008 and subsequently all parties signed the agreement. The signed version was distributed in the meeting of June 25, 2009 to all partners. Later during the project the DLCST study was added to the agreement with full consent of all original partners (see item 0 on page 8).

consortium meetings. Every six months a full-day meeting was held to discuss progress and to take decisions on relevant issues, related to either the organization or to scientific issues. The meetings included reporting to the General Assembly. The coordinator organised these meetings with all consortium members: it was planned to have bi-annual meetings, but the members deemed that for one particular meeting insufficient data were available for an in-depth discussion and postponed that meeting. As result six meetings were held on Jan 8th / June 25th / Nov 25th 2008, June 25th 2009, April 8th / Dec 16th 2010. Each meeting was chaired by the coordinator, who made minutes and which were distributed to the members for approval. Every month the two Dutch partners met in Groningen to discuss issues related to the discovery database and the replication. The coordination of flow of data, building the databases, DNA-sample logistics etc etc were discussed in 1-2 hour meetings. The COPACETIC coordinator was present; minutes were made.

consortium structure. No partners left the consortium, one was added. A few relevant items are worth mentioning: a] the change of DKFZ to the Heidelberg University, b] the inclusion of

AstraZeneca subsidiaries and c] the addition of the Danish Lung Cancer Screening Trial as replication cohort

COPACETIC started with the German DKFZ as partner: prof Kauzcor headed that institution, but he switched to the Heidelberg University. It meant that DKFZ was taken out as a partner and replaced immediately by the Heidelberg University. The Technical Annex was adopted and updated to reflect this change and new forms A were signed.

AstraZeneca consists of many local subsidiaries which are all legal entities and according to EU-rules, dissemination amongst AZ subsidiaries of information leads to drafting and signing of agreements between AZ subsidiaries. This is an unwanted situation and it was decided to ass two AstraZeneca subsidiaries to the consortium agreement as affiliates: AstraZeneca UK Limited and AstraZeneca AB.

Halfway the project it was discovered through the AstraZeneca members that in Denmark the Danish Lung Cancer Screening Trial (DLCST) was carried out. That study, organised by dr. Asger Dirksen and dr. Jesper Holst Pedersen, strongly resembles the COPACETIC underlying Nelson study. They included the same type of heavy smoker of similar age, measured the same parameters as in the Nelson study (low dose CT-scan / spirometry / blood samples for DNA). After some discussion within the consortium, informal talks between the coordinator and the DLCST were held. The mutual interest was high and it was decided that the DLCST cohort could be added to the COPACETIC project. The most important goal of this cooperation was to increase the size of the replication cohorts with emphasis on the replication of the SNPs associated with emphysema, as such cohorts are scarce. As spirometry data were available, DLCST also acted as replication cohort for obstruction. The DLCST finalised its data collection at the moment the cooperation started and additional funding was not necessary. In the near future that cooperation will be continued.

cooperation with other groups. The dissemination of data, as explained in section **Error! Reference source not found. Error! Reference source not found.**, resulted in several invitations to cooperate. The most frequent question to COPACETIC is to serve as a replication cohort for other COPD genetics studies. A few examples are

- the COPDgene study in the USA
- the DeCramer study into emphysema/diffusion capacity

- the Eclipse study in Canada/UK
- the group of Silverman, Harvard University in the USA

cooperation with the NELSON study. As known the Nelson study forms the basis under COPACETIC as the former study has in place all logistic activities and performs the CT-scanning, needed to obtain the emphysema scores. The basis for the cooperation is a free exchange of information and data, which boiled down to a COPACETIC access to DNA-samples (where needed), CT-scans and demographic data (e.g. smoking history). In return NELSON gained access to the GWAS outcome and can use these data to answer research questions they consider relevant. The NELSON steering committee also has right to nominate authors in the COPACETIC publications (and visa versa). Each party acknowledged that the one is and stays the owner of the data exchanged and no third party use is allowed without permission.

Despite the above starting point, some members of the NELSON steering committee started to consider the understanding as unbalanced: they valued their data higher as in the beginning and halted the exchange of demographic data. Several discussions within the NELSON steering committee and with the COPACETIC spokesman lead to an adaptation of the above free exchange principle: for an amount of € 75000 the demographic data were transferred. This completed the negotiations between NELSON and COPACETIC and from that point on no further discussions were necessary.

dissemination. In the first period of the project a restricted dissemination policy was adopted as COPACETIC was in a data collection phase: news letters were not issued. Two immediate major outward bound activities were undertaken:

- a. press releases to inform the lay public on having succeeded in obtaining the grant: this was widely cited in the national Dutch press and appeared on many websites
- b. a poster illustrating the design and goals of COPACETIC was presented at the ERS conference in Berlin. The first major activity was a presentation at the September 2009 Vienna ERS conference: the ERS organized a special session to present and discuss FP6 and FP7-projects.
- c. Following that first major presentation many posters / oral presentations at symposia / congresses were displayed or given
- d. as planned, regular newsletters were issued on schedule in the later phases of the project

e. the COPACETIC research project was presented to the Dutch lay public on a Dutch Astma Fund event in September 2009 where ~800 Astma Foundation members and their family were present. Prof Postma from the UMCG addressed the lay public in 2010 with a presentation on the causes of COPD and the genetic background of the susceptibility.



Figure 1 Astma Foundation presentation to the Dutch lay public on COPACETIC

website. The COPACETIC website is developed by the subcontract partner of the coordinating UMCU, the Dutch Asthma Foundation. It was decided to design two websites: one for professionals and one for the lay public. The former one was developed first. The site informs the professional on the starting points of COPACETIC, newsletters can be downloaded and researchers can communicate with the COPACETIC researchers. The website is operational since 13-03-2009 and is updated regularly. The website was expanded with a section for PhD students. The PhD students disseminate their activities, results etc via weblogs to the public in an attempt to heighten their profile.

The lay public site became operational after the professional site turned functional. The design of this website was considered much more difficult compared to the professional one. The site has to cope with a fundamental characteristic of COPD: it is a blue collar disease and patients are

known to reside in the lower classes of society. They are often not able to read/understand English and language specific sites are obligatory. Advice how to structure the lay public site was sought from prof. Kaptein from the chronic illness psychology department of the Leiden University, the Netherlands.

The website informs on COPD, DNA and genes, the COPACETIC research project and the researchers involved. The information is available in English and the languages from all participating partners (English, Dutch, Danish, German, Polish and Swedish).

Some statistics on the website: there were ~1340 visitors per year which averages ~100 / month and the time on site was ~1.5 minutes. 68% of the visitors found the website via a search engine, 19% knew the URL, and 13% found the site through referral sites. The visitors came from 50 different countries and the top 10 was: Netherlands, Poland, United States, Sweden, Denmark, Germany, United Kingdom, Belgium, Finland and last India.

work package 2

The objectives of work package 2 for the second reporting period are listed as:

- 1. to collect CT scans, pulmonary function tests scan and blood samples of the ~4000 subjects in the NELSON cohort in the discovery database that will serve as the basis for the genetic studies in WP 3*
- 2. to collect clinical and biological data in the replication cohorts to be used to replicate the findings obtained via the genome wide scan in subjects in the discovery database*
- 3. blood samples replication cohorts collected and ready to be transferred to WP4;*
- 4. pulmonary function test data and demographic data replication cohorts ready to be used in WP4*

ad 1] building the discovery database. The COPACETIC study is based on the NELSON study. The latter means that the UCMU / UMCG Radiology and Pulmonary departments, which take part in the Nelson study, were up and running in collecting data at the start of the COPACETIC project. The Nelson study started data collection in 2004 and the last entry was in 2008/2009. COPACETIC started 01/01/2008, and data collection could therefore close according to scheme..

After closure, the pulmonary function databases of the UMCU / UMCG were merged into one single pulmonary function database, which contains 6136 pulmonary function records and obtained in 3784 subjects. 1826 subjects were included via the UMCU and the remaining 1958 via the UMCG.

Also after data collection closure, from the UCMU / UMCG radiology databases CT-scans were extracted. Due to the size (approx 0.5 Gb per scan) downloading and storing the ~6200 scans needed for COPACETIC was a lengthy and cumbersome process: only batches of scans was transferred to the UMCU Image Science Institute for processing. This institute developed the ImageExplorer software to estimate the extent and localisation of emphysema. These results were made available in Excel-format, which files were translated into SPSS datasets and these were merged with the pulmonary function database.

Regarding the emphysema estimation a problem emerged. In short, comparison of the emphysema severity in UMCG / UMCU based subjects learned that, on the average, UMCG subjects suffer from more extensive emphysema. The UMCU Image Science Institute technicians learned that the reconstruction algorithms are the most probable culprits and that their ImageExplorer software needed adaptations. In the mean time CT-scans were validated visually. The solution reached was based on the noted differences between the two CT-scan systems: it was found that that a systemic difference in density

measurements did exist. A Philips machine will always deliver other HU-levels per voxel as a Siemens machine. A correction factor was implemented: the air in the human trachea can easily be measured and the density is known by definition ($=-1000$ HU) and the difference between -1000 HU and the actual measured value denotes that correction factor. All voxels in each CT-scan were corrected on an individual basis, which necessitated recalculation of all available UMCU / UMCG CT-scans. The approach to emphysema severity was also adopted: the approach, where the percentage lung volume with a sub-threshold density (e.g. <-950 HU) is estimated, was replaced by one where the HU-level below which 15% of the voxel density distribution falls, was estimated. This approach is less sensitive to scanner differences and shows a normal distribution. The percentage volume approach has a highly right skewed distribution, which is hard to handle statistically

In the UMCG blood samples were already obtained at year 1 of the NELSON study for all subjects. These samples were stored under suitable conditions for later DNA-extraction. In the UMCU blood samples were obtained in the last year of the NELSON study ($=$ year 4). This difference in logistics means that it is possible that subjects present in the database with year 1 PFT and CT-scan, but who dropped out and did not return for the last year 4 measurements, do not contribute to the blood sample databank. As the number of DNA-samples planned to be subjected to the GWAS-procedure is less than the number of available DNA-samples and drop-out was random, this minor lack of blood samples did not prove to be a problem.

The pulmonary function /emphysema score database could be expanded by smoking related parameters from the NELSON study database. Parameters, like pack years, current/ex-smoker, smoking starting age, etc etc could be added. This allows e.g. a correction of the genetic background of COPD susceptibility for differences in smoking characteristics.

ad 2] building the replication databases. From 6 replication cohorts demographic data and DNA-samples needed to be sent to the Genetics Department of the UMCG for processing. For each of the replication cohorts we summarise the status.

- a. *Poland:* the Jagiellonian University School of Medicine in Krakow (prof. Nizankowska-Mogilnicka) performed a COPD screening, following the Burden of Obstructive Lung Disease (BOLD) Initiative. As result in the Krakow area a cohort of 603 subjects became available and 487 provided pre- and post- bronchodilator spirometry with blood

samples; DNA was extracted. These data /specimens were already available prior to the start of COPACETIC. The samples proved to be of very good quality. The transfer of the pulmonary function data was without problems.

- b. *Denmark*: the Copenhagen City Heart Study (CCHS) (prof. Jorgen Vestbo) sampled approx 8750 subjects of which 1220 were diagnosed as COPD-patients. Blood samples of these subjects were already obtained prior to the start of COPACETIC and DNA was extracted. After the start of COPACETIC it became clear that medical-ethical restrictions applied: the CCHS patient consent did not allow shipment of DNA to third parties. The solution found is straightforward: prof Vestbo closely cooperates with prof. Nordestgaard of the Herlev University Hospital, who is heavily involved with the CCHS. The replication of the significant SNPs will be done by prof. Nordestgaard in house on equipment equal to that used in UMCG Genetics Department on the same platform. This guarantees equivalent quality, suitable outcome and compliance with the medical-ethical restrictions. The outcome of this replication will be shared with the consortium freely and completely.
- c. *the Netherlands*: the Vlagtwedde/Vlaardingen study sampled approx 2500 subjects of which 633 were diagnosed as COPD-patients. Blood samples of these subjects were obtained prior to the start of the COPACETIC and DNA is extracted. Pulmonary function data are available and located at the UMCG. Later other cohorts could be added in silico: the Doetinchem, Rucphen and the Glucold cohorts
- d. *Germany*: the Heidelberg University intended to build a cohort of ~2000 smoking subjects sampled from the general population, who will undergo CT-scanning and pulmonary function testing (LUSI study). The inclusion criteria of this study are similar to the NELSON study. Inclusion unfortunately proved to be slower as expected: the approach by interesting subjects via news paper advertisements was not as efficient as hoped for. The strict >40 pack years criterion also was a significant negative factor and during 2008 the LUSI steering committee decided to adopt a less strict inclusion criterion, including smokers with >20 pack years. The inclusion time was lengthened and samples arrived in the UMCG Genetics Departments later as anticipated, but still in time for full analysis. The project, as a whole, was not jeopardized by this slower LUSI inclusion.
- e. *Sweden*: the AstraZeneca EUROSCOP study was a 3-year long multicenter study on patients with mild COPD, all of which were active smokers. The COPD-patients were fully characterised via pulmonary function testing. DNA samples are available from 700

COPD patients from 9 different countries and 500 controls. The samples and pulmonary function data are in place. As with partner b] Astra-Zeneca is restricted in disseminating samples and data and prior to the start of this COPACETIC project one agreed on a similar approach as for b]: AstraZeneca will perform replication of the most significant SNPs in-house and the outcome of this replication will be shared with the consortium freely and completely.

- f. *Denmark*: the Danish Lung Cancer Screening Study is a five year longitudinal study with the same goal as the Nelson / LUISI study, the early detection of lung cancer via low-dose CT-scans. The measurements in this study encompass yearly spirometry and CT-scans, next to blood samples for DNA extraction. The study was up and running at the moment the cooperation started and the year 1 baseline data / samples were efficiently transported to the UMCG Genetics Departments. In total ~1300 samples were received in Groningen.

work package 3

The objectives of work package 3 for the second reporting period are listed as:

1. frequency $\sim 300,000$ SNPs of isolated DNA samples in discovery data-base determined;
2. *databases containing the genome wide scan and the CT/pulmonary function test (=COPD diagnosis) data in the discovery database merged;*
3. *selection of the ~ 400 relevant SNPs in the discovery database*

ad 1 SNP frequency estimation and QC. The genome-wide scan was performed using Illumina 610 Quad BeadChips, containing 620901 probes. In total 3082 DNA samples from the Nelson study cohort were hybridized and the dataset was subjected to a stringent quality control to exclude samples and SNPs performing less well in the assay. The Hardy-Weinberg equilibrium was checked and those SNPs not in HWE ($=p < 0.0001$) were removed from the analysis. Missingness per individual and per genotype was also investigated and every SNP sample with a missingness $> 5\%$ was excluded. SNPs with minor allele frequency $< 5\%$ were removed as well. Population stratification was investigated and pairwise identity by state (IBS) distances were calculated. Based on that, ethnic outliers, related individuals and duplicates were identified and removed.

ad 2 the genome-wide association on airway obstruction in the discovery cohort. The final dataset so comprised 1030 airway obstruction cases and 1799 controls (to increase the statistical power of the analysis, blood bank controls were included). Cases were defined as a $FEV_1/FVC < 0.7$; controls as $FEV_1/FVC > 0.70$ and $FEV_1 (\%pred) > 90\%$ (the latter definition selected subjects with near normal or even a high normal FEV_1 to increase the contrast). Association tests were performed via PLINK. The genomic inflation factor turned out to be 1.01, indicating a lack of population stratification. The resulting Q-Q plot indicated a clear deviation of the expected number of significant SNP-frequency differences between cases and controls (see Figure 2). The distribution of association signals per chromosome is represented in a Manhattan plot (see Figure 3). A p-value of 10^{-4} was selected as threshold: all SNPs with a p-value below that threshold were selected for the replication analysis. 312 SNPs were selected.

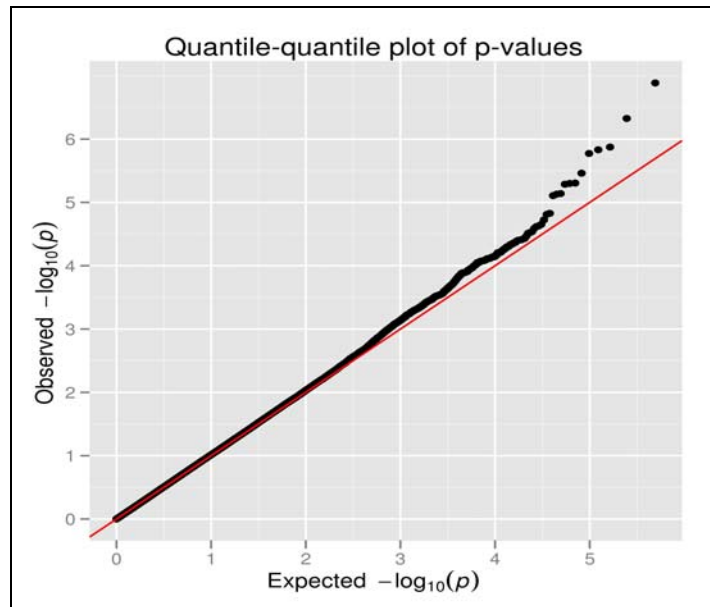


Figure 2 Quantile-quantile plot – p value distribution from the airway obstruction association tests.

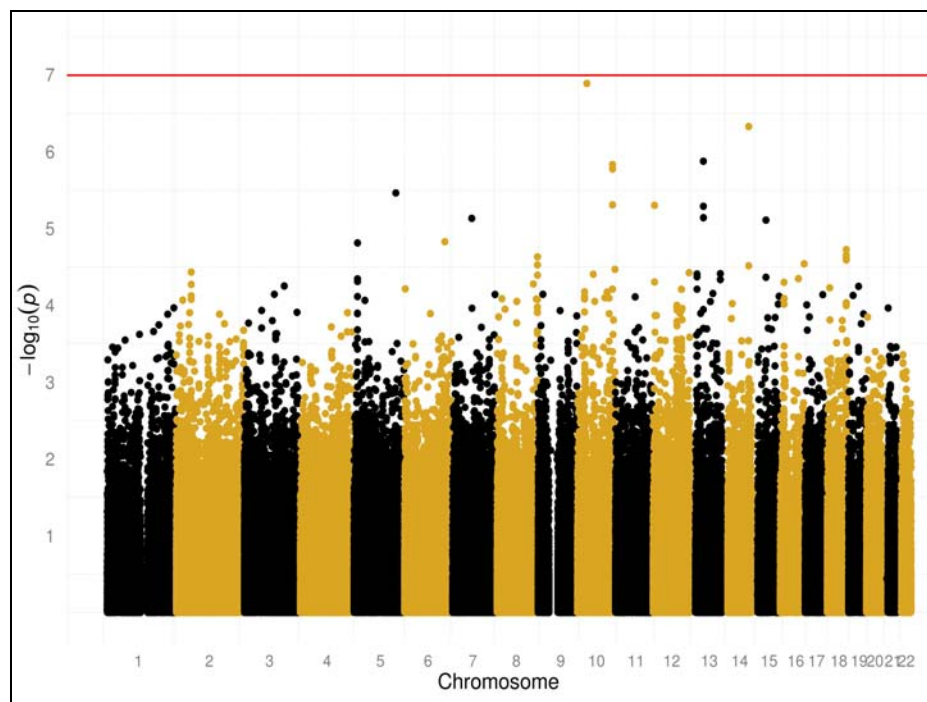


Figure 3 Manhattan plot – association signals distribution from the airway obstruction analysis

4. *the genome-wide association on emphysema in the discovery cohort.* For the same samples as described above, the emphysema phenotype was available. For none of the available approaches to emphysema estimation, consensus how to dichotomize into cases / controls exists and a continuous approach was adopted. Linear regression, adjusting for age and pack-years was performed in PLINK. The genomic inflation factor turned out to be 1.05, indicating a lack of

population stratification. The resulting Q-Q plot indicated a deviation of the expected number of significant SNP-frequency differences between cases and controls (see Figure 1). The distribution of association signals per chromosome is represented in a Manhattan plot (see Figure 5). A p-value of 5×10^{-4} was selected as threshold: all SNPs with a p-value below that threshold were selected for the replication analysis. 71 SNPs were selected.

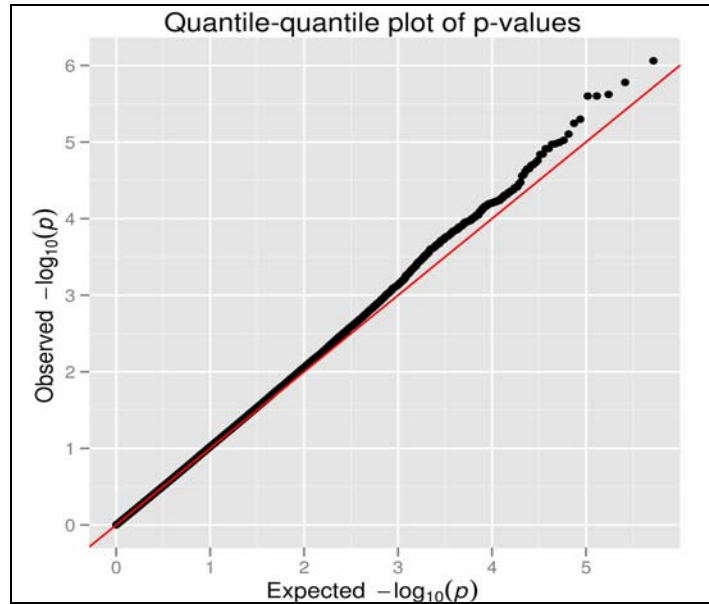


Figure 4 Quantile-quantile plot – p value distribution from the emphysema association tests.

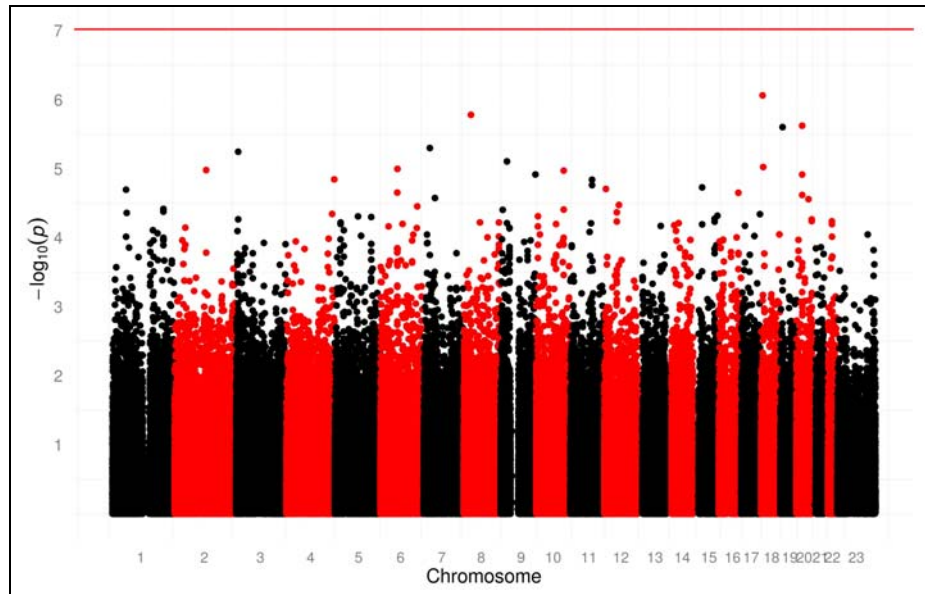


Figure 5 Manhattan plot – association signals distribution from the emphysema analysis

work package 4

The objectives of work package 4 are listed as:

1. *isolated DNA in blood samples of replication cohorts;*
2. *frequency 400 SNPs of isolated DNA samples in replication database determined.*
3. *databases containing the genome wide scan and the CT/pulmonary function test data (=COPD diagnosis) in the replication database merged;*
4. *~ 30 most significant SNPs determined*

The first three objectives were all achieved and below we report on the replication analysis outcome.

Ad 1 the replication of airway obstruction associated SNP. As known, 312 associated SNPs were selected for replication. Samples were genotyped with an Illumina Golden Gate custom array. To assure proper genotype calling, samples from each cohort were clustered separately. Next, each cohort underwent quality control to assure valid and accurate data. 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. Table 4 represents the number of cases and controls from each cohort as well as number of SNPs that were included in final analysis.

The meta-analysis replicated 10 SNPs: these showed below 5% p-values, although none showed genome-wide significance ($p < 5 \times 10^{-8}$).

<i>cohort</i>	<i>cases</i>	<i>controls</i>	<i>SNPs</i>
BOLD study Poland	88	310	293
CCHS Denmark	1104	3362	287
DLCST Denmark	641	561	293
Doetinchem, the Netherlands	155	743	292
Glucold	94	221	291
Rucphen, the Netherlands	89	69	284
Vlaardingen/Vlagentwedde cohort, the Netherlands	234	1058	289
Total for meta	2405	6324	261*

Table 4 Number of obstructive cases; controls and number of SNPs available for analysis after QC. * = the number of SNPs present in all cohorts.

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
10	rs2601751	0.000248	0.00260	6.623×10^{-6}	1.15	FAM107B
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

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
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 5 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.

ad 2 the replication of emphysema associated SNPs. As known, 71 associated SNPs were selected for replication. The same approach as for obstruction was used here with the difference that again a continuous analysis was used. Table 6 lists the number of cases and controls from each cohort as well as number of SNPs that were included in final analysis. The meta-analysis replicated 7 SNPs: these showed below 5% p-values, although none showed genome-wide significance ($p < 5 \cdot 10^{-8}$).

<i>cohort</i>	<i>samples</i>	<i>SNPs</i>
DLCST, Denmark	1370	65
LUI SI, Germany	904	68

Table 6 Number of samples and number of SNPs available for analysis after QC.

<i>chromosome</i>	<i>SNP</i>	<i>p-value from GWAS</i>	<i>p-value from replication</i>	<i>p-value from meta-analysis</i>	<i>odds-ratio</i>	<i>gene</i>
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 7 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 the geneticists that a threshold of $1 \cdot 10^{-8}$ has to be passed. That threshold is based on the need for a Bonferroni correction due to multiple testing in GWAS analysis. As the consortium agreed 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.

As the current results are still preliminary, the decision was also taken to postpone work on WP5 and WP 6 till final data are available.

work package 5

The objectives of work package 5 are listed as:

- 1. candidate molecular phenotypes for COPD determined*
- 2. replicated expressed genes in replication tissue bank*
- 3. mapped pathways to top ~ 30 SNPs of WP4*

RNA was isolated from frozen PAXgene tubes for 142 controls, 198 emphysema and 241 obstructive cases. RNA has also been isolated from 98 subjects for which the phenotype needs to be checked. All the RNA samples were hybridized to HT-12 arrays. Gene expression profiles for 48.000 transcripts have been generated for these cases.

The outcome of the GWAS / replication analysis, as discussed under work package 4, was deemed not to be decisive during the Dec 16th consortium meeting and it was decided that further work on the objectives of this work package can only proceed when more decisive results from work package 4 are available. This will not impede future activities as the work is mostly in silico.

work package 6

The objectives of work package 6 are listed as:

1. *the objective of this work package is to assess the diagnostic value of the SNPs in separating COPD- from non-COPD-subjects and to build a prediction rule.*

The outcome of the GWAS / replication analysis, as discussed under work package 4, was deemed not to be decisive during the Dec 16th consortium meeting and it was decided that further work on the objectives of this work package can only proceed when more decisive results from work package 4 are available. This will not impede future activities as the work is mostly in silico.

4.1.4. additional research projects not described in the Technical Annex

The Nelson study obtained data on the absence / presence of so called chronic mucus hypersecretion and the ‘re-use’ of the GWAS data in the discovery cohort lead to a side project into the genetic susceptibility to chronic mucus hypersecretion.

The Nelson study obtained CT-scans and lung function tests in year 1 and year 4 of that project: longitudinal changes in both emphysema severity and lung function therefore are available, enabling a side project into the study of determinants of lung function decline.

chronic mucus hypersecretion

Chronic bronchitis, a sub-phenotype of COPD, is characterized by chronic cough and mucus hypersecretion (CMH) and is often accompanied by breathlessness. CMH is defined here as the presence of sputum production during at least 3 months in two consecutive years without another explaining origin. Patients with COPD suffering from CMH have a significantly accelerated FEV₁ decline and a higher risk of hospitalization than those without these symptoms. Moreover, individuals with CMH have a 4-fold risk of mortality compared to those without CMH. So far it is not understood why CMH is a risk factor for COPD development, and neither why this constitutes such a risk of accelerated lung function decline and mortality in COPD patients. A plausible explanation for this phenomenon is the presence of a genetic predisposition. As the data from the GWAS in the discovery cohort were available as well as data on CMH from questionnaires, a genome wide association study is easily done.

	<i>Controls</i> <i>N=1795</i>	<i>CMH</i> <i>N=717</i>
age	60.2	60.4
smoking		
packyears (median)	34.2	38.7*
current %	47.5	74.2*
lung function		
FEV ₁ % predicted	100.3	93.5*
FEV ₁ /FVC %	72.9	69.2*
COPD cases (FEV ₁ /FVC <70%) %	31.0	46.7*

Table 8 CMH cases and controls in thhe discovery cohort (*= significant differences between CMH cases and controls).

77 SNPs associated with CMH with a p-value <10⁻⁴, of which 5 SNPs had a p-value <10⁻⁵. Many SNPs were localized close to “promising” genes in relation to the known biological pathways

involved in mucus hypersecretion, or grouped close to the same gene. The Manhattan plot is shown below.

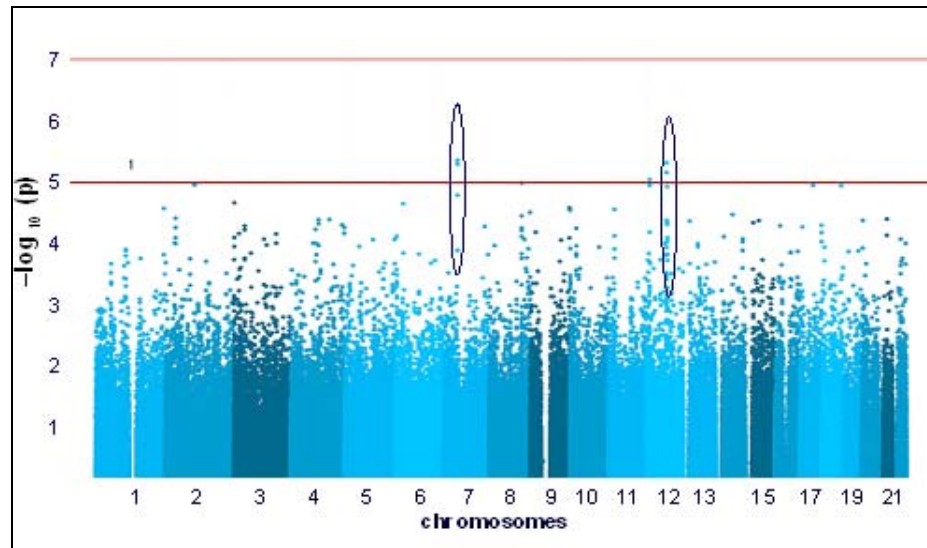


Figure 6 Manhattan plot for CMH in the discovery cohort.

The 71 top SNPs could be replicated in 6 cohorts: Heidelberg and Poland, Rucphen, GLUCOLD, Vlagtwedde/Vlaardingen and Doetinchem: 4 SNPs replicated with a p-value $<1^{0.5}$ and 2 SNPs with a p-value $<10^{-4}$. A next round of replication of these SNPs in the Rotterdam cohort (4000 subjects), in the CCHS (1600 subjects) and in the LifeLines cohort (8000 subjects) is foreseen.

<i>CHR</i>	<i>SNP</i>	<i>P</i>	<i>OR</i>	<i>closest gene</i>
18	rs8086262	9.35×10^{-7}	1.308	CBLN2
1	rs6677529	2.28×10^{-6}	1.365	NOS1AP
3	rs6782856	5.56×10^{-6}	1.292	ITGA9
3	rs6796546	6.29×10^{-6}	1.267	CTDLSPL

Table 9 Replicated SNPs for the presence of CMH.

The chromosome number, SNP identifier, p-value from the replication, odds ratio and gene identifier are given.

determinants of lung function decline

The first studies investigated whether the extent / distribution of emphysema in the discovery cohort associates with stronger lung function decline. 2085 males (mean age 59.8 years) were included in the first part. The mean (SD) baseline emphysema severity (15% HU level method, denoted as Perc15) was -934.9HU (19.5). A lower Perc15 correlated with a lower FEV₁ ($r=0.12$) at baseline ($p<0.001$). Linear mixed model analysis showed that a lower Perc15 significantly related to a stronger declined in FEV₁ ($p<0.001$) after follow-up. Participants without baseline

airway obstruction, but developing it after follow-up, had significantly lower mean (SD) Perc15 values at baseline than those who remained non-obstructive: -934.2HU (17.1) versus -930.2HU (19.7) ($p < 0.001$). Baseline emphysema severity is related to lower baseline lung functions and stronger rates of lung function decline, even in those without airway obstruction.

The next part investigated whether the distribution of emphysema is associated with lung function decline. 587 participants underwent CT-scanning of the lungs and pulmonary function testing at baseline next to a follow up measurement after a median 2.9 years. The lungs were automatically segmented based on anatomically defined lung lobes. Severity of emphysema was automatically quantified per anatomical lung lobe. Linear mixed models, correcting for age, height, BMI, packyears and smoking status, were used to assess the association of emphysema distribution and FEV₁/FVC-decline. Participants with upper lobe predominant emphysema had a lower FEV₁/FVC after follow-up compared to participants with lower predominant emphysema ($p = 0.001$), independent of the total extent of emphysema. Heavy smokers with upper lobe predominant emphysema have a more rapid decrease in FEV₁/FVC than those with lower lobe predominant emphysema. Upper lobe predominant emphysema may be a different phenotype than lower lobe predominant emphysema.

The second line focused on the relation between baseline lung function and subsequent decline. Subjects were classified by their entry FEV₁/FVC: group 1 $> 70\%$; group 2 $< 70\%$, but $>$ lower limit of normal (LLN) and group 3 $< \text{LLN}$. Differences in FEV₁/FVC, FEV₁, MEF₅₀ and Perc15 decline / increase between these groups were assessed using multiple linear regression and one-way ANOVA. Over three years, mean (SD) FEV₁/FVC, FEV₁, and MEF₅₀ decline in group 1 was 3.1% (1), 0.21 L (0.07) and 0.39 L/s (0.27), respectively. Decline in group 3 was 0.15 L (0.08), 2.4 (1.1) and 0.12 (0.2) in FEV₁/FVC, FEV₁ and MEF₅₀, respectively. Mean (SD) emphysema progression in group 1 was 3.7 (0.4) HU and was 9.1 (0.7) in group 3. Decline in all lung function parameters was highest in group 1 when compared to group 3, but emphysema progression was highest in group 3 (p all < 0.001). So called ‘non-diseased’ subjects according to the GOLD and LLN approaches (group 1) show the steepest decline in lung function, however progression of emphysema was greatest in those with $< \text{LLN}$. The LLN approach selects the smallest number of subjects with a rapid lung function decline, but with the largest emphysema progression.

Next the relation of the baseline diffusion capacity for CO to the decline in lung function was investigated. The association between Kco at baseline with progression of emphysema and lung function decline was assessed by multiple linear regression. 522 participants were included with a mean (SD) age of 60.1 (5.4) years. A lower baseline Kco was significantly related to an increase of

CT-quantified emphysema and a more rapid decline in FEV₁/FVC. A one standard deviation (0.25) lower Kco value at baseline, predicted a 1.6 HU lower Perc15 and a 0.78% lower FEV₁/FVC after follow-up ($p < 0.001$). A lower baseline Kco value is independently associated with a more rapid progression of emphysema and lung function decline in heavy smokers.

The third line investigated the effect of the duration of smoking cessation on lung function decline and the increase of CT-quantified emphysema. Smoking status at enrolment (≥ 5 , 1-5, ≤ 1 year of smoking cessation or current smoking) was assessed. Change in lung function and emphysema severity was analyzed by multiple linear regression adjusting for age, height, baseline pulmonary function/emphysema severity, packyears, years in study and study center. Current smokers were used as reference. The groups ' ≥ 5 years' and '1-5 years' smoking cessation at enrolment of the study showed significantly lower decline in all lung function parameters ($p < 0.03$) than current smokers. The group ' ≤ 1 year' smoking cessation at enrolment was not significantly different from current smokers. Emphysema increase was significantly slower in the group that quit 1-5 years ($p < 0.045$), but was not likely of any clinical relevance. Smoking cessation stabilized lung function decline after ≥ 4 year cessation (> 1 year at enrolment plus 3-year follow-up), but it does not stabilize emphysema development during 3-years follow-up.

A last line focused on whether SNPs in the nicotinic acetylcholine receptor (nAChR) subunit genes are associated with increased lung function decline. RS1051730 and rs8034191 were genotyped in a population-based cohort of 1,226 heavy smokers (COPACETIC) with 3-year follow-up and a hospital-based cohort of 893 COPD patients (LEUVEN). All participants underwent pulmonary function tests and computed tomography (CT) of the chest at baseline. Lung function decline and emphysema progression was assessed over a median follow-up of 3 years in COPACETIC. Smokers homozygous for the rs1051730 A-allele had a more pronounced FEV₁/FVC decline compared to GG carriers (1.9%; $p = 0.014$). Former smokers with AA genotypes did not exhibit a more severe decline (0.54%; $p = 0.317$). Similar data were observed for GG genotypes of rs8034191 (2.2%; $p = 0.002$ for smokers and 0.75%; $p = 0.249$ for former smokers). In addition, in clinically-diagnosed COPD patients, the number of homozygous carriers of the rs1051730 A-allele gradually increased with increasing COPD severity. Variants of the nAChR genes are associated with accelerated lung function decline in current, but not in former smokers. The accelerated decline in lung function led to an increased risk of developing clinically important airflow obstruction.

4.1.5. The potential impact, the main dissemination activities and exploitation of results

The potential impact

Chronic Obstructive Pulmonary Disease (COPD) is a pulmonary disease characterised by airway obstruction on one hand and destruction of lung tissue on the other. The disease is relentlessly progressive and reduces life expectancy. The damage done to the lung tissues is irreversible: stabilising the disease is the best one can do at the moment. Predictions on the COPD burden to society reveal that it will become a leading cause of morbidity and mortality in the Western society. According to World Health Organization (WHO) estimates, 80 million people have moderate to severe COPD. More than 3 million people died of COPD in 2005, corresponding to 5% of all deaths globally. Total deaths from COPD are projected to increase by more than 30% in the next 10 years unless urgent action is taken (www.who.int).

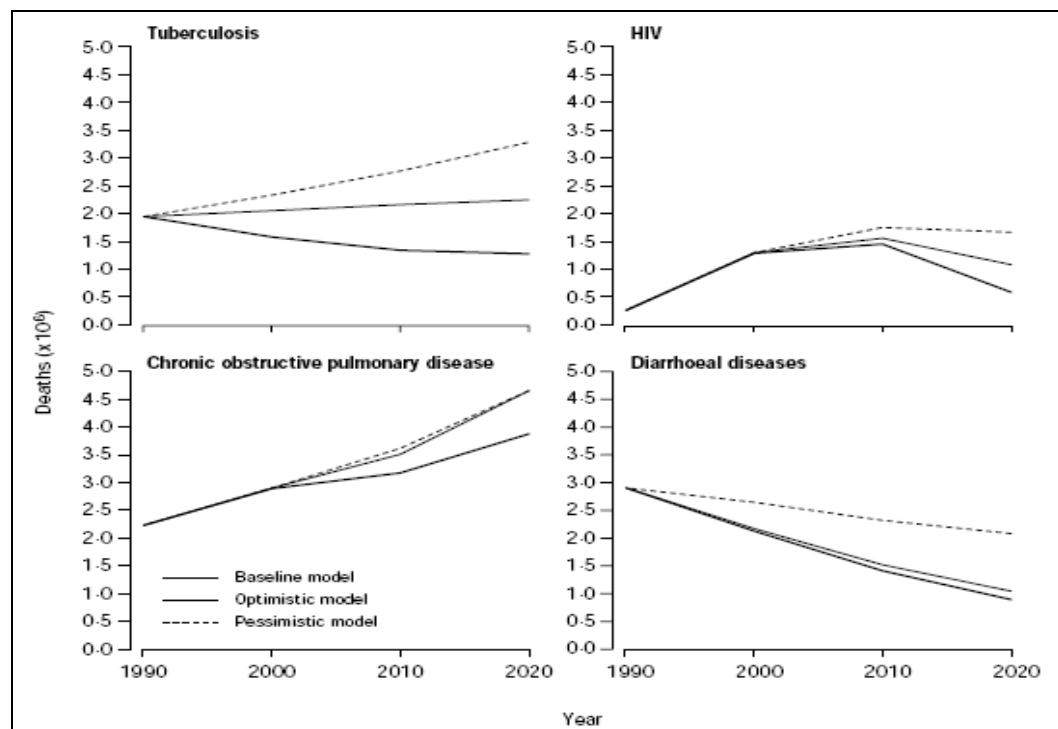


Figure 7 WHO estimates of total number of deaths due to tuberculosis, COPD, HIV and diarrhoea.

The important risk factor COPD is known: smoking. The problem with the disease is that COPD emerges over very long periods of time and affected subjects slowly adapt to the increasing severity of the disease by reducing their daily activities to the decreasing lung function.

Only when the loss of that is so severe and undeniable, medical help is sought in a too advanced stage of the disease.

The next problem with COPD is that not all smokers will suffer from it: the so called healthy smokers is a well known phenomenon opposing severely disabled smokers, dying at a much younger age. Tobacco load is a bad predictor for these two extremes and the evidence that a genetic background is responsible for this difference is accumulating.

The added value of the results of this project to the population at large is an obvious one: the possible prevention of COPD by advising susceptible subjects not to start smoking as the risks of developing COPD are high. The efficacy of specifically targeted campaigns to revert smoking habits could increase considerably. The health costs related to the morbidity and mortality of COPD can be reduced by magnitudes when we indeed succeed in preventing the development of COPD in susceptible subjects. Balancing the investment in terms of research costs and the diseases related costs over e.g. the next 25 years is a simple task with a highly positive outcome. For the risk groups not yet diagnosed with COPD accurate and reliable diagnosis of the disease in the earliest stages possible is of the utmost importance as prevention of further development of the disease is in the best interest of the patient. The current diagnostic approach with the sole use of pulmonary function testing is 'dangerous' as subjects in whom tissue destruction is present without concomitant airflow limitation are being missed. The challenge of the medical world is to identify a smoker as 'endangered' as early as possible in order to start risk-modifying therapy.

The impact of the COPACETIC project lies foremost in the definition of the genetic factors responsible for the difference in morbidity despite similar other risk factors as pack years smoking. As genetic factors can be tested at a young age, even before any smoking habit developed, the opportunities to prevent COPD are obvious. By means of simple tests it will be possible that susceptible smokers will become known and such subjects never must start smoking. Obviously smoking is a habit to be avoided at all times, but the hurdles to start smoking in this society are so low that it is a fairy tale to expect that world wide smoking will cease.

Even in those who started a smoking habit the outcome of this project will be helpful. Susceptible smokers often display a much faster decline of e.g. lung function than others and again genetic factor are held responsible for this phenomenon. In smokers it will be of the utmost importance to locate those rapid decliners and focus therapies on these subjects in order to stop them continuing smoking. At this moment stopping a smoking habit is the only known

therapy which prevents further decline of the lung function, an increase of symptoms and loss of quality of life. That therapy will be much more efficient when it is geared towards the group with the highest risk of either contracting the disease or of rapid decline of bodily functions.

The COPACETIC project has several advantages. One of the most important ones is that the outcome in terms of genetic factors is not biased. Many other studies studied genetic susceptibility in cohorts of non-, light and severe smokers. Rephrased the outcome can be explained in two ways: 1] the genetic factors reported are responsible for the noted differences in incidence or 2] the difference in pack years is responsible. The latter is very well possible and undermines any genetic influence: the difference in pack years noted could very well be the result of effective advertising in susceptible groups of e.g. youngsters. COPACETIC however selected subjects who are (or were) all smokers and consumed the same minimal amount of tobacco. Any possible bias due to difference in smoking is simply impossible and the outcome is much more valid. In that sense COPACETIC can be viewed upon as a new start following other less suited studies.

The outcome of this study can be translated into a risk-profile and knowing that profile for individuals a probability can be calculated to contract COPD or to suffer from an augmented decline of lung function. That probability can be determined at young age and used to guide measures / therapies to preventing the start of a smoking habit or to increase efficiency of stopping smoking therapies. Selection of high risk groups is and will always be essential in preventive medicine.

Another important factor is that COPACETIC did not consider COPD to be a simple disease characterised by obstruction only. COPD is a mixture of obstruction and emphysema and there is mounting evidence that these two phenomena are caused by different pathological processes. The baseline data of COPACETIC firmly point at this: many subjects show evidence of obstruction without emphysema and visa versa. If this study measured only one of the two phenomena we would generate results which have partial validity only. Conceive that only SNPs were found for obstruction and a risk profile was built on that, the above procedure of determining the prior risk of contracting COPD would be of a limited value. It is very well possible that a low risk profile for obstruction combines with a high risk profile for emphysema.

Measuring emphysema was therefore from the start on an important aspect of this project and much was learned. An important aspect was that inter-scanner differences prevent adequate centre-to-centre comparisons as each manufacturer appears to implement specific algorithms to measure voxel density. The correction factor designed based on the measurement of tracheal air density is a step forwards in characterising emphysema: future experiments will benefit from this approach significantly.

Next to the option to design tests for susceptibility, the notion that the reported outcome of this project will enhance the pathophysiologic thinking /science is evident. Knowing the 'causative genes' is the first step to unravel the mechanisms behind COPD. The malfunctioning genes will be closely scrutinised and examined to find which part of their normal function is defective or what proteins are not generated any more. What are the missing normal functions of protein and how does that loss influence the loss of pulmonary function. The reported outcome of this COPACETIC project will be the start of many new research projects digging deeper into the pathophysiology of COPD. Needless to say that it is hoped for that such research will also end up in effective drugs preventing further loss of pulmonary functions.

A last remark on the impact on collaboration world wide. COPACETIC as leading European study is considered as one of the major players in this field of research. It can be mirrored easily with a US based study and already frequent contact between the principal investigators were established. It is to be expected that in the near future these will intensify and the leading role of the EU in this field will only be amplified.

Dissemination

In the tables on the pages below we list the extensive number of oral presentations, poster and workshops based on the COPACETIC project. Some peer-review reports in high ranking journals already were published and many are to come. In the review of WP4 we stated that the research is still ongoing in order to strengthen the outcome of this project. In other words the main publication is not yet due but the consortium is very much confident that that major publication will be prepared and submitted in a foreseeable period of time. It is reassuring that the consortium received invitations to submit the outcome by the very top ranking journals in the medical field.

It goes without saying that publication of the main results will be the start of a whole series of related activities geared both to the professional public and to the lay public. Until that moment

we are forced to restrict outward bound activities in order not to jeopardise the publication of main results: policies of the top-ranking journals are clear in this.

Exploitation of results

As just stated, the research is still ongoing in order to strengthen the outcome of this project. Until that moment and being able to estimate the full the impact, exploitation is not a very realistic activity. Having said that some informal discussions with the industrial partner in this consortium are on-going and will be upgraded at the moment final results are available. Indeed the construction of test kits is an option.

For the very same reason of not having all results available, discussions on application for patents have been postponed: the consortium will decide on that later.

4.2 Use and dissemination of foreground

4.2.1. Section A (public)

<i>LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS (in chronological order)</i>										
no.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers (if available)	Is / will open access provided to this publication?
1	Common and different genetic background for rheumatoid arthritis and coeliac disease	Coenen M.J.H.	Human Molecular Genetics	12	Oxford Journals	Oxford	2009	4195-4203	ISSN 09646906	no
2	Meta-analyses on suspected chronic obstructive pulmonary disease genes: A summary of 20 years' research	Smolonska J.	ARJCCM	24	ATS	New York	2009	618-631	ISSN 1073449X	no
3	The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema	Lambrechts D	ARJCCM	24	ATS	New York	2010	486-493	ISSN 1073449X	no

<i>LIST OF DISSEMINATION ACTIVITIES (in chronological order)</i>								
No.	type of activities	main leader	title	date	place	type of audience	size of audience	countries addressed
1	poster	P. Zanen	COPACETIC: unravelling the genetics of COPD	12 -19 Sep 2008	Berlin	pulmonologists	~15000	world
2	workshop	M. Owsijewitsch	Comparison of an anatomical and a non-anatomical approach in distribution analysis of COPD associated emphysema	October 2008	Iowa	physicians	100	World
3	congress	M. Owsijewitsch	Distribution of emphysema between rind and core region of the lung in COPD patients	December 2008	Chicago	radiologists	70	World
4	congress	M. Owsijewitsch	Phenotyping of COPD: Comparison of an anatomical and a non-anatomical approach in studying the distribution of emphysema in upper and lower lung	March 2009	Vienna	radiologists	70	Europe
5	congress	S. Ley	Airway remodelling in COPD: Quantitative analysis of airway tapering and geometry – initial results	September 2009	Vienna	physicians	80	Europe
7	oral presentation	C. Wijmenga	first results of COPACETIC	12 -19 Sep 2009	Vienna	pulmonologists	~15000	world
8	poster	J. Smolonska	COPACETIC, a genome-wide association study on chronic obstructive pulmonary disease (COPD)	20 – 24 Oct 2009	Honolulu	geneticists	~5000	world
9	congress	M. Owsijewitsch	COPD Phenotyping in CT: Comparison of Visual and Densitometric	December 2009	Chicago	physicians	100	World

Emphysema Extent Assessment								
10	workshop	Kauczor-HU	COPD phenotyping	2-5 Feb 2010	Washington	physicians	80	USA
11	conference	Kauczor-HU	COPD phenotyping	8-11 April 2010	Yokohama	radiologists	250	Japan
12	poster	F. Mohamed Hoessein	detecting lung function decline in heavy smokers: choosing between FEV ₁ /FVC <70% or <LLN	15-19 May 2010	New Orleans	pulmonologists	~15000	world
13	oral presentation	A. Dijkstra	susceptibility to chronic mucus hypersecretion: a genome-wide association study	15-19 May 2010	New Orleans	pulmonologists	~15000	world
14	summer school	Kauczor-HU	COPD phenotyping	5-9 July 2010	Greifswald	physicians	70	Germany
15	poster	F. Mohamed Hoessein	effect of smoking cessation on lung function in a cohort of heavy smokers	18 -22 Sep 2010	Barcelona	pulmonologists	~15000	World
16	poster	J. Smolonska	COPACETIC, a genome-wide association study on chronic obstructive pulmonary disease (COPD)	18 -22 Sep 2010	Barcelona	pulmonologists	~15000	world
17	workshop	Kauczor-HU	Population-based imaging	5 Nov 2010	Rotterdam	physicians	50	Europe
18	poster	C. Mol	correction of quantitative emphysema measures with density calibration based on measurements in the trachea	28 Nov - 3 Dec 2010	Chicago	radiologists	~20000	world

4.2.2. Section B

LIST OF APPLICATIONS FOR PATENTS, TRADEMARKS, REGISTERED DESIGNS, ETC.

Type of IP Rights:	Confidential	Foreseen embargo date dd/mm/yyyy	Application reference(s) (e.g. EP123456)	Subject or title of application	Applicant (s) (as on the application)
--------------------	--------------	-------------------------------------	---	---------------------------------	---------------------------------------

No patents, trademarks, registered designs, etc. were applied for at the closure of the project

Type of Exploitable Foreground	Description of exploitable foreground	Confidential	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Not applicable								

4.3 Report on societal implications

A General Information	
Grant Agreement Number:	201379
Title of Project:	COPD Pathology: Addressing Critical gaps, Early Treatment & diagnosis and Innovative Concepts
Name and Title of Coordinator:	P. Zanen MD, PhD
B Ethics	
1. Did your project undergo an Ethics Review (and/or Screening)? If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports?	yes yes ¹
2. Please indicate whether your project involved any of the following issues:	
RESEARCH ON HUMANS	
• Did the project involve children?	
• Did the project involve patients?	
• Did the project involve persons not able to give consent?	
• Did the project involve adult healthy volunteers?	✓
• Did the project involve Human genetic material?	✓
• Did the project involve Human biological samples?	✓
• Did the project involve Human data collection?	✓
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	
• Did the project involve Human Foetal Tissue / Cells?	
• Did the project involve Human Embryonic Stem Cells (hESCs)?	
• Did the project on human Embryonic Stem Cells involve cells in culture?	
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	
PRIVACY	
• Did the project involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	✓
• Did the project involve tracking the location or observation of people?	
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	
• Were those animals transgenic small laboratory animals?	
• Were those animals transgenic farm animals?	
• Were those animals cloned farm animals?	
• Were those animals non-human primates?	
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	
• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	
DUAL USE	
• Research having direct military use	
• Research having the potential for terrorist abuse	
C Workforce Statistics	

¹ This project made use of the recruitment, data and logistics of the Nelson study. That study acquired the METC approval and therefore communicated with those bodies.

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).		
Type of Position	Number of Women	Number of Men
Scientific Coordinator	0	1
Work package leaders	4	1
Experienced researchers (i.e. PhD holders)	5	9
PhD Students	2	1
Other	-	-
4. How many additional researchers (in companies and universities) were recruited specifically for this project?		3
Of which, indicate the number of men:		1
D Gender Aspects		
5. Did you carry out specific Gender Equality Actions under the project?		no
6. Which of the following actions did you carry out and how effective were they?		
7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?		
<ul style="list-style-type: none"> Yes- the COPACETIC underlying Nelson study included foremost males: the subjects had to be at a high risk for COPD/lung cancer and much more males are at risk due to their significantly higher tobacco consumption. 		
E Synergies with Science Education		
8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?		
<ul style="list-style-type: none"> No 		
9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?		
<ul style="list-style-type: none"> Yes- a website (www.copacetic-study.eu) 		
F Interdisciplinarity		
10. Which disciplines (see list below) are involved in your project?		
<ul style="list-style-type: none"> Radiology, pulmonary, epidemiology and genetics 		
G Engaging with Civil society and policy makers		
11a Did your project engage with societal actors beyond the research community? (if 'No', go to Question 14)		no
11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?		
11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?		
12. Did you engage with government / public bodies or policy makers (including international organisations)		
13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?		
13b If Yes, in which fields?		

13c If Yes, at which level?		
H Use and dissemination		
14. How many Articles were published / accepted for publication in peer-reviewed journals?		
To how many of these is open access provided?		
How many of these are published in open access journals?		
How many of these are published in open repositories?		
To how many of these is open access not provided?		
Please check all applicable reasons for not providing open access:		
<input type="checkbox"/> publisher's licensing agreement would not permit publishing in a repository <input type="checkbox"/> no suitable repository available <input type="checkbox"/> no suitable open access journal available <input type="checkbox"/> no funds available to publish in an open access journal <input type="checkbox"/> lack of time and resources <input type="checkbox"/> lack of information on open access <input type="checkbox"/> other:		
15. How many new patent applications ('priority filings') have been made?		0
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	0
	Registered design	0
	Other	0
17. How many spin-off companies were created / are planned as a direct result of the project?		0
<i>Indicate the approximate number of additional jobs in these companies:</i>		-
18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:		0
<input checked="" type="checkbox"/> None of the above / not relevant to the project		
19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (FTE = one person working fulltime for a year) jobs:		10

I Media and Communication to the general public	
20. As part of the project, were any of the beneficiaries professionals in communication or media relations?	
<input type="radio"/> Yes <input checked="" type="radio"/> No	
21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?	
<input type="radio"/> Yes <input checked="" type="radio"/> No	
22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?	
<input checked="" type="checkbox"/> Press Release <input type="checkbox"/> Media briefing <input type="checkbox"/> TV coverage / report	<input checked="" type="checkbox"/> Coverage in specialist press <input checked="" type="checkbox"/> Coverage in general (non-specialist) press <input checked="" type="checkbox"/> Coverage in national press

<input type="checkbox"/> Radio coverage / report <input type="checkbox"/> Brochures /posters / flyers <input type="checkbox"/> DVD /Film /Multimedia	<input type="checkbox"/> Coverage in international press <input type="checkbox"/> Website for the general public / internet <input type="checkbox"/> Event targeting general public (festival, conference, exhibition, science café)
23 In which languages are the information products for the general public produced?	
<input type="checkbox"/> Language of the coordinator <input type="checkbox"/> Other language(s)	<input type="checkbox"/> English