|  |
| --- |
| I. Project Title |
| Genes related to Genetic And Rare lung Diseases (GARD) as risk for LC | **Date: 2018-12-03** |
| **II. Project Group Investigators (List Names and Institutes)** *(for external investigator- it is required to include at least 1 ILCCO/TRICL investigator in the Project Group)* | **Primary Contact****Information** |
| Albert Rosenberger, Ph.D., UMGHeike Bickeböller, Ph.D. , UMGJulia Heck, Ph.D., UCLAChristopher Amos, Ph.D. , Baylor College | **Name:** Dr. Albert Rosenberger**Email:****arosenb@gwdg.de****Tel: ++49 (551) 39-14044** |
| III. Background |
| The US Genetic and Rare Diseases (GARD) Information Center labels 139 uncommon diseases as at least related to the lung [1]. This set of diseases is quit heterogeneous, some affects children, newborns or even unborn; other mainly affects older adults; some are considered as (more or less) mono-genetic, like Alpha-1 antitrypsin deficiency (AATD), Cystic fibrosis (CF), or the Birt-Hogg-Dube syndrome; others are multi-genetic in nature, like idiopathic pulmonary fibrosis (IPF) / Cryptogenic fibrosing alveolitis (CFA) or sarcoidosis (SAR). We want to investigate whether genes associated with these GARDs contribute to the genetic susceptibility of lung cancer (LC). Therefore we cut down the list of GARDs to 62 lung diseases that are not mainly environmentally caused, but with associated or causing genes reported in the literature or relevant data bases (ORPHANET [2], MONARCH [3], OMIM [4]). We distinguish between n=18 lung specific GARDs and n=44 multi-organ GARDs. We then tracked all related genes, excluding genes already associated with LC, genes of the HLA region and genes not covered by the OncoArray. Finally we achieved 72 genes of lung specific GARDs and 82 genes of multi-organic GARDs (see figure 1). Some of these GARDs are associated with an increased risk of LC or cancer in general. * For AATD, e.g. an increased risk for LC of individuals with any kind of AATD phenotypes was observed compared to those with non-deficient AAT variants (OR = 4.51, 95% CI = 1.66-12.29) [5]. Restricted to passive smokers the risk increase was even stronger (OR=12.1, 95% CI: 1.2-123) [6].
* For CF an increased risk for gastrointestinal cancer SIR=8.13 (95% CI: 6.48-10.21) [7] has been observed, although no increased risk was seen for all cancers combined(SIR=1.1, 95% CI: 1.0-1.3) [8]. It has pointed out, that the number of LCs was - in turn - lower than expected. [9]
* It is still unclear, if SAR patients have to face an increased risk for LC. Estimates of the relative risk range between RR=3.24 (95% CI: 0.84-1.71) [22] and RR=0.60 (95% CI: 0.42-0.85, n=5768) [36].
* Also higher risk for LC was observed in CFA/IPF patients, e.g. a SMR=7.4 (95% CI: 5.42-9.88), which is stronger in current smokers and given an exposure to asbestos [10].

For more detail see “Background in detail”. |
| IV. Specific Aims |
| **Aim 1: GARD gen-set investigation** Aim 1.1: To examine potential associations of all genes related to the selected GARDs with LC with respect to sex, age, ethnicity, smoking or histological subtype.Aim 1.2: To examine a potential “joint association” within the set of selected gene (genes sets); this includes selecting the “significance driving genes” in terms of leading edge markers/genes of a gene-set enrichment analysis (GSEA).Aim 1.3: To examine differences of “joint associations” between the 72 genes of lung specific GARDs and 82 genes of multi-organic GARDs, respectively by further dividing these sub-gene-sets into inflammatory- and non-inflammatory related genes. **Aim 2: selected disease investigation**Aim 2.1: In case an association is found for genes related to AATD, we want to investigate potential gene-gene interactions of SERPINA1 and ELANE, PPIC, PPP4R4 or other genes of the gene family HGNC739: Serpin peptidase inhibitors (SERPIN, see table 1).Aim 2.2: In case an association is found for genes related to CF, we want to concentrate the investigation of the CFTR gene and examine gene-gene interactions with other IPF-related genes or other genes of the gene family HGNC807: ATP binding cassette subfamily C (ABCC) (see table 2).Aim 2.3: In case an association is found for genes related to SAR, we want to concentrate the investigation of the three SAR-related genes HLA-DRB1, BTNL2 and ANXA11 and examine gene-gene interactions Aim 2.4: A deeper investigation of CFA/IPF-related genes, is part of the ILCCO/TRICL-proposal “Role of surfactant genes in lung cancer risk” by Jennifer Luyapan, Chis Amos et al. |
| V. Methods |
| **For the Aims 1.1 to 1.3*** We will first perform single- and multi-marker-association analyses, adjusted for sex, age, ethnicity, lung conditions (e.g. COPD), smoking and markers of known genetic lung cancer loci. Nesting of estimates within the study or geographic region is considered to mimic random effects meta-analysis. If appropriate, these models will be fitted conditional to lung conditions, sex or smoking, and stratified to histological subtypes. All SNPs within LD-blocks covering at least one of the 72 genes of lung specific GARDs or 82 genes of multi-organic GARDs are selected to be analysed.

We want further perform gene-set enrichment analysis and apply network-based kernel machine test to investigate interactions within these genes. This will be carried out stratified and to compare for genes of lung specific GARDs, genes of multi-organ GADS, as well as inflammatory and non-inflammatory related genes.(The set of 72 genes of lung specific GARDs contains 8 “surfactant genes”, investigated by Jennifer Luyapan et al.. We will investigate these only as part of the proposed gene set).* In parallel we will screen the literature for markers/loci being of specific interest in the considered GARDs (“going beyond the level of genes”). We will look for accurate proxies if not genotyped by the OncoArray, or request for additional genotyping in available samples after power considerations, according to Immunology Database and Analysis Portal (ImmPort) [11].

**For the Aims 2.1 to 2.3**We will carry out deeper analysis for the AATD, CF and SAR, in the case associations are found in related genes. This includes multi-SNP haplotype analysis [12] and taking potential interaction with lung-relevant haplotypes of HLA-region [13] and other known lung cancer loci’s into account. This can be a valuable step in particular for AATD, since Thun, et al. (2013) [14] have demonstrated that associations of rare variants of SERPINA1A with COPD are easily been masked by common variants. We may complete the investigations with a mediation analysis [15], aiming to distinguish between direct and indirect effects, considering COPD or other lung conditions as mediator.For AATD we will also plan to investigate an association between a “predicted AATD-serum level” with LC that is based on a genetic score of Thun et al. [14]) A replication of findings will be carried out on the basis of the non-OncoArray genotype data of the ILCCO/TRICL consortium (data will be requested later as needed). |
| **VI. Materials or variables needed from the study Pi\*S (Please separate out the required vs optional components)** |
| The following will be required from each participating case-control or cohort study:**Phenotypic information:**1. ID number
2. Case/control status
3. Cases: histological subtype
4. Sex
5. Cases: age at diagnosis / non-cases: age at recruitment
6. Ethnicity
7. Study location / name / design
8. smoking (smoking status, pack years, age at first cigarette, exposure to environmental tobacco smoke/passive smoker)
9. Lung conditions (previous COPD, bronchitis, asthma, emphysema, others) and age at diagnosis
10. Other comorbid conditions including history of liver diseases (if available)
11. Personal and family cancer history (if available)
12. Effect measure modifiers of AATD related diseases, such as alcohol use, diet, physical activity (if available)

**Genotypic information (genotyped):**1. Genotypes of marker in GARD related genes
2. Genotypes of markers of the gene family HGNC739 (Serpin peptidase inhibitors)
3. Genotypes of markers of the gene family HGNC807 (ATP binding cassette subfamily C (ABCC))
4. Genotypes of marker in known lung cancer loci
5. Haplotypes of the HLA-region: 25 to 35 Kb at chromosome 6 (NCBI build 37), as achieved applying SNP2HLA v1.0.3 in the project “Fine mapping of MHC region in lung cancer“ by Ferreiro-Iglesias, et al. (2018) [13]
 |
| **VII. Time line** |
| We aim to get access to relevant data from participating studies in 1st quarter of 2019. We will request for additional genotyping after screening the literature for markers/loci being of specific interest in the considered GARDs, within the 1st quarter of 2019. Data pooling, checks and analysis of aims 1.1 to 1.3 are planned for spring/summer 2019. In case of results value to be published: writing a manuscript in Autumn /Winter 2019Analysis of aims 2.1 to 2.3 are planned for Autumn 2019 to Spring 2020. Writing a manuscript in Spring /Summer 2020 |
| **VIII. Funding Sources and Declaration of Conflict of Interests.** *To ensure full transparency and to protect collaborating study Pi\*S, ILCCO/TRICL requires the Project Leaders to disclose any circumstances that could give rise to a potential conflict of interests related to the proposed project activity in particular, or to LC and/or tobacco products in general, including but not limited to funding sources, employment and consulting, board membership and investment interests within the last 5 years.*  |
| None of the investigators has a conflict of interests.There is an Alpha-1 Foundation which provides small grants. Due date for in-cycle applications is February 2019, and there are out-of-cycle applications as well. |
| **IX. Other remarks** (e.g. **publication** plan, etc.) |
| There is a little overlap to the project “Role of surfactant genes in lung cancer risk” by Jennifer Luyapan, Chris Amos et al. (August 28, 2018). This was taken into account to avoid duplicate investigations.This proposal replaces two (former) proposals 1. Lung cancer risk among carriers of SERPINA1 variants, by Julia Heck (3 Sept 2016, later withdrawn)
2. SERPINA1-induced Alpha-1 antitrypsin deficiency as a risk for lung cancerby Albert Rosenberger (12 July 2018)

The results are planned to be published irrespective of whether association will be observed or rejected, as well as presented at national or international conferences.  |

Figure 1: selection of diseases and genes of interest



# Background in detail:

### Alpha-1 antitrypsin deficiency (AATD)

Alpha-1 antitrypsin deficiency (AATD) is a widespread autosomal genetic disorder predisposing individuals to develop chronic obstructive pulmonary disease (COPD) and emphysema, may be linked to asthma, and it is among the most common causes of liver diseases in children. AATD is the most frequent cause of COPD among nonsmokers. Persons with this disorder have functional genetic polymorphisms in *SERPINA1* (14p32.13; *serpin family A member 1* , also known as protease inhibitor Pi) causing inadequate production of the protease inhibitor α-1-antitrypsin, which protects lung tissue elastin from neutrophil elastase-mediated destruction by binding and inhibiting elastase. Neutrophil elastase is involved in bronchoconstriction and airway hyperresponsiveness, and may contribute to airway inflammation. Low AAT serum levels fail to protect pulmonary tissue from enzymatic degradation [14]. The prevalence of AATD in European populations is 1% to 2%, but this is potentially underestimated [16].

Genes coding for an excess of neutrophil elastase have previously been associated with LC [17]. Carcinogenesis may occur due to inflammation, a relevant process in LC, particularly among patients with COPD [18]. Alternatively, because some tumors are capable of synthesizing and excreting protease to facilitate invasion during neoplasia, AATD may favor the invasion of healthy tissue by neoplastic cells. The presence of AAT could protect the healthy tissue by opposing its growth or by inhibiting the circulating proteases.

Several infrequent but functional genetic polymorphisms in the SERPINA1 gene are known to substantially reduce concentration of alpha1-antitrypsin (AAT) in the blood.

There are four prominent variants of AAT, named according to the pace of AAT moving in an electrophoretic field (F=fast, M=medium, S=slow, Z=very slow). The Pi\*M allele is most commonly carried by 85%-90% of Caucasians [6,19]. However, the low-frequent alleles Pi\*Z and Pi\*S cause more than 95% of AATD cases. Pi\*S is most common in persons from the Iberian peninsula, and Pi\*Z alleles more common in Northern Europe. An estimated 9.6% of European-origin ILCCO participants will be carriers of at least one deficient allele. The prevalence of these alleles is believed to be low in persons of African, Asian, or South American descent, although few prevalence surveys have been published of those regions. Much of the research on AATD has focused on Pi\*ZZ or Pi\*SS genotypes as they have the poorest health outcomes, yet little is known about the health risk of heterozygotes. Due to low rates of testing [20], heterozygotes are rarely diagnosed, and as such, have not been the focus of most AATD research.

Pi\*Z and Pi\*S are also associated with an increased risk of early onset chronic obstructive pulmonary disease (COPD), with lung function (e.g. FEV1 in ever-smokers) and the risk of lung other diseases (e.g. asthma); or more generally with AATD-phenotypes (e.g. circulating AAT levels) [14,21-24]. Genetic interaction with ELANE (19p13.3; *elastase, neutrophil expressed*, also known as ELA2), PPP4R4 (next to SERPINA1; *protein phosphatase 4 regulatory subunit 4*) or PPIC (5q23.2; *peptidylprolyl isomerase C*) are supposed [24,25].

Several studies have examined cancer risk among persons with either heterozygous or homozygous AATD deficiency, with most studies focusing on liver carcinoma, of which there is a fairly well-demonstrated risk. LC and COPD share a common etiology (Brenner 2012), and despite the link between AATD and COPD, few studies have examined whether AATD is related to later LC risk, with most published studies having small sample sizes. In recent years a potential association of SERPINA1 with LC was investigated. The results are controversial and may be confounded by COPD or smoking.

Topic, et al. (2006) [5] reported an (almost significant) increase risk for LC of individuals with any kind of AATD phenotypes (Pi\*M, Pi\*S, Pi\*Z and others). In this Serbian study of 186 patients, the risk to develop squamous cell LC was strongest elevated comparted to those with non-deficient AAT variants (OR = 4.51, 95% CI = 1.66-12.29). Topic, et al. (2011) [25] further identified the Pi\*M1 allele (a subtype of Pi\*M) as risk factor for haematological malignancies (HM). Torres-Duran, et al. (2015) [6] observed the Pi\*S allele in non-smoking, non-COPD cases to increase the LC risk (OR=4.64, 95% CI: 1.08-19.92) and even stronger in passive smokers (OR=12.1, 95% CI: 1.2-123); compared to those with normal genotype Pi\*M. They were not able to examine the risk from Pi\*Z alleles perhaps due to low prevalence. Yang, et al. (2008) [19] observed, in the largest study done to date included 1856 US LC patients, an increased LC risk for any carrier of Pi\*S or Pi\*Z (OR=1.7, 95% CI: 1.2-2.4) which was strongest in COPD-positive never-smokers (OR=5.9, 95%CI: 2.7-12.8); as it was observed by Enewold, et al. (2012) [26] but only regarding NSCLC in COPD-positive African-Americans (OR=7.4, 95% CI: 1.03-53). Much earlier Schwartz, et al. (1998) [27] reported an OR of 13.8 (95% CI: 1.4-136) for non-smoking carriers of at least one risk allele under the age of 60, but no thereafter (OR=1.7, 95% CI: 0.3-8.3).

In contrast, no association with LC was observed in a register-based cohort study based on the Swedish National AAT Deficiency Register and covering the period 1991-2014 (SMR=0.9, 95% CI: 0.4-1.7), perhaps because most patients who have been tested, diagnosed, and listed in a Register are Pi\*ZZ, and they are more likely to die at an earlier age of other AATD-related diseases. An association with respiratory diseases (SMR=48, 95% CI: 43-54) was observed instead [28]. Also El-Akawi, et al. (2006) [29] missed to identify any Pi\*S or Pi\*Z in bronchial tissues taken from a series of 100 LC patients. In contrast, Yang, et al. (1999) [30] observed that 12.3% out of a series of 260 patients carried an AATD allele, which was significantly higher than expected.

In summary, a potential association of SERPINA1 with LC seems complex with regard to sex, age, ethnicity, smoking and COPD-status. Stratified effects, confounding but also mediated association may explain the inconsistency of previous findings, since COPD is known to be associated with LC [31]

Table 1: Serpin peptidase inhibitors gene family

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| proved Symbol | Approved Name | Previous Symbols | Synonyms | Chromosome | Cluster |
| SERPINC1 | serpin family C member 1 | AT3 | ATIII, MGC22579 | 1q25.1 |  |
| AGT | angiotensinogen | SERPINA8 |  | 1q42.2 |  |
| SERPINE2 | serpin family E member 2 | PI7 | PN1, GDN, PNI, nexin | 2q36.1 |  |
| SERPINI1 | serpin family I member 1 | PI12 |  | 3q26.1 | 1 |
| SERPINI2 | serpin family I member 2 | PI14 | PANCPIN, TSA2004, MEPI, pancpin | 3q26.1 | 1 |
| SERPINB1 | serpin family B member 1 | ELANH2 | EI, PI2, anti-elastase | 6p25.2 | 2 |
| SERPINB6 | serpin family B member 6 | PI6, DFNB91 | PTI, CAP | 6p25.2 | 2 |
| SERPINB9 | serpin family B member 9 | PI9 | CAP3 | 6p25.2 | 2 |
| SERPINE1 | serpin family E member 1 | PLANH1, PAI1 | PAI | 7q22.1 |  |
| SERPING1 | serpin family G member 1 | C1NH | C1IN, C1-INH, HAE1, HAE2 | 11q12.1 |  |
| SERPINH1 | serpin family H member 1 | CBP1, CBP2, SERPINH2 | HSP47, colligen | 11q13.5 |  |
| SERPINE3 | serpin family E member 3 |  |  | 13q14.3 |  |
| **(PPP4R4)\*** |  |  |  |  |  |
| **SERPINA1** | serpin family A member 1 | PI | AAT, A1A, PI1, alpha-1-antitrypsin, A1AT, alpha1AT | 14q32.13 | 3 |
| SERPINA10 | serpin family A member 10 |  | PZI, ZPI | 14q32.13 | 3 |
| SERPINA11 | serpin family A member 11 |  |  | 14q32.13 | 3 |
| SERPINA12 | serpin family A member 12 |  | OL-64, Vaspin | 14q32.13 | 3 |
| SERPINA13P | serpin family A member 13 | SERPINA13 | UNQ6121 | 14q32.13 | 3 |
| SERPINA2 | serpin family A member 2 | PIL, SERPINA2P | ATR, ARGS | 14q32.13 | 3 |
| SERPINA3 | serpin family A member 3 | AACT | ACT | 14q32.13 | 3 |
| SERPINA4 | serpin family A member 4 | PI4 | KST, KAL, KLST, kallistatin | 14q32.13 | 3 |
| SERPINA5 | serpin family A member 5 | PLANH3, PCI | PAI3, PROCI | 14q32.13 | 3 |
| SERPINA6 | serpin family A member 6 | CBG |  | 14q32.13 | 3 |
| SERPINA9 | serpin family A member 9 |  | CENTERIN, SERPINA11b, GCET1 | 14q32.13 | 3 |
| SERPINF1 | serpin family F member 1 | PEDF | EPC-1, PIG35 | 17p13.3 | 4 |
| SERPINF2 | serpin family F member 2 | PLI | API, ALPHA-2-PI, A2AP, AAP | 17p13.3 | 4 |
| SERPINB10 | serpin family B member 10 | PI10 | Bomapin | 18q21.3 | 4 |
| SERPINB11 | serpin family B member 11  |  | EPIPIN | 18q21.33 | 5 |
| SERPINB12 | serpin family B member 12 |  | YUKOPIN | 18q21.33 | 5 |
| SERPINB13 | serpin family B member 13 | PI13 | HUR7, hurpin, headpin | 18q21.33 | 5 |
| SERPINB3 | serpin family B member 3 | SCC, SCCA1 | T4-A, HsT1196 | 18q21.33 | 5 |
| SERPINB4 | serpin family B member 4 | SCCA2 | PI11, LEUPIN, SCCA-2, SCCA1 | 18q21.33 | 5 |
| SERPINB5 | serpin family B member 5 | PI5 | Maspin | 18q21.33 | 5 |
| SERPINB7 | serpin family B member 7 |  | MEGSIN | 18q21.33 | 5 |
| SERPINB2 | serpin family B member 2 | PLANH2, PAI2 | HsT1201 | 18q21.33-q22.1 | 6 |
| SERPINB8 | serpin family B member 8 | PI8 | CAP2 | 18q22.1 | 6 |
| SERPIND1 | serpin family D member 1 | HCF2 | HC-II, HLS2, HC2, D22S673 | 22q11.21 |  |
| SERPINA7 | serpin family A member 7 | TBG |  | Xq22.3 |  |

\* not member of the Serpin peptidase inhibitors gene family

### Idiopathic pulmonary fibrosis (IPF)/ Cryptogenic fibrosing alveolitis (CFA).

Idiopathic Pulmonary Fibrosis (IPF), a fibrosing interstitial lung disease (ILD) of unknown aetiology, primarily affects older adults and leads to a progressive decline in lung function and quality of life. With a prevalence of less than 1 case per 5000 patients and median survival of 3–5 years, IPF remains the most common and deadly of the idiopathic interstitial pneumonias (IIPs). [10,32]

Roughly 70–80% of individuals with IPF endorse a history of cigarette smoking, which has long been an established IPF risk factor. Not surprisingly, about 30% of IPF patients have concurrent pulmonary emphysema, including 8–27% with ≥10% emphysematous involvement throughout the lungs.

Compared to those in the general population, individuals with IPF have an increased risk of developing LC. Le Jeune, et al. (2007) [33] investigated a series of 1064 incident cases of IPF from the UK and reports a incidence rate ratio for overall cancer of IR=1.51 (95% CI = 1.20-1.90). This was largely due to a marked increase in the incidence of LC (IR = 4.96; 95% CI 3.00-8.18). In a prospective cohort of 890 IF- cases and 5884 controls the RR for LC was estimates as 7.31 (95% CI: 4.47-11.93), which remained when restricting to current smokers. A detailed investigation is given by Harris, et al. (2010) [10]. First they sum up the 14 case series reporting (pooled) 17% of CFA/IPF patients developing LC. However, this may be overestimated given the observation of 9% in 588 patients in the British Thorax Survey (BTS). A comparison with the general population yielded an SMR=7.4 (95% CI: 5.42-9.88). Strongest in younger (SMR=10.72), current smokers (SMR=14.67) and given an exposure to asbestos (SMR=10.67, of at least 10 years SMR=18.58). Thus they conclude that “*… in the light of the questionable temporality of this association, the confounding effects of smoking, the high rate of asbestos exposure among this cohort and the possibility of inherent ascertainment bias highlighted by Berkson,* [a] *causal relationship of CFA*[/IPF] *and LC remains questionable*.” It is also unclear, if the LC risk in CFA/IPF patients correlates positively [34] or negatively [10] with the duration of IPF (time since diagnosis). The link between IPF and a history of cigarette smoking or exposure to asbestos may explain only a portion of the increased LC risk; particularly occupational risk factors have been identified yet: e.g. stone cutting/polishing (OR = 3.9, 95% CI: 1.2, 12.7); hairdressing (OR = 4.4, 95% CI: 1.2, 16.3); farming and gardening (OR = 2.73, 95%CI = 1.47–5.10) [32,35-37].

### Sarcoidosis(SAR)

Sarcoidosis is a multi-system disorder; it is also a systemic inflammatory disease of unclear etiology typically affecting the lymphatic system and the lungs. The presence of non-caseating granuloma is the histopathological hallmark of the disease. The annual incidence of SARamong adults in is about 1 case per 10,000 people. SARaffects people of all racial and ethnic groups and occurs at all ages. However, the incidence is higher northern European countries (5 to 40 cases per 100,000 people) and about 3.5-times higher in African than in European Americans (35.5 AA: 10.9 EA cases per 100,000 people). In general the incidence peaks age 20 to 39 years; in Japan it peaks in the third, within African Americans in the fourth decade of life. In Scandinavia, the incidence in women appears to be bimodal, with one peak at 25 to 29 years of age and another at 65 to 69 years of age.

The overall mortality of patients with SARwas not different from the general population (SMR: 0.90; 95% CI, 0.74–1.08) [38]. The etiology of SARis unknown, but it is hypothesized that a combination of genetic and environmental factors contribute to its occurrence [39].

A history of ever smoking cigarettes was less frequent among SARpatients than control subjects (OR=0.62, 95% CI: 0.50-0.77) [40]. This observation is consistent with several earlier observations ([41-43]and a recent investigation in 345 incident cases of SARand 345 controls, where the risk reduction in current smokers compared to never smokers (OR=0.34, 95% CI: 0.23 – 0.50) was more elevated than if former smokers (OR=0.68; 95% CI, 0.45 – 1.01) [44].

Bonifazi, et al. (2015) [45] combined 8 cohort studies on the LC risk in a meta-analyse including 27,268 and SARpatients and 153 cancer cases. The calculated a pooled relative LC risk of RR=1.20 (95% CI: 0.84-1.71), however observed strong heterogeneity among the studies. The individual RRs of the most recent studies ranged from RR=3.24 (95% CI: 0.84-1.71, n=1153 from the UK [33]), over RR= 0.96 (95% CI: 0.71-1.27, n=10037 from Sweden[46]) to RR=0.60 (95% CI: 0.42-0.85, n=5768 from the US [47]). However, there was also an inverting trend by time since diagnosis: with an RR=1.73 (95% CI, 1.05-2.85) until the fourth year, RR=1.53 (95% CI, 0.87-2.68) between fifth and ninth year, and RR=0.48 (95% CI, 0.34-0.67) from the tenth year onwards.

### Cystic fibrosis (CF)

Cystic fibrosis (CF) is a common autosomal and recessively inherited condition and the commonest genetic cause of bronchiectasis in the Caucasian population. The incidence in Europe is on average 1 in 2000-3000 new-borns, but even where populations appear relatively homogeneous, there may be marked local and regional variations. In France, for example, there is a very high incidence of CF in Northwest Brittany and a lower incidence in the South. The incidence of CF in North America is about 1 in 3500, in Australia about 1 in 2500 [48,49].

The putative gene is the CF transmembrane conductance regulator (CFTR) [50] CFTR gene mutations have been well characterized in most European populations. The Δ508 CFTR mutation is the most common mutation causing CF. In central, northern, western, and north-eastern Europe, Δ508 has a frequency of about 70% within CF cases [51], but this varies from a maximum of 100% in the isolated Faroe Islands, to a minimum of about 20% in Turkey [48]. Other CF-causing alleles are substantially heterogeneous, with fewer than 20 mutations occurring at a worldwide frequency of more than 0.1% [52]. Some mutations can reach a higher frequency in certain populations, due to a founder effect in religious, ethnic or geographical isolates [53].

Maisonneuve, et al. (2013) [8] investigated 41188 CF-patients between 1990 and 2009 with respect to cancer. The reported no significantly increased overall cancer risk compared to expectations based on the SEER program (SIR=1.1, 95% CI: 1.0-1.3). However, the risk for a digestive tract cancer was elevated (SIR=3.5, 95% CI: 2.6-4.7). This confirms earlier observations [9], and was confirmed by a resent meta-analysis on 99255 patients, which reports a pooled SIR=8.13 (955 CI: 6.48-10.21) for gastrointestinal cancer [7]. However, the nub of matter is that the number of 3 cancers in the respiratory, intrathoracic organs (ICD-9-CM: 160-165, this includes the lung) is lower than the expected 7.4. This has previously observed by Schoni, et al. (1996) [9] as well, who stated: “*Interestingly, this excess* [in overall cancers] *is confined to the gastrointestinal tract, whereas only one tumour* [out of 39] *originated from the lung was found. If the CF gene was itself responsible for cancer in these patients, then one would have expected to see a more uniform increase in the risk of cancer in various organ systems*.“

Table 2: ATP binding cassette subfamily C (ABCC)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Approved symbol | Approved name | Previous symbols | Synonyms | Chromosome |
| ABCC1 | ATP binding cassette subfamily C member 1 | MRP,MRP1 | GS-X | 16p13.11 |
| ABCC2 | ATP binding cassette subfamily C member 2 | CMOAT | DJS,MRP2,cMRP | 10q24.2 |
| ABCC3 | ATP binding cassette subfamily C member 3 | MRP3,cMOAT2,EST90757,MLP2,MOAT-D | 17q21.33 |
| ABCC4 | ATP binding cassette subfamily C member 4 | MRP4,EST170205,MOAT-B,MOATB | 13q32.1 |
| ABCC5 | ATP binding cassette subfamily C member 5 | MRP5,SMRP,EST277145,MOAT-C | 3q27.1 |
| ABCC6 | ATP binding cassette subfamily C member 6 | ARA,PXE | MRP6,EST349056,MLP1,URG7 | 16p13.11 |
| CFTR | cystic fibrosis transmembrane conductance regulator | CF,ABCC7 | MRP7,ABC35,TNR-CFTR,dJ760C5.1,CFTR/MRP | 7q31.2 |
| ABCC8 | ATP binding cassette subfamily C member 8 | SUR,HRINS | HI,PHHI,SUR1,MRP8,ABC36,HHF1,TNDM2 | 11p15.1 |
| ABCC9 | ATP binding cassette subfamily C member 9 | SUR2,CMD1O | 12p12.1 |
| ABCC10 | ATP binding cassette subfamily C member 10 | EST182763,MRP7,SIMRP7 | 6p21.1 |
| ABCC11 | ATP binding cassette subfamily C member 11 | MRP8 | 16q12.1 |
| ABCC12 | ATP binding cassette subfamily C member 12 | MRP9 | 16q12.1 |
| ABCC13 | ATP binding cassette subfamily C member 13 (pseudogene) | PRED6,C21orf73,ABCC13P | 21q11.2 |

### Genomic lung cancer loci

Genomic lung cancer loci are markers listed in Supplement S2 to McKay, et al. (2017) [54] (restricted to genome-wide significance of p<5x10-8), or markers mentioned as associated with LC by Timofeeva, et al. (2012) [55]; Timofeeva, et al. (2009) [56], Rosenberger, et al. (2017) [57], Brenner, et al. (2013) [58], Brenner, et al. (2015) [59], Ji, et al. (2018) [60]; Truong, et al. (2010) [61], Truong, et al. (2010) [62], Wang, et al. (2014) [63] and Feng, et al. (2018) [64].

### Requested genotypes of 89 markers lung cancer loci

### Identified by McKay, et al. (2017) [54] either mentioned in table 2 or from supplementary S2 (marker associations with p<5\*10-7 and max(OR,1/OR)≥ 1.125)or others(main genetic lung cancer loci or next genotyped marker):

rs71658797 (1p31.1, AK5 FUBP1) genotyped

rs34517439 (1p31.1, DNAJB4 FUBP1 GIPC2 and others) genotyped

rs1056836 (2p22.2, CYP1B1) genotyped

rs722864 (2q31.1, MAP3K20) genotyped

rs7636839 (3q28, TP63) as proxy for rs13080835

rs402710 (5p15.33, CLPTM1L) genotyped

rs4530805 (5p15.33, MIR4457 TERT) genotyped

rs61574973 (5p15.33, MIR4457) as proxy for rs60622800

rs61574973 (5p15.33, MIR4457) as proxy for rs6554758

rs4635969 (5p15.33, MIR4457 TERT) as proxy for rs7446461

rs7705526 (5p15.33, TERT) genotyped

rs2853677 (5p15.33, TERT) genotyped

rs4635969 (5p15.33, MIR4457) genotyped

rs61574973 (5p15.33, CLPTM1L MIR4457) genotyped

rs4975616 (5p15.33, CLPTM1L MIR4457) genotyped

rs421629 (5p15.33, CLPTM1L) genotyped

rs380286 (5p15.33, CLPTM1L) genotyped

rs467095 (5p15.33, CLPTM1L) genotyped

rs40181 (5p15.33, CLPTM1L LINC01511) genotyped

rs3130477 (6p21.33, HCP5 MICA) as proxy for rs114002231

rs3130288 (6p21.32, PPT2-EGFL8) as proxy for rs115860703

rs1233385 (6p22.1, GABBR1 MOG) as proxy for rs115870917

rs886422 (6p21.33, DDR1) as proxy for rs116551911

rs3094604 (6p21.33, HCP5 MHC) as proxy for rs116822326

chr6\_27815639\_A\_G (6p22.1, HIST1H2BN) as proxy for rs13194781

rs2734986 (6p22.1, GABBR1 MOG) as proxy for rs138488080

rs3116813 (6p22.1, HCG4 IFITM4P) as proxy for rs147680653

rs2233956 (6p21.33, C6orf15) as proxy for rs2233956

rs34158769 (6p22.2, BTN3A2 HIST1H4H) genotyped

rs34493019 (6p22.2, HIST1H1E HIST1H2AC) as proxy for rs36109883

rs3769201 (6p21.33, CSNK2B) genotyped

rs439553 (6q27, RNASET2) as proxy for rs6920364

rs2187668 (6p21.32, HLA-DQA1) as proxy for rs74942078

rs17598658 (6p22.2, HIST1H2BD HIST1H2BE) genotyped

rs200484 (6p22.1, HIST1H2AI HIST1H2AJ HIST1H2AK and others) genotyped

rs188015 (6p22.1, HIST1H2BO OR2B2) genotyped

rs13197574 (6p22.1, AL3589331 OR1F12P1 RP1265C2481 and others) genotyped

rs34662244 (6p22.1, AL3589331 OR1F12P1 RP1265C2481 and others) genotyped

rs13213152 (6p22.1, ZKSCAN3 ZSCAN12) genotyped

rs3094222 (6p21.33, C6orf15 PSORS1C1) genotyped

rs1800629 (6p21.33, LTA TNF) genotyped

rs3132445 (6p21.33, MSH5 MSH5-SAPCD1 MSH5SAPCD1) genotyped

rs3131378 (6p21.33, MSH5 MSH5-SAPCD1 MSH5SAPCD1 and others) genotyped

rs1270942 (6p21.33, CFB DOM3Z RDBP and others) genotyped

rs2640726 (8p21.2, EPHX2) as proxy for rs2741354

rs2439312 (8p12, NRG1) as proxy for rs4236709

rs11780471 (8p21.1, CHRNA2) genotyped

rs1333040 (9p21.3, CDKN2A CDKN2BAS) genotyped

rs7045771 (9p21.3, MIR31HG MTAP) as proxy for rs7868253

rs7029287 (9p21.3, CDNK2A) as proxy for rs885518

rs62560775 (9p21.3, CDKN2B-AS1 CDKN2BAS) genotyped

rs11591710 (10q24.3, OBFC1) genotyped

rs11598840 (10q24.33, OBFC1) genotyped

rs10501831 (11q21, MTMR2) genotyped

rs1056562 (11q23.3, AMICA1 MPZL2) genotyped

rs10849605 (12p13.33, RAD52) as proxy for rs55791720

rs12816367 (12p13.33, RAD52) as proxy for rs7953330

rs11571815 (13q13.1, BRCA2) as proxy for rs11571818

rs11571833 (13q13.1, BRCA2) as proxy for rs11571833

rs185038874 (13q13.1, PDS5B) as proxy for rs185038874

rs1051730 (15q25.1, CHRNA3) genotyped

rs34606419 (15q21.2, GALK2) as proxy for rs11632038

rs16969968 (15q25.1, CHRNA3) as proxy for rs146009840

rs951266 (15q25.1, CHRNA5) as proxy for rs17486195

rs690367 (15q15.3, TP53BP1) as proxy for rs2602141

rs55781567 (15q25.1, CHRNA5) as proxy for rs55781567

rs55781567 (15q25.1, CHRNA5) as proxy for rs55853698

rs1378214 (15q21.1, SEMA6D) as proxy for rs66759488

rs2036527 (15q25.1, CHRNA5 PSMA4) as proxy for rs72740955

rs2413932 (15q21.1, SECISBP2L) as proxy for rs77468143

rs2413932 (15q21.1, SECISBP2L) as proxy for rs79040073

rs8034191 (15q25.1, HYKK) as proxy for rs8034191

rs8040868 (15q25.1, CHRNA3 SCAPER) genotyped

rs17483929 (15q25.1, IREB2 RP11650L1211) genotyped

rs56117933 (15q25.1, AGPHD1 CHRNA3 CHRNA5 and others) genotyped

rs2036527 (15q25.1, AGPHD1 CHRNA3 CHRNA5 and others) genotyped

rs12914385 (15q25.1, CHRNA3) genotyped

rs6495309 (15q25.1, CHRNA3 CHRNB4 RP11335K521) genotyped

rs10851907 (15q25.1, CHRNA3 CHRNB4 RP11335K521) genotyped

rs77719127 (15q25.1, ADAMTS7 MORF4L1) genotyped

rs17244648 (15q25.1, CTSH MORF4L1) genotyped

rs2644898 (19q13.2, CYP2A6 RAB4B-EGLN2) as proxy for rs11083569

rs11879413 (19q13.2, CYP2A6 RAB4B-EGLN2) genotyped

rs1709084 (19q13.2, CYP2A13 CYP2F1) genotyped

rs11670760 (19q13.2, CYP2A6) as proxy for rs56113850

rs11670760 (19q13.2, CYP2A6 RAB4B-EGLN2) as proxy for rs60446182

rs3761121 (20q13.33, RTEL1) as proxy for rs41309931

rs111105 (22q13.1, KCNJ4) as proxy for rs138396

rs17879961 (22q12.1, CHEK2 CHEK2-I157T) genotyped

### HLA-Region

The major histocompatibility complex (MHC: HLA-Region) corresponds to the genomic coordinates of 29677984 (GABBR I) to 33485635 (KIFC1) in the human genome build 36.3 of the National Center for Biotechnology Information (NCBI) map viewer [65]. We retrieved the list all genes within this region from the ENSEMBL database [66].

### Immuno-related genes

A list of 4667 immune-related genes is given by Immunology Database and Analysis Portal (ImmPort) [11].

### Genotypes of markers of GARD related genes

**gene cytoBand position from to number of SNPs**

KIAA0319L 1p34.3 35628789 36154217 34

GSTM1 1p13.3 110210671 110260742 27

RAP1A 1p13.2 111968121 112329551 45

GBA 1q22 155094013 155263186 42

GORAB 1q24.2 170456064 170564545 8

NCF2 1q25.3 183492205 183563224 34

SLC26A9 1q32.1 205857654 205914885 39

IL10 1q32.1 206924536 207109104 28

TGFB2 1q41 218458058 218690577 56

WDR35 2p24.1 20081474 20205541 53

IFT172 2p23.3 27502055 27844601 22

SRD5A2 2p23.1 31658174 31873839 32

REL 2p16.1 60920795 61276263 34

RETSAT 2p11.2 85537573 85595863 35

SFTPB 2p11.2 85861325 85897864 29

IL1B 2q13 113514159 113606560 20

TTC21B 2q24.3 166598478 166927659 38

LRP2 2q31.1 169936727 170267627 62

COL3A1 2q32.2 189526890 189935843 48

STAT4 2q32.3 191854874 192060780 26

BMPR2 2q33.1 203184161 203488449 16

CTLA4 2q33.2 204648661 204818924 17

CPS1 2q34 211186894 211653300 62

SLC11A1 2q35 218867708 219295198 73

TGFBR2 3p24.1 30586611 30746464 98

CASR 3q21.1 121796768 122009175 26

MYLK 3q21.1 123283519 123690404 38

IFT122 3q21.3 129151667 129269961 26

IFT80 3q25.33 159819567 160155861 20

TERC 3q26.2 169461571 169490183 21

CCDC39 3q26.33 180242703 180993424 71

HES1 3q29 193801606 193908761 23

TCTEX1D2 3q29 195955472 196068626 12

DOK7 4p16.3 3447156 3634526 27

EVC2 4p16.2 5505090 5713275 37

EVC EVC2 4p16.2 5698814 5717026 5

EVC 4p16.2 5710924 5866819 43

SLC34A2 4p15.2 25615723 25729339 19

WDR19 4p14 38987825 39301572 28

FAM13A 4q22.1 89532682 90062555 70

NFKB1 4q24 103374154 103557077 19

NEK1 4q33 170234781 170609251 25

SLC9A3 5p15.33 423385 551579 40

DNAH5 5p15.2 13619926 13964121 70

NIPBL 5p13.2 36759602 37159877 18

COMMD10 5q23.1 115401830 115791669 58

LOX 5q23.2 121260543 121489122 19

PPIC 5q23.2 122072632 122406325 29

CSF2 5q31.1 131107692 131546280 43

IL12B 5q33.3 158538588 158858898 40

DSP 6p24.3 7497110 7611118 21

CDKN1A 6p21.2 36613812 36657116 65

TSPYL1 6q22.1 116560766 116673487 13

RSPH4A 6q22.1 116893146 117054307 65

CTGF 6q23.2 132212773 132298141 14

CCR6 6q27 167313144 167591956 44

DNAH11 7p15.3 21484209 21947138 132

IL6 7p15.3 22654968 22831954 38

ELN 7q11.23 73230808 73524207 29

CAV1 7q31.2 116055528 116217657 19

CFTR 7q31.2 116928371 117358025 28

IRF5 7q32.1 128548463 128709789 18

WDR60 7q36.3 158525086 158782275 31

SFTPC 8p21.3 21964561 22030969 11

SPAG1 8q22.2 100904277 101304017 36

RAD21 8q24.11 117850148 117914100 21

LRRC6 8q24.22 133549399 133707408 23

IL33 9p24.1 6185360 6282511 58

DNAI1 9p13.3 34333855 34535000 16

TGFBR1 9q22.33 101813506 101998431 19

MUSK 9q31.3 113371069 113566357 42

GLE1 9q34.11 131198146 131306567 2

GLE1 WDR34 9q34.11 131219357 131453854 11

WDR34 9q34.11 131393940 131555366 6

TTF1 9q34.13 135169525 135312351 31

TSC1 9q34.13 135658616 135850574 33

ARMC4 10p12.1 27967566 28300818 41

PRKG1 10q11.23 52723140 54190663 219

SFTPA2 10q22.3 81298307 81322717 16

SFTPA1 SFTPA2 10q22.3 81322840 81343103 4

SFTPA1 10q22.3 81345246 81627553 12

ANXA11 10q22.3 81776483 81971019 23

ACTA2 10q23.31 90662086 90806623 24

TCTN3 10q24.1 97317158 97685352 31

SMC3 10q25.2 112249114 112392162 14

MUC2 11p15.5 915396 1106419 31

MUC2 MUC5AC 11p15.5 1111164 1111474 2

MUC5AC MUC5B 11p15.5 1112806 1245500 13

MUC5B TOLLIP 11p15.5 1246941 1313242 14

TOLLIP 11p15.5 1293601 1369465 5

FAM111B 11q12.1 58838704 58944104 114

RELA 11q13.1 65399528 65551417 27

EFEMP2 11q13.1 65571524 65785149 130

GSTP1 11q13.2 67345320 67356131 19

DYNC2H1 11q22.3 102926689 103437292 49

MFAP5 12p13.31 8759007 8832203 16

MARS 12q13.3 57853335 57951283 19

RAP1B 12q15 68914239 69208578 32

BBS10 12q21.2 76704479 76856220 10

HMGB1 13q12.3 30894735 31231872 32

TNFSF11 13q14.11 43054094 43227195 20

LACC1 13q14.11 44327751 44489247 12

MIR17HG 13q31.3 91971967 92050246 21

NKX21 14q13.3 36514392 37020175 23

MDGA2 14q21.3 47241493 48170086 114

DNAAF2 14q21.3 49953480 50211578 26

IFT43 14q24.3 76348241 76573554 24

FBLN5 14q32.12 92220869 92439059 35

PPP4R4 14q32.12 94563265 94757200 24

SERPINA1 14q32.13 94766726 94892438 33

JAG2 14q32.33 105582014 105695645 19

EIF2AK4 15q15.1 40113107 40360741 47

SMAD3 15q22.33 67322068 67647218 54

STRA6 15q24.1 74413911 74556797 25

IFT140 16p13.3 1546443 1705488 30

TSC2 16p13.3 2006099 2154235 41

ABCA3 16p13.3 2263836 2406044 21

PARN 16p13.12 14405428 14975292 29

MYH11 16p13.11 15676644 16041576 42

NOD2 16q12.1 50633443 50769181 20

DNAAF1 16q23.3 84122442 84215455 23

CYBA 16q24.3 88638854 88727733 22

FLCN 17p11.2 17075318 17147787 22

CSF3 17q21.1 38068621 38211121 35

CCDC103 17q21.31 42693848 43002225 37

MAPT 17q21.31 43463493 44866570 1340

GOSR2 17q21.32 44958937 45195850 24

ITGA3 17q21.33 48103935 48217569 19

COL1A1 17q21.33 48221920 48309874 14

DNAI2 17q25.1 72246466 72343155 19

CCDC40 17q25.3 77959459 78154360 69

PRTN3 19p13.3 789890 850175 3

ELANE PRTN3 19p13.3 819997 856145 9

ELANE 19p13.3 849014 867872 7

GNA11 19p13.3 3054567 3133853 18

DPP9 19p13.3 4654757 4783024 24

C3 19p13.3 6643942 6732982 30

CCDC151 19p13.2 11433169 11596297 17

FAM98C 19q13.2 38877923 38901728 11

TGFB1 19q13.2 41734666 41903861 52

RELB 19q13.32 45447178 45594595 25

CCDC114 19q13.33 48764695 48866386 16

TTYH1 19q13.42 54905127 54961672 48

SLC2A10 20q13.12 45269556 45378226 13

RTEL1 20q13.33 62268333 62330416 66

ARFRP1 RTEL1 20q13.33 62310963 62369895 71

ARFRP1 20q13.33 62327996 62369997 1

RSPH1 21q22.3 43795697 43938203 32

CSF2RB 22q12.3 37228277 37347959 32

OFD1 Xp22.2 13725261 13864977 15

FIGF Xp22.2 15256031 15472910 24

CYBB Xp11.4 37178504 37722202 11

KDM6A Xp11.3 44643615 45051111 29

HDAC8 Xq13.1 71521188 71881237 21

ATP7A Xq21.1 76399729 77497469 48

SLC6A14 Xq23 115480782 115668150 17

GPC4 Xq26.2 132194187 132551518 18

GPC3 GPC4 Xq26.2 132494646 132862190 12

GPC3 Xq26.2 132667773 133226706 17

 *total: 6499 Marker*

### Genotypes of markers of the gene family HGNC739 (Serpin peptidase inhibitors) (including SERPINA1 and PPP4R4):

all markers of chromosome 1 between position 11776915 (rs74657824) and 11828364 (rs77347935)

all markers of chromosome 1 between position 173837516 (chr1\_173837516\_A\_C) and 174040594 (rs16846667)

all markers of chromosome 1 between position 230808005 (rs10864766) and 230878561 (rs2493151)

all markers of chromosome 2 between position 224825779 (chr2\_224825779\_C\_T) and 224945517 (rs16865540)

all markers of chromosome 3 between position 167121000 (rs1403646) and 167668104 (rs9290330)

all markers of chromosome 6 between position 2799060 (chr6\_2799060\_A\_C) and 3002323 (chr6\_3002323\_C\_G)

all markers of chromosome 7 between position 100731026 (exm645170) and 100858178 (chr7\_100858178\_C\_T)

all markers of chromosome 11 between position 57232349 (chr11\_57232349\_A\_G) and 57447561 (rs1783819)

all markers of chromosome 11 between position 75255123 (rs12788428) and 75297884 (rs588955)

all markers of chromosome 11 between position 118959846 (newrs79983883) and 119003845 (chr11\_119003845\_A\_G)

all markers of chromosome 13 between position 51867290 (rs2153512) and 52008238 (chr13\_52008238\_C\_T)

all markers of chromosome 14 between position 94581722 (rs74536787) and 95129190 (rs7151768)

all markers of chromosome 17 between position 1576369 (rs11078563) and 1692140 (rs8066706)

all markers of chromosome 18 between position 61055369 (rs6567338) and 61751058 (rs2849308)

all markers of chromosome 22 between position 21111628 (chr22\_21111628\_C\_T) and 21307058 (rs2285547)

all markers of chromosome 23 between position 105153172 (exm1651328) and 105451287 (exm1651515)

These are in total 581 genotyped markers.
Access to imputed markers of the same regions could potentially be useful.

### Genotypes of markers of the gene family HGNC807 (ATP binding cassette subfamily C (ABCC)

all markers of chromosome 3 between position 183637722 (rs3805114) and 183733784 (rs4148557) : ABCC5

all markers of chromosome 3 between position 183730116 (rs1879259) and 183730116 (rs1879259) : ABCC5AS1

all markers of chromosome 6 between position 43396856 (rs4714684) and 43418733 (rs1214749) : ABCC10

all markers of chromosome 7 between position 117111985 (chr7\_117111985\_) and 117286896 (chr7\_117286896\_) : CFTR

all markers of chromosome 10 between position 101542578 (kgp4895132) and 101611277 (rs8187709) : ABCC2

all markers of chromosome 11 between position 17414570 (rs8192690) and 17496516 (kgp2580890) : ABCC8

all markers of chromosome 12 between position 21950433 (kgp27630655) and 22094815 (rs10770872) : ABCC9

all markers of chromosome 13 between position 95672105 (kgp5542930) and 95953517 (kgp97387) : ABCC4

all markers of chromosome 16 between position 16053729 (rs4148333) and 16236783 (kgp16378240) : ABCC1

all markers of chromosome 16 between position 16243542 (kgp25657567) and 16301530 (kgp16314222) : ABCC6

all markers of chromosome 16 between position 18599980 (chr16\_18599980\_) and 18607586 (kgp7834585) : ABCC6P1

all markers of chromosome 16 between position 48117180 (kgp16269839) and 48189363 (chr16\_48189363\_) : ABCC12

all markers of chromosome 16 between position 48200841 (kgp4328820) and 48266831 (rs11865596) : ABCC11

all markers of chromosome 17 between position 48723585 (rs12051822) and 48768515 (rs11656685) : ABCC3

all markers of chromosome 21 between position 15613985 (rs2822502) and 15729775 (rs2822623) : ABCC13

These are in total 499 genotyped markers.
Access to imputed markers of the same regions could potentially be useful.

References

1. (2018) Genetic and Rare Diseases Information Center (GARD) - lung diseases. Gaithersburg, MD.

2. (2018) ORPHANET

3. Mungall, C.J.*, et al.* (2017) The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res*, **45**, D712-D722.

4. Amberger, J.S.*, et al.* (2017) Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. *Curr Protoc Bioinformatics*, **58**, 1 2 1-1 2 12.

5. Topic, A.S.*, et al.* (2006) Association of moderate alpha-1 antitrypsin deficiency with lung cancer in the Serbian population. *Arch Med Res*, **37**, 866-70.

6. Torres-Duran, M.*, et al.* (2015) Alpha-1 Antitrypsin Deficiency and Lung Cancer Risk: A Case-Control Study in Never-Smokers. *J Thorac Oncol*, **10**, 1279-1284.

7. Yamada, A.*, et al.* (2018) Gastrointestinal cancer risk in cystic fibrosis: more exploration is needed - Authors' reply. *Lancet Oncol*, **19**, e333.

8. Maisonneuve, P.*, et al.* (2013) Cancer risk in cystic fibrosis: a 20-year nationwide study from the United States. *J Natl Cancer Inst*, **105**, 122-9.

9. Schoni, M.H.*, et al.* (1996) Cancer risk in patients with cystic fibrosis: the European data. CF/CSG Group. *J R Soc Med*, **89 Suppl 27**, 38-43.

10. Harris, J.M.*, et al.* (2010) Cryptogenic fibrosing alveolitis and lung cancer: the BTS study. *Thorax*, **65**, 70-6.

11. ImmPort.

12. Stram, D.O. (2017) Multi-SNP Haplotype Analysis Methods for Association Analysis. *Methods Mol Biol*, **1666**, 485-504.

13. Ferreiro-Iglesias, A.*, et al.* (2018) Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. *Nature Communications*, **9**, 3927.

14. Thun, G.A.*, et al.* (2013) Causal and synthetic associations of variants in the SERPINA gene cluster with alpha1-antitrypsin serum levels. *PLoS Genet*, **9**, e1003585.

15. VanderWeele, T.J. (2016) Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health*, **37**, 17-32.

16. Biedermann, A.*, et al.* (2006) Alpha-1-Antitrypsin-Mangel – eine versteckte Ursache der COPD: Überblick über Pathogenese, Diagnostik, Klinik und Therapie. *Dtsch Arztebl International*, **103**, 1828-.

17. Yang, P.*, et al.* (2005) Alpha1-antitrypsin and neutrophil elastase imbalance and lung cancer risk. *Chest*, **128**, 445-52.

18. Bozinovski, S.*, et al.* (2016) COPD and squamous cell lung cancer: aberrant inflammation and immunity is the common link. *Br J Pharmacol*, **173**, 635-48.

19. Yang, P.*, et al.* (2008) Alpha1-antitrypsin deficiency carriers, tobacco smoke, chronic obstructive pulmonary disease, and lung cancer risk. *Arch Intern Med*, **168**, 1097-103.

20. Aboussouan, L.S.*, et al.* (2009) Detection of alpha-1 antitrypsin deficiency: a review. *Respir Med*, **103**, 335-41.

21. Turner, A.M. (2013) Fifty years on: GWAS confirms the role of a rare variant in lung disease. *PLoS Genet*, **9**, e1003768.

22. Topic, A.*, et al.* (2012) Alpha-1-antitrypsin deficiency in Serbian adults with lung diseases. *Genet Test Mol Biomarkers*, **16**, 1282-6.

23. Obeidat, M.*, et al.* (2011) A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. *PLoS One*, **6**, e19382.

24. Busch, R.*, et al.* (2017) Genetic Association and Risk Scores in a Chronic Obstructive Pulmonary Disease Meta-analysis of 16,707 Subjects. *Am J Respir Cell Mol Biol*, **57**, 35-46.

25. Topic, A.*, et al.* (2011) Alpha-1-antitrypsin phenotypes and neutrophil elastase gene promoter polymorphisms in lung cancer. *Pathol Oncol Res*, **17**, 75-80.

26. Enewold, L.*, et al.* (2012) SERPINA1 and ELA2 polymorphisms are not associated with COPD or lung cancer. *Anticancer Res*, **32**, 3923-8.

27. Schwartz, A.G.*, et al.* (1998) Alpha-1-antitrypsin carrier status and lung cancer risk among nonsmokers. *Am J Epidemiol.*, **147**.

28. Tanash, H.A.*, et al.* (2016) Cause-specific mortality in individuals with severe alpha 1-antitrypsin deficiency in comparison with the general population in Sweden. *Int J Chron Obstruct Pulmon Dis*, **11**, 1663-9.

29. El-Akawi, Z.J.*, et al.* (2006) Relationship between alpha-1 antitrypsin deficient genotypes S and Z and lung cancer in Jordanian lung cancer patients. *Saudi Med J*, **27**, 181-4.

30. Yang, P.*, et al.* (1999) Alpha1-antitrypsin deficiency allele carriers among lung cancer patients. *Cancer Epidemiol Biomarkers Prev*, **8**, 461-5.

31. Brenner, D.R.*, et al.* (2011) Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. *PLoS One*, **6**, e17479.

32. Oldham, J.M.*, et al.* (2017) Comorbid Conditions in Idiopathic Pulmonary Fibrosis: Recognition and Management. *Front Med (Lausanne)*, **4**, 123.

33. Le Jeune, I.*, et al.* (2007) The incidence of cancer in patients with idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Respir Med*, **101**, 2534-40.

34. Ozawa, Y.*, et al.* (2009) Cumulative incidence of and predictive factors for lung cancer in IPF. *Respirology*, **14**, 723-8.

35. Paolocci, G.*, et al.* (2018) Occupational risk factors for idiopathic pulmonary fibrosis in Southern Europe: a case-control study. *BMC Pulm Med*, **18**, 75.

36. Baumgartner, K.B.*, et al.* (2000) Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. *Am J Epidemiol*, **152**, 307-15.

37. Balestro, E.*, et al.* (2016) Immune Inflammation and Disease Progression in Idiopathic Pulmonary Fibrosis. *PLoS One*, **11**, e0154516.

38. Ungprasert, P.*, et al.* (2016) Epidemiology of Sarcoidosis 1946-2013: A Population-Based Study. *Mayo Clin Proc*, **91**, 183-8.

39. Reich, J.M. (2016) Epidemiology of Sarcoidosis. *Mayo Clin Proc*, **91**, 822-3.

40. Newman, L.S.*, et al.* (2004) A case control etiologic study of sarcoidosis: environmental and occupational risk factors. *Am J Respir Crit Care Med*, **170**, 1324-30.

41. Douglas, J.G.*, et al.* (1986) Sarcoidosis: a disorder commoner in non-smokers? *Thorax*, **41**, 787-91.

42. Hance, A.J.*, et al.* (1986) Smoking and interstitial lung disease. The effect of cigarette smoking on the incidence of pulmonary histiocytosis X and sarcoidosis. *Ann N Y Acad Sci*, **465**, 643-56.

43. Valeyre, D.*, et al.* (1988) Smoking and pulmonary sarcoidosis: effect of cigarette smoking on prevalence, clinical manifestations, alveolitis, and evolution of the disease. *Thorax*, **43**, 516-24.

44. Ungprasert, P.*, et al.* (2016) Smoking, obesity and risk of sarcoidosis: A population-based nested case-control study. *Respir Med*, **120**, 87-90.

45. Bonifazi, M.*, et al.* (2015) Sarcoidosis and cancer risk: systematic review and meta-analysis of observational studies. *Chest*, **147**, 778-791.

46. Ji, J.*, et al.* (2009) Cancer risk in hospitalized sarcoidosis patients: a follow-up study in Sweden. *Ann Oncol*, **20**, 1121-6.

47. Boffetta, P.*, et al.* (2009) A cohort study of cancer among sarcoidosis patients. *Int J Cancer*, **124**, 2697-700.

48. (2004) *The Molecular Genetic Epidemiology of Cystic Fibrosis: Report of a Joint Meeting of WHO/ECFTN/ICF(M)A/ECFS* World Health Organization.

49. Sanders, D.B.*, et al.* (2012) Comparing age of cystic fibrosis diagnosis and treatment initiation after newborn screening with two common strategies. *J Cyst Fibros*, **11**, 150-3.

50. Southern, K.W. (2007) Cystic fibrosis and formes frustes of CFTR-related disease. *Respiration*, **74**, 241-51.

51. Bobadilla, J.L.*, et al.* (2002) Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. *Hum Mutat*, **19**, 575-606.

52. Castellani, C.*, et al.* (2008) Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *J Cyst Fibros*, **7**, 179-96.

53. Lao, O.*, et al.* (2003) Spatial patterns of cystic fibrosis mutation spectra in European populations. *Eur J Hum Genet*, **11**, 385-94.

54. McKay, J.D.*, et al.* (2017) Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*, **49**, 1126-1132.

55. Timofeeva, M.N.*, et al.* (2012) Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet*, **21**, 4980-95.

56. Timofeeva, M.N.*, et al.* (2009) CYP450 polymorphisms as risk factors for early-onset lung cancer: gender-specific differences. *Carcinogenesis*, **30**, 1161-9.

57. Rosenberger, A.*, et al.* (2017) Gene-set meta-analysis of lung cancer identifies pathway related to systemic lupus erythematosus. *PLOS ONE*, **12**, e0173339.

58. Brenner, D.R.*, et al.* (2013) Hierarchical modeling identifies novel lung cancer susceptibility variants in inflammation pathways among 10,140 cases and 11,012 controls. *Hum Genet*, **132**, 579-89.

59. Brenner, D.R.*, et al.* (2015) Identification of lung cancer histology-specific variants applying Bayesian framework variant prioritization approaches within the TRICL and ILCCO consortia. *Carcinogenesis*, **36**, 1314-26.

60. Ji, X.*, et al.* (2018) Identification of susceptibility pathways for the role of chromosome 15q25.1 in modifying lung cancer risk. *Nature Communications*, **9**, 3221.

61. Truong, T.*, et al.* (2010) Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J Natl.Cancer Inst.*, **102**, 959-971.

62. Truong, T.*, et al.* (2010) International Lung Cancer Consortium: coordinated association study of 10 potential lung cancer susceptibility variants. *Carcinogenesis*, **31**, 625-633.

63. Wang, Y.*, et al.* (2014) Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet*, **46**, 736-41.

64. Feng, Y.*, et al.* (2018) Novel genetic variants in the P38MAPK pathway gene ZAK and susceptibility to lung cancer. *Mol Carcinog*, **57**, 216-224.

65. Shiina, T.*, et al.* (2009) The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet*, **54**, 15-39.

66. Cunningham, F.*, et al.* (2015) Ensembl 2015. *Nucleic Acids Res*, **43**, D662-9.