Identification of genetic variants in the TNF promoter associated with COPD secondary to tobacco smoking and its severity

Abstract: Chronic obstructive pulmonary disease (COPD) is an inflammatory disease that arises in response to noxious particles or gases. Associations of genetic polymorphisms in TNF have been reported in Asians and Caucasians, but not in Mestizo populations. A case-control study was conducted in two stages: in the first stage, patients with COPD (COPD group, n=165) and smokers without disease (SNC group, n=165) were included and the TNF promoter sequence was determined using direct sequencing. In the second stage, the identified polymorphisms were validated by real-time polymerase chain reaction (PCR) in COPD (n=260) and SNC (n=506). In the first stage, 11 different sets of “contig” alignments were determined, of which contig 10 was found to be associated with susceptibility ($P=5.0E-04$, OR [odds ratio] =3.64) and contig 1 with Global Initiative for COPD (GOLD) greater grade ($P=1.0E-02$, OR =3.82). The single nucleotide polymorphisms found in this region were individually identified; the GA genotypes of rs1800629 ($P=0.038$, OR =2.07), rs56036015 ($P=0.0082$, OR =3.18), and rs361525 ($P=1.0E-02$, OR =4.22) were higher in the COPD group vs the SNC group; after second-stage validation, rs1800629 ($P=6.0E-03$, OR =2.26) and rs56036015 ($P=1.0E-03$, OR =2.54) are maintained. There are genetic variants in the TNF promoter associated with increased risk of COPD secondary to smoking and with a higher GOLD grade in the Mexican Mestizo population.

Keywords: lung, cigarette smoking, SNP, GOLD, Mexican population

Introduction

Chronic obstructive pulmonary disease (COPD) is defined by the Global Initiative for COPD (GOLD) as a common, preventable, and treatable disease characterized by persistent limitation of airflow, which is usually progressive and associated with an increased chronic inflammatory reaction in response to noxious particles or gases in the lung.1 Cigarette smoking is the main risk factor for developing COPD; however, only 10%–20% of smokers develop the disease, suggesting that other factors influence its development. Some studies estimate that smoking accounts for 15% of the decline in lung function, while genetic factors account for up to 55%.2,2 The worldwide prevalence of COPD varies between 5% and 10%, and in Mexico City, it is 7.8%.4 COPD is largely related to inflammation, both systemic and localized,5 and therefore, various studies have focused on measuring the levels of inflammatory biomarkers and their relation to the clinical severity of the disease.6 Some of the markers that have been tested are interleukin 1-beta (IL-1β), IL-6, IL-8, and tumor necrosis factor-alpha (TNF-α). TNF-α is a critical molecule in inflammation regulation because it induces a cascade of other inflammatory cytokines, chemokines, and growth factors.7 In addition, murine models have found increased levels of mRNA of TNF-α and some...
metalloproteinases in alveolar walls and airways of mice exhibiting overexpression of TNF-α transgene in SP-C producing cells.8

Genetic variations in the TNF promoter region have been associated with differences in phenotypic expression and with a wide range of infectious, oncological, and autoimmune diseases.9–11 Certain studies of the genetic association with COPD reveal some relevance of single nucleotide polymorphisms (SNPs) to TNF in Asian populations, but not in Caucasian populations.12–14 Therefore, the aim of the present investigation was to determine the set of polymorphisms found in the TNF promoter region, its possible associations with COPD secondary to smoking, and the clinical severity in the Mexican Mestizo population.

Material and methods

Study population

A case-control study was performed in two stages; 1,096 participants were included; all were evaluated via lung function tests considering a ratio of forced expiratory volume in 1 second/forced vital capacity (FEV1/FVC) <70% as COPD; individuals older than 50 years of age and with a smoking rate of ≥5 packs/year were included. They underwent a background questionnaire of inherited pathologies, whereby subjects who reported suffering some type of lung and/or chronic inflammatory disease were excluded, as were subjects with non-Mexican ancestry. All participants were recruited from COPD and smoking-cessation support clinics, both belong to Department of Research in Smoking and COPD of Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas (INER) at México City. The subjects were first invited to participate, were asked to sign a letter of informed consent, and then were given a document assuring the privacy of personal data, both evaluated and approved by the Institutional Committee on Science and Bioethics in Research (protocol #B10-12). The STREGA (STrengthening the REporting of Genetic Association) guidelines were taken into consideration in the design of this genetic association study.15

Stages

In the first stage, 165 patients with a diagnosis of COPD secondary to smoking (COPD group) and 165 smokers without COPD (SNC group) were included. Direct sequencing of the promoter region of the TNF gene was performed. In the second stage, validation of the association of the polymorphisms identified in the first step was performed using allelic discrimination via real-time polymerase chain reaction (RT-PCR); 763 smokers were included, of whom 260 were in the COPD group and 506 in the SNC group.

DNA extraction

The DNA was extracted from peripheral blood cells via venipuncture, using the commercial BDtract Genomic DNA isolation kit (Maxim Biotech, San Francisco, CA, USA). The DNA was then quantified by UV absorption spectrophotometry at the 260 nm wavelength using a NanoDrop system (Thermo Scientific, Wilmington, DE, USA).

Selection and design of the regions to be sequenced

The selected regions of interest were chosen based on previous reports from the available literature, which identifies the regulatory regions in which different transcription factors are found.16–18 The region of interest consisted of approximately 300 bp, and the sequence that encompassed this region was selected from the National Center for Biotechnology Information (NCBI) database. Using NCBI’s Primer-BLAST tool, the primers were designed with GC% <65%, Tm (melting temperature) <60°C, and a primer size of <25 bp. The resultant size of the sequenced region was approximately 350 bp.

Bidirectional sequencing

Sequencing was performed using the commercial BigDye Direct Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA), for which the DNA concentration was adjusted to 4 ng/μL, and the integrity of the DNA was verified by the electrophoretic shift in a 2% gel. PCR amplification was performed under the following amplification program: 95°C for 10 min for 1 cycle; 96°C for 3 s, 62°C for 15 s, and 68°C for 30 s for 35 cycles; and 72°C for 2 min and 4°C until the next step. The sequencing reaction was performed with the following cycles: 37°C for 15 min, 80°C for 2 min, and 96°C for 1 min for 1 cycle; 96°C for 10 s, 50°C for 5 s, and 60°C for 75 s for 25 cycles, and 4°C until purification. The product obtained was purified via ethanol–EDTA precipitation. The capillary injection was performed in an ABI 3130 Genetic Analyzer (Applied Biosystems) using the DT3130POP7{BD} v2.mob mobility file, in the RapidSeq36_POP7_1 analysis module, injecting 1–1.5 kV for 5–10 s, collecting data for 1,800 s, and reviewing them with the Sequencing Analysis Software v5.4 (Applied Biosystems).

Sequence alignment

To identify the sets of alignment, we used the “contig assembly” algorithm in the Sequencher 5.1 (Gene Codes,
Ann Arbor, MI, USA) software, using “assemble automatically” with a “minimum match percentage” of 80 and a “minimum overlap” of 100 bp. The reference sequence was imported from the NCBI database with data obtained from the Ensembl database (Reference Sequence: NC_000006.12), the full sequence is included in Supplementary material, the set of samples that shared a consensus sequence were defined as “contig”. Polymorphisms and their position within the sequence were determined, and all were assigned taking into account the overlap of the peaks in the electropherograms.

Genotyping of the SNPs
The allelic discrimination of SNPs was performed using the commercial TaqMan probes (Applied Biosystems) at a concentration of 20×, using the technique of RT-PCR in a 7300 Real Time PCR System kit (Applied Biosystems), and the analysis was performed by the SDS (sequence detection software) version 1.4 software (Applied Biosystems).

Statistical analysis
The differences between the two groups under study were evaluated by determining and comparing the allele, genotype, and haplotype frequencies. Statistical significance was assessed using the SPSS v15.0 (SPSS Inc., Chicago, IL, USA) and Epi Info 7.1.4.0 (Centers for Disease Control and Prevention, Atlanta, GA, USA) statistical software, taking into account the overlap of the peaks in the electropherograms. Through comparative alignment, it was possible to determine 11 different alignment assemblies (contigs), and a comparison of the frequencies of these assemblies between the study groups was performed; contig 10 was increased in the COPD group relative to the SNC group, obtaining a statistically significant difference (P=0.0005, OR =3.64, 95% CI =2.03–6.71) (Table 2). The sequences of the different identified contigs are found in Table S1.

Table 1 Demographic and lung function variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>First stage</th>
<th>Second stage</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COPD (n=165)</td>
<td>SNC (n=165)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>112 (67.87)</td>
<td>107 (64.84)</td>
<td></td>
</tr>
<tr>
<td>Female (%)</td>
<td>53 (32.12)</td>
<td>58 (35.15)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>63.96±2.92</td>
<td>60.07±4.63</td>
<td></td>
</tr>
<tr>
<td>Tobacco index (packyears)</td>
<td>55.16±4.8</td>
<td>51.87±1.7</td>
<td></td>
</tr>
<tr>
<td>GOLD grades (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>27 (16.36)</td>
<td>44 (16.92)</td>
<td>NS</td>
</tr>
<tr>
<td>II</td>
<td>74 (44.84)</td>
<td>104 (41.15)</td>
<td>NS</td>
</tr>
<tr>
<td>III</td>
<td>47 (28.48)</td>
<td>71 (27.30)</td>
<td>NS</td>
</tr>
<tr>
<td>IV</td>
<td>17 (10.30)</td>
<td>38 (14.61)</td>
<td>NS</td>
</tr>
<tr>
<td>Measures of lung function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV₁ref (%)</td>
<td>58.70 (14–129)</td>
<td>97.54 (58–140)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FVC ref (%)</td>
<td>85.06 (31–160)</td>
<td>97.53 (61–140)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV₁/FVC (%)</td>
<td>51.82 (22–70)</td>
<td>79.88 (71–91)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Notes: *Mean ± SD; †Postbronchodilator values; ‡minimum and maximum values are shown.
Abbreviations: COPD, chronic obstructive pulmonary disease; SNC, smokers without COPD; NS, Not significant; GOLD, Global Initiative for COPD; FEV₁, forced expiratory volume in 1 second; FVC, forced vital capacity; SD, standard deviation.

Results
The clinical characteristics of the subjects included in both stages are presented in Table 1. The sex ratio (female: male) for the participants in the first stage was 1:2.1 in the patient group and 1:1.8 in the SNC group; in the second stage, the ratio in the COPD group was 1:2.3 and in the SNC group was 1:2.0. In both stages, these differences were not statistically significant. The average age in the COPD groups of both stages was slightly higher than in the SNC group (around 3 years, P=not significant [NS]). The smoking rate was higher in the patient group compared with the smokers without COPD. For both groups, the largest number of subjects was found in GOLD grades II and III (~70%), with GOLD II being the most frequent (first stage =44.84%, second stage =41.15%, P=NS). Differences in the measures of lung function were found due to the specific characteristics of the case and control selection.

First stage analysis
Sets of alignment (contigs)
Through comparative alignment, it was possible to determine 11 different alignment assemblies (contigs), and a comparison of the frequencies of these assemblies between the study groups was performed; contig 10 was increased in the COPD group relative to the SNC group, obtaining a statistically significant difference (P=0.0005, OR =3.64, 95% CI =2.03–6.71) (Table 2). The sequences of the different identified contigs are found in Table S1.
Table 2  Frequencies of the TNF alignment assemblies

<table>
<thead>
<tr>
<th>TNF contig</th>
<th>COPD</th>
<th>SNC</th>
<th>P-value</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CF (%)</td>
<td>CF (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>34.55</td>
<td>12.65</td>
<td>5.0E-04</td>
<td>3.64</td>
<td>2.03–6.71</td>
</tr>
<tr>
<td>2</td>
<td>13.33</td>
<td>13.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12.73</td>
<td>12.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12.37</td>
<td>12.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10.91</td>
<td>10.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>&lt;10</td>
<td>43</td>
<td></td>
<td></td>
<td>&lt;10</td>
</tr>
</tbody>
</table>

Note: Only those contigs that had a frequency >10% are shown.
Abbreviations: COPD, chronic obstructive pulmonary disease; SNC, smokers without COPD; CF, “contig” frequency; OR, odds ratio; 95% CI, 95% confidence intervals.

Stratification by GOLD grade

Stratification of the patient group was conducted by GOLD grade, grouping the grades of lesser severity (GI and GII) under the name G1 and the grades of greater severity (GIII and GIV) as G2. When comparing G2 vs G1, a statistically significant difference in contig 1 was obtained, being increased in the greater severity group (P = 0.010, OR = 3.56, 95% CI = 1.30–10.33). When comparing the G1 and G2 groups with the SNC group, the associated assembly was 10 (P = 0.000017, OR = 3.82, 95% CI = 1.99–7.43; P = 0.0042, OR = 2.92, 95% CI = 1.35–6.24, respectively; Table 3).

Determination of SNPs by sequencing

Six SNPs were identified in the studied region, which included the region from base −350 to −1 of the TNF promoter. In the analysis by genotype, an association with genotype GA was found for three of the six identified SNPs: rs1800629 (P = 0.038, OR = 2.07, 95% CI = 1.04–4.25; pB = 0.228), rs56036015 (P = 0.0082, OR = 3.18, 95% CI = 1.32–8.46; pB = 0.049), and rs361525 (P = 0.001, OR = 4.220, 95% CI = 1.71–11.84; pB = 0.006), showing increased frequency in COPD compared to the SNC group. In addition, a haplotype analysis was conducted in which all identified polymorphisms were included, none was found to be associated with the disease. The P-value for Hardy–Weinberg equilibrium (HWE) for each SNP analyzed in sequencing stage is: rs1800629 HWE = 0.442, rs56036015 HWE = 0.749, rs4987027 HWE = 0.906, rs673 HWE = 0.906, rs361525 HWE = 0.0038, rs41297589 HWE = 0.458.

Second stage analysis

Determination of the SNPs by allelic discrimination

Six SNPs, which had previously been identified by sequencing, were genotyped by RT-PCR; the results of the frequency of the genotypes were compared using the χ² test and are presented in Table 4.

The GA genotype of rs1800629 was increased in the group of patients with COPD (10.4%) compared with smokers without COPD (4.8%). Also, the GA genotype of rs56036015 had a higher frequency in the COPD group (11.39%) compared with the SNC group (4.8%); in both cases, the differences were statistically significant and were associated with increased risk (OR > 2.0). The frequencies of the genotypes for the six SNPs in both stages are presented in Table 4. The characteristics of the identified SNPs are presented in Table 5.

Discussion

COPD is a lung disease of multifactorial origin, and different immunological pathways involved in its pathogenesis have
been described;\textsuperscript{19} different altered biomarkers have also been reported, particularly those related to inflammation. Among these is TNF-\(\alpha\), which plays a mediator role in the immune response to gram-positive bacteria, amplifies inflammation via the regulation of different genes, and is related to the systemic effects of COPD. Several studies have demonstrated a relationship between metabolic disorders and high levels of TNF-\(\alpha\);\textsuperscript{20} it is associated with accelerated metabolism and with chronic debilitating diseases, resulting in loss of weight, musculoskeletal proteins, and adipose tissue.\textsuperscript{21,22} Increased TNF-\(\alpha\) production leads to a raise in the inflammatory and prooxidative responses, and it is also thought to play an important role in respiratory and systemic disease characteristics.\textsuperscript{23} These levels have been found to be increased during infective exacerbations,\textsuperscript{24} and it has also been associated with systemic manifestations such as decreased body mass index\textsuperscript{25} and abnormal energy expenditure during rest.\textsuperscript{26}

This cytokine is encoded by the TNF gene located on the short arm of chromosome 6 (6p21.3); multiple polymorphisms throughout the gene have been described; in this regard, various studies have evaluated the contribution of SNP-type polymorphisms in the promoter region, given their involvement in transcriptional regulation. In the present study, alignment assemblies (contigs) in the TNF promoter region associated with susceptibility and with GOLD severity in COPD were characterized; in addition, polymorphisms contained in these sets were individually analyzed; three SNPs associated with the establishment of the disease were identified, two of which retained the association after validation in a second population with the same clinical and demographic characteristics.

This present study is the first to use this methodology that reports alignment assemblies, which are polymorphisms that are shared by a group of individuals and are associated

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr P</th>
<th>Change</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1800629</td>
<td>31543031</td>
<td>G/A</td>
<td>-308</td>
</tr>
<tr>
<td>rs56036015</td>
<td>31543065</td>
<td>G/A</td>
<td>-274</td>
</tr>
<tr>
<td>rs4987027</td>
<td>31543094</td>
<td>G/A</td>
<td>-245</td>
</tr>
<tr>
<td>rs673</td>
<td>31543095</td>
<td>G/C</td>
<td>-244</td>
</tr>
<tr>
<td>rs361525</td>
<td>31543101</td>
<td>G/A</td>
<td>-238</td>
</tr>
<tr>
<td>rs41297589</td>
<td>31543262</td>
<td>T/A</td>
<td>-77</td>
</tr>
</tbody>
</table>

Abbreviations: SNP, single nucleotide polymorphisms; Chr P, chromosome position; NP, nucleotide position.
with COPD. Prior work has been performed in which haplotypes associated with the disease have been found; in 2008, Gingo et al. conducted a study in Caucasians of CO, USA, analyzing five SNPs in the TNF promoter and finding that allele A at the −308 position and A at the −237 position are associated with increased risk, whereas allele T at −857 is associated with decreased risk when comparing the less severe grades vs the more severe ones. Interestingly, the researchers also found the haplotype CC/TT (−1031/−863) with a decreased OR when comparing patients against smokers without COPD. In contrast, in a Greek population, a microarray was analyzed with five SNPs in the TNF promoter in patients with COPD secondary to smoking and bronchiectasis; however, the researchers found no association with individual polymorphisms, whereas with the haplotype TCCGG (−1013C/A, −863C/A, −857C/T, −308G/A, −238G/A), a discrete association was observed when compared with a population group included in the study. In a Caucasian population in the USA, an analysis of 10 tag SNPs in TNF and LTA was performed, and no individual association or association with a certain haplotype was found. Regarding the severity of the disease and its association with genetic variants, the findings are controversial; in a US Caucasian population, a study was conducted in which no TNF polymorphisms were found associated with the rate of decline in lung function. Consistent with this finding, in a Caucasian population in Turkey, no association was found between rs1800629 and the prognosis, as determined by the decrease in FEV\textsubscript{1} and body mass index in the monitoring of patients after 1 year; however, genotype GA showed an association with exacerbations and hospitalization. Interestingly, in Spanish Caucasians, allele A of TNF−863 was associated with an improved FEV\textsubscript{1} and a low-BODE (Body mass index, Obstruction, Dyspnea, Exercise capacity) index. Similar to these findings, in a study conducted in ten different centers in the US, considering ten SNPs in LTA and four in TNF, no association was found with clinical severity variables. In this paper, we found alignment assemblies associated with the GOLD grades of the greatest severity, which is the first report of its type in a Mexican Mestizo population. In our population, GOLD II and III had the highest prevalence, whereas the lower frequency grades were I and IV. In this report, we grouped the grades of lesser severity (GOLD I and II) and of greater severity (GOLD III and IV) in the two groups of included cases, with most subjects found in grades II and III. GOLD II was the most frequent, which is a behavior reported in our population.

Individually, three SNPs associated with susceptibility were identified, of which rs1800629 and rs361525 have already been studied in COPD, with inconclusive results. For SNP TNF−308G/A (rs1800629, a guanine to adenine [G → A] change), allele A (the minor allele) reportedly increases the rate of gene transcription\textsuperscript{10} and has been associated with multiple diseases. In a Taiwanese population, an association of the GA genotype with chronic bronchitis and with an increased inflammatory response was reported;\textsuperscript{11} the results of this work were included in a meta-analysis by Zhang et al.\textsuperscript{16} in which it was shown that the −308 GA genotype is associated with risk in Asian but not in Caucasian populations; this association is preserved when stratified by smoking. However, these results have not been replicated in other populations. In 2012, a study was conducted in a Taiwanese population, in which TNF−863 (rs1800630) associated with an increase in the FEV\textsubscript{1}/FVC ratio and body mass index was identified. Another polymorphism reported in COPD is rs361525 GA, which increases the protein expression in lung secretions and bioactivity with neutrophilic inflammation, increasing chemotaxis as well as being associated with clinical features of the disease, including its progression. This polymorphism did not retain the association during the validation of the findings in the second stage; our working group previously found a relationship between the GA genotype of this SNP and COPD secondary to biomass-burning smoke, an association that is not maintained when comparing both clinical entities of COPD (Reséndiz-Hernández et al unpublished data, 2015). Regarding SNP rs56036015, there are no association studies regarding COPD or other diseases; interestingly, in our investigation, this association was preserved in both stages.

Together these polymorphisms have not been described in any prior work; in our investigation, it was possible to report them, given that the methodology used enabled finding them using an alignment algorithm, which encompasses this group of polymorphic variants located in the region of transcription factor binding. The binding of these factors depends largely on the nucleotide sequences and on the variations of the same; consequently, the analysis of regulatory regions is important for elucidating the molecular mechanisms that are related to the pathogenesis of the disease.

Among the principal limitations of our study is the difficulty in differentiating the main clinical phenotypes of the disease (chronic bronchitis and emphysema), a fact that may bias the impact of the results, since they affected different biological pathways and molecules involved in these. Regarding our findings, we found that polymorphism rs361525
shows a deviation to HWE, this may be related to the low frequency of the minor allele homozygote (AA) and heterozygous (GA) in Mestizo populations, additionally, rs1800629 does not maintain a significant $P$-value after analysis by multiple testing ($pB = 0.228$); however, the rest of genotypes associated conserve the association value ($pB < 0.05$).

It would be worthwhile to perform a subsequent bioinformatic analysis and additional functional studies, which relates the identified sequences, particularly those associated with greater severity, with the potential transcription factors, and identifies their possible correlation with both local (ie, in the lung microenvironment) and systemic expression levels to conclusively determine whether the SNPs identified are associated with COPD development.

Acknowledgments
This paper constitutes a partial fulfillment of the Graduate Program in Biological Sciences of the National Autonomous University of México (UNAM). JM Reséndiz-Hernández wishes to acknowledge the scholarship (#233765) and financial support provided by the National Council of Science and Technology (CONACyT), and UNAM.

Author contribution
Conception and design, acquisition of data, and analysis and interpretation of data: JMRH, RHS, ARV, and RFV. Drafting the article or revising it critically for important intellectual content: JMRH, RHS, RJHZ, GVA, LCB, MVU, AC, ARV, and RFV; final approval of the version to be published: JMRH, RHS, RJHZ, GVA, LCB, MVU, AC, ARV, and RFV. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: JMRH, RHS, RJHZ, GVA, LCB, MVU, AC, ARV, and RFV. All authors contributed toward data analysis, drafting and revising the paper.

Disclosure
The authors report no conflicts of interest in this work.

References
2. Lee PN, Fry JS. Systematic review of the evidence relating FEV1 decline to giving up smoking. BMC Med. 2010;8:84.


### Table S1 Sequence of contigs identified

<table>
<thead>
<tr>
<th>Contig</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| 1      | GAAATGGAGCGCAATAGGTTTTGAGGGGCATGAGACGGGTTCAGGCTTACACACAATCATCGTCTGGGAGGAAAGACCCCTGTGAATCGAGCGAGGAGGAGTGGGATGTAGGGATGCGGGTACTGCTGTGTTCTGCTGCTCAGCCGCTTACACCGAATCGAGGCAAGACCCCAAGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 2      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 3      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 4      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 5      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 6      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 7      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 8      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 9      | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG
| 10     | GAAATGGAGGCAATAGGTTTTGAGGGCATGGGACCCGGTTCAGGCTTACACACAATAATCATCGTCTGGGAGGAAAGACCCCGCCCTCTGGAGATCGAGGCTCTCTCCTCCAGATGAGCTCTTCTGCTTACACAGAGGAGTTTCCCGCTGGTGAATACTTCCACGACGAGCCACTCGCTCTCCAGAGCTGCTGTTTGAGGAGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAGGAGATGAGGGATGGTTATCCAGATGTTCTGGCCACCAAGGACATATAAGGGAGCTTTGGCGACACCAAGCAGCAGCAAGCACAGAG

(Continued)
### Table S1 (Continued)

<table>
<thead>
<tr>
<th>Contig</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>GAAATGGAGGCAATAGGTGTgTGAAGGGGATTGACGGGGTTTCAGCTCAGGGTCTACACACAAATCAGTCAGTGCCAGAAGACCCCTCCACTACATCGGAGCAG GAGATGGGGGTGAGGGATCTCTGGATGCTTGTGCTCCCCTACTTTCCAATCCCCGCCCCCGCATGGAAGAAACCGAGACAGAAGTGCAAGGCCCCACTACGCTTCCTC CAGATGACGTCATGGGTTCCTCCACACAGGAAGTTTCCCGCTGATGATTCTTTCCGCTCCCTCCTGGCCCAAGGACATATAAGGCAAGTTGTGGGCACACCCAGG CCAACGAGACG</td>
</tr>
</tbody>
</table>

**Notes:** Reference Sequence: NC_000006.12 => gi|224589818:31543000-31543350 Homo sapiens chromosome 6, GRCh37.p5 Primary Assembly, NCBI Reference Sequence: NC_000006.12 GAAATGGAGGCAATAGTTTTGAGG GCAATGGGCAGGGTTCCAGCCTCAGGGTCTACACACAAATCAGTCAGTGCCAGAAGACCCCTCCACTACATCGGAGCAG GAGATGGGGGTGAGGGATCTCTGGATGCTTGTGCTCCCCTACTTTCCAATCCCCGCCCCCGCATGGAAGAAACCGAGACAGAAGTGCAAGGCCCCACTACGCTTCCTC CAGATGACGTCATGGGTTCCTCCACACAGGAAGTTTCCCGCTGATGATTCTTTCCGCTCCCTCCTGGCCCAAGGACATATAAGGCAAGTTGTGGGCACACCCAGG CCAACGAGACG.