Identification of FGF7 as a novel susceptibility locus for chronic obstructive pulmonary disease

John M. Brehm,[1-3,5,7] Koichi Hagiwara,[9] Y Tesfaigzi Y,[15] S Bruse,[15] Thomas J. Mariani,[10] Soumyaroop Bhattacharya,[10] Nadia Boutaoui,[1] John P. Ziniti,[2] Manuel E. Soto-Quiros,[8] Lydiana Avila,[8] Michael H. Cho, [2,3,5] Blanca Himes,[2] Augusto A. Litonjua, [2,3,5,7] Francine Jacobson,[6] Per Bakke,[12] Amund Gulsvik,[12] Wayne H Anderson,[13] David A. Lomas,[14] Erick Forno, [11] Soma Datta,[2] Edwin K. Silverman,[2-5] and Juan C. Celedón[1,2-5,7]

[1]Division of Pediatric Pulmonary Medicine, Allergy, and Immunology, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA; [2]Channing Laboratory, [3]Division of Pulmonary/Critical Care Medicine, [4]Center for Genomic Medicine, [5]Department of Medicine, and [6]Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts; [7]Department of Medicine, Harvard Medical School, Boston, Massachusetts; [8]Division of Pediatric Pulmonology, Hospital Nacional de Niños, San José, Costa Rica; [9]Saitama Medical University Hospital and Institute, Saitama, Japan; [10]University of Rochester Medical Center, Rochester, NY, [11]Division of Pediatric Pulmonology, Dept. of Pediatrics, University of Miami, Miami, FL, [12]Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway; [13]GlaxoSmithKline Research and Development, Research Triangle Park, North Carolina, USA; [14]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK; [15]Lovelace Respiratory Research Institute, Albuquerque, NM

Supplementary Methods

Study Populations

Costa Rica: The primary study population consisted of 58 subjects with (cases) and 57 subjects without (controls) COPD in the Genetic Epidemiology of COPD in Costa Rica (GECCOR) study. Cases were recruited from patients attending four adult hospitals in San José (Costa Rica) and their affiliated clinics, and through newspaper ads. Control subjects were recruited from individuals attending a smoking-cessation clinic at the Institute for Pharmaco-dependency in San José, and through newspaper ads. All of these control subjects had no addictions other than to nicotine as determined by phone questionnaire. To ensure their descent from the founder population of the Central Valley of Costa Rica (which is predominantly of Spanish and Native American ancestry), all participants were required to have at least six great-grandparents born in the Central Valley of Costa Rica. Additional inclusion criteria for cases were age 21 to 71 years, physician-diagnosed COPD, ≥ 10 pack-years of cigarette smoking, and an FEV1 ≤ 65% predicted and an FEV1/FVC ratio of ≤ 70% after bronchodilator administration. Controls were recruited on the basis of the same criteria for age and smoking history, but they had to have no physician-diagnosed COPD and normal spirometry. Exclusion criteria for cases and controls included history of chronic pre-existing chronic lung disease (e.g., bronchiectasis). Pre-existing lung disease was determined by physician diagnosis, CT scans (41/58 cases), and subject questionnaire. Controls who had significant dyspnea, cough, or wheezing without a formal pulmonary diagnosis were also excluded.

Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-Points (ECLIPSE; SCO104960, NCT00292552): Details of the ECLIPSE study have been described previously[1] Briefly, ECLIPSE is a 3-year longitudinal study of COPD. Cases were diagnosed by post-bronchodilator spirometry[2] as GOLD Stage II or greater (forced expiratory volume in 1 second (FEV1) < 80% predicted and FEV1/ forced vital capacity (FVC) < 0.7) and without severe alpha-1 antitrypsin deficiency; smoking controls had normal lung function (post-bronchodilator FEV1 >85% predicted and FEV1/FVC > 0.7). Both cases

and controls were limited to subjects of self-reported white ethnicity, between ages 40-75 and were required to have at least a 10 pack-year smoking history. As the study was designed to evaluate COPD-related endpoints, recruitment was weighted towards cases; a total of 1839 cases and 196 smoking controls were genotyped.

National Emphysema Treatment Trial (NETT): Details of the National Emphysema Treatment Trial, a clinical trial to evaluate lung volume reduction surgery, have been published previously[3]. NETT subjects were enrolled at 17 clinical centers based on severe airflow obstruction by spirometry[2] (FEV1 ≤ 45% predicted[4]) and emphysema on computed tomographic (CT) imaging of the chest. The NETT Genetics Ancillary Study contains a subset of the original cohort with blood available for genotyping. After providing written informed consent, NETT participants provided a blood sample for DNA extraction for genetic studies of COPD. The study was approved by the institutional review boards at participating NETT centers. A total of 382 self-reported white subjects without severe alpha-1 antitrypsin deficiency were included in this study.

Normative Aging Study (NAS): The Normative Aging Study is an ongoing longitudinal study of healthy men established in 1963 and conducted by the Veterans Administration (VA) [5]. Briefly, men from the greater Boston area, ages 21 to 80 years, enrolled in the study after an initial health screening determined that they were free of known chronic medical conditions. Since enrollment, the participants have undergone comprehensive clinical examinations at 5-year intervals for those < 52 years old and at 3-year intervals for those > 52 years old. Selection of controls for COPD genetic studies from this population has been described previously[6]; self-reported white control subjects had DNA available for genotyping, a history of at least 10 pack-years of cigarette smoking and no evidence for airflow obstruction by spirometry[2] at their most recent visit (FEV1 > 80% predicted[7] and FEV1/FVC > 90% predicted[4]. Anonymized data were used, as approved by the institutional review boards of Partners Healthcare System and the Boston VA. A total of 453 subjects meeting enrollment criteria were

included in this study.

Norway: Details on the Bergen, Norway case-control study have been described previously[8]. Subjects were recruited from Bergen, Norway, and in contrast to the NAS and ECLIPSE studies, were required to have > 2.5 pack years of smoking history. GOLD Stage II or greater COPD cases were diagnosed by post-bronchodilator spirometry2 (FEV1 < 80% predicted and FEV1/FVC < 0.7), while controls had normal spirometry (FEV1 > 80% predicted and FEV1/FVC > 0.7). Subjects with alpha-1 antitrypsin genotypes PI ZZ, ZNull, Null-Null or SZ were excluded. All subjects gave informed consent, and the appropriate institutional review boards approved the study. A total of 933 cases and 919 controls of self-identified white ethnicity were included in this study.

Lovelace Smokers Cohort: Details on the Lovelace Smokers Cohort have been published previously.[9] Our top two SNPs in *FGF7* were replicated using a Taqman assay (Applied Biosystems, Foster City, CA). All subjects gave informed consent, and the institutional review board of the Lovelace Respiratory Institute approved the study. COPD was classified based on an FEV1/FVC ratio below the 5th percentile of the predicted value, also referred to as the lower limit of normal (LLN).[10]

Genotyping and Quality Control

Genotyping of NETT,NAS,Norway, and ECLIPSE has been described in detail previously [11].

For the Costa Rica cohort, genotyping was performed on the Illumina Quad 610 platform (Illumina, Inc; San Diego,CA) at the Channing Laboratory, Brigham and Women's Hospital. Cases and controls were randomly allocated in batches, which included at least one replicate sample. First pass cleaning using two-channel intensity was performed using Beadstudio. Although subjects with <95% of markers genotyped, and SNPs with a call frequency < 95% were to be removed a priori, all genotyped subjects passed with an average call rate of 99.87%. The remaining markers were cleaned following Illumina guidelines.

SNPs with a call frequency between 0.9 and 0.99 and a cluster separation below 0.3 were manually inspected, with a threshold set for each workspace below which no SNPs were deemed unambiguous. We additionally examined SNPs with a low ABR mean, ABT mean, and heterozygote excess. X, Y, and mitochondrial SNPs were excluded from analysis.

Subject and Marker Cleaning

For the Costa Rica cohort, BeadStudio workspaces were exported to ped file format and further quality control assessment was performed using Python (www.python.org) and R (www.r-project.org) scripts in conjunction with PLINK (1.05) [12]. Quality control of genotyping data was assessed by subject and by marker. Subject data were excluded after examining missingness, reproducibility and discordances, relatedness, sex, and inbreeding. Subjects with discordance > 1% in replicate genotyping were discarded. Relatedness was examined using rgGrr[13] and estimated IBD in PLINK, using a cutoff of 0.12513. Within each set of related subjects, unrelated individual(s) were chosen based on phenotypic criteria. In addition to testing for relatedness within each cohort, a test for relatedness between cohorts was performed prior to construction of the primary (merged) dataset. Sex assignment was based on X homozygosity estimates, with those > 0.8 as male, and < 0.2 as female. Discordant samples were removed. Inbreeding coefficients > |0.2| were removed. Finally, among samples run in replicate, the sample with the higher passing genotype rate was chosen.

Markers were excluded based on missingness, minor allele frequency, reproducibility, and Hardy-Weinberg equilibrium. Markers with missingness > 5% were excluded. A strict minor allele frequency cutoff was not specified; instead, markers that were monoallelic or singleton in each dataset were excluded. Within replicates, SNPs showing significant discordance (examining the distribution of discordances) were excluded. Markers with Hardy-Weinberg deviation – P value less than .01 – were also excluded. Differential missingness between cases and controls was assessed, but not used to exclude markers at this stage. A summary of the quality control measures is shown in Supplementary

Table 1.

Construction of RCHH

Regions of conserved homozygosity haplotype (RCHHs) were identified using the method described by Miyazawa et. al. [14]. In brief, for any given individual all heterozygous SNPs were ignored and the SNP location was scored with the value of the allele for that subject. Subjects are compared only across SNPs that are scored. RCHH are defined by runs of SNPs that share the same allele at the homozygous locations across multiple subjects, ignoring heterozygous SNPs. The size of the shared segments between any two individuals was set at 3.0 cM, which is a size designed to reduce the false positive and false negative rates of discovery. A theoretical ancestral segment is then constructed from the largest subgroup of subjects sharing a particular RCHH (see Supplementary Figure 1). While any two subjects much have at least 3.0 cM of sharing, the size may be much smaller when comparing across multiple subjects (Supplementary figure 2). The total number of cases and controls sharing this ancestral allele is used to calculate a P value based on a standard normal distribution.

Supplementary Acknowledgments

Supplementary References

1. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J* 2008;31(4):869-73.
2. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med* 1995;152(3):1107-36.
3. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348(21):2059-73.
4. Crapo RO, Morris AH. Standardized single breath normal values for carbon monoxide diffusing capacity. *Am Rev Respir Dis* 1981;123(2):185-9.
5. Bell B, Rose CL, Damon A. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* 1972;3:5-17.
6. Hersh CP, Demeo DL, Lange C, Litonjua AA, Reilly JJ, Kwiatkowski D, et al. Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am J Respir Cell Mol Biol* 2005;33(1):71-8.
7. O'Connor GT, Sparrow D, Weiss ST. A prospective longitudinal study of methacholine airway responsiveness as a predictor of pulmonary-function decline: the Normative Aging Study. *Am J Respir Crit Care Med* 1995;152(1):87-92.
8. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, Anderson WH, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am J Respir Crit Care Med* 2007;176(2):167-73.
9. Sood A, Stidley CA, Picchi MA, Celedon JC, Gilliland F, Crowell RE, et al. Difference in airflow obstruction between Hispanic and non-Hispanic White female smokers. *COPD* 2008;5(5):274-81.
10. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26(5):948-68.
11. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;42(3):200-2.
12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 2007;81(3):559-75.
13. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. GRR: graphical representation of relationship errors. *Bioinformatics* 2001;17(8):742-3.
14. Miyazawa H, Kato M, Awata T, Kohda M, Iwasa H, Koyama N, et al. Homozygosity Haplotype Allows a Genomewide Search for the Autosomal Segments Shared among Patients. *The American Journal of Human Genetics* 2007;80(6):1090-102.

Supplementary Table 1 – Genotyping quality control for Costa Rica

Illumina Platform	Quad 610
Genotyped Subjects	115
Missingness > 5%	0
Discordance	0
Relatedness	0
Gender mismatch	0
Passing subjects	115
Genotyped markers	619,372
Minor allele frequency = 0 / Missingness > 5%	40,640
Less than 2 subjects with minor allele	16,665
Non-Reproducible	0
HWE < 0.01	3,138
Passing markers	558,929

Supplementary Table 2 – Characteristics of cohorts

|	Costa Rica		NETT/NAS,ECLIPSE,Norway combined cohort	
	Cases	Controls	Cases	Controls
No. subjects	58	57	2940	1380
Sex (# male)	33	37	1903	910
Mean age (in years)	60	46	65	60
Mean pack-years	55	35	47	28
Mean FEV1 % predicted (in liters)	43	101	46	98
Mean FEV1/FVC	0.53	0.83	.45	.79

Supplementary Table 3 – Top 20 P values representing 100 RCHH

Location (chromosome and position in base pairs using hg18 coordinates)	Cases N (%)	Controls N (%)	P value	Size (in bp)
6:121788169-122413818	12 (21%)	0 (0%)	2.20E-04	6.3E+05
6:123119324-123274804	12 (21%)	0 (0%)	2.20E-04	1.6E+05
6:123668845-123668845	12 (21%)	0 (0%)	2.20E-04	0.0E+00
7:151892012-151903280	12 (21%)	0 (0%)	2.20E-04	1.1E+04
7:151904404-152295138	14 (24%)	1 (2%)	2.60E-04	3.9E+05
6:121660856-121783240	11 (19%)	0 (0%)	4.20E-04	1.2E+05
6:122414162-123118271	11 (19%)	0 (0%)	4.20E-04	7.0E+05
6:123277300-123664764	11 (19%)	0 (0%)	4.20E-04	3.9E+05
7:151823054-151887401	11 (19%)	0 (0%)	4.20E-04	6.4E+04
15:30302218-30713368	13 (22%)	1 (2%)	5.00E-04	4.1E+05
6:120847633-121657937	10 (17%)	0 (0%)	7.80E-04	8.1E+05
6:123669986-123911605	10 (17%)	0 (0%)	7.80E-04	2.4E+05
7:151524608-151820728	10 (17%)	0 (0%)	7.80E-04	3.0E+05
8:8469618-8476478	10 (17%)	0 (0%)	7.80E-04	6.9E+03
17:4832680-5666242	10 (17%)	0 (0%)	7.80E-04	8.3E+05
15:28124439-29225595	12 (21%)	1 (2%)	9.40E-04	1.1E+06
15:30300525-30301633	12 (21%)	1 (2%)	9.40E-04	1.1E+03
15:30714768-30721385	12 (21%)	1 (2%)	9.40E-04	6.6E+03
6:120356562-120847245	9 (16%)	0 (0%)	1.40E-03	4.9E+05
6:123915133-123935587	9 (16%)	0 (0%)	1.40E-03	2.0E+04
8:8482301-8491941	9 (16%)	0 (0%)	1.40E-03	9.6E+03
15:30748430-30764393	9 (16%)	0 (0%)	1.40E-03	1.6E+04
17:4746831-4830503	9 (16%)	0 (0%)	1.40E-03	8.4E+04
17:5666418-5673154	9 (16%)	0 (0%)	1.40E-03	6.7E+03
20:7112725-7318428	9 (16%)	0 (0%)	1.40E-03	2.1E+05
7:152299155-152322133	11 (19%)	1 (2%)	1.70E-03	2.3E+04
15:28115352-28116500	11 (19%)	1 (2%)	1.70E-03	1.1E+03
15:30723787-30746003	11 (19%)	1 (2%)	1.70E-03	2.2E+04
10:2951101-2969537	16 (28%)	4 (7%)	2.30E-03	1.8E+04
1:54240754-54358003	8 (14%)	0 (0%)	2.70E-03	1.2E+05
6:120203694-120354567	8 (14%)	0 (0%)	2.70E-03	1.5E+05
6:123938996-123954610	8 (14%)	0 (0%)	2.70E-03	1.6E+04
7:151325371-151523278	8 (14%)	0 (0%)	2.70E-03	2.0E+05
8:8493973-8538640	8 (14%)	0 (0%)	2.70E-03	4.5E+04
13:37187973-37868825	8 (14%)	0 (0%)	2.70E-03	6.8E+05
15:30766905-30782048	8 (14%)	0 (0%)	2.70E-03	1.5E+04
20:6808895-7112218	8 (14%)	0 (0%)	2.70E-03	3.0E+05
20:7319819-7499917	8 (14%)	0 (0%)	2.70E-03	1.8E+05
15:29227096-30300468	12 (21%)	2 (4%)	3.10E-03	1.1E+06
7:152326396-152436044	10 (17%)	1 (2%)	3.20E-03	1.1E+05
8:8450177-8467882	10 (17%)	1 (2%)	3.20E-03	1.8E+04
15:28091067-28113480	10 (17%)	1 (2%)	3.20E-03	2.2E+04
21:16580023-16650095	10 (17%)	1 (2%)	3.20E-03	7.0E+04
10:2969826-2976450	15 (26%)	4 (7%)	4.00E-03	6.6E+03

8:8405647-8411849	13 (22%)	3 (5%)	4.90E-03	6.2E+03
1:54359237-54379392	7 (12%)	0 (0%)	4.90E-03	2.0E+04
3:195215725-195237176	7 (12%)	0 (0%)	4.90E-03	2.1E+04
6:20108027-21710531	7 (12%)	0 (0%)	4.90E-03	1.6E+06
6:120164641-120203646	7 (12%)	0 (0%)	4.90E-03	3.9E+04
6:158414763-159470841	7 (12%)	0 (0%)	4.90E-03	1.1E+06
6:166716423-166938711	7 (12%)	0 (0%)	4.90E-03	2.2E+05
7:106220834-109629284	7 (12%)	0 (0%)	4.90E-03	3.4E+06
7:130062836-130088545	7 (12%)	0 (0%)	4.90E-03	2.6E+04
7:151325334-151325334	7 (12%)	0 (0%)	4.90E-03	0.0E+00
8:8538732-8577480	7 (12%)	0 (0%)	4.90E-03	3.9E+04
8:141202813-142094332	7 (12%)	0 (0%)	4.90E-03	8.9E+05
11:58979902-61009776	7 (12%)	0 (0%)	4.90E-03	2.0E+06
11:125802819-126097478	7 (12%)	0 (0%)	4.90E-03	2.9E+05
11:126163084-126355974	7 (12%)	0 (0%)	4.90E-03	1.9E+05
13:21298032-21305888	7 (12%)	0 (0%)	4.90E-03	7.9E+03
13:36237281-36283266	7 (12%)	0 (0%)	4.90E-03	4.6E+04
13:37140481-37185893	7 (12%)	0 (0%)	4.90E-03	4.5E+04
13:37873623-37973865	7 (12%)	0 (0%)	4.90E-03	1.0E+05
14:20026284-20062727	7 (12%)	0 (0%)	4.90E-03	3.6E+04
14:20117472-20117472	7 (12%)	0 (0%)	4.90E-03	0.0E+00
15:30782135-30810778	7 (12%)	0 (0%)	4.90E-03	2.9E+04
15:45975580-49182801	7 (12%)	0 (0%)	4.90E-03	3.2E+06
15:56434430-56450814	7 (12%)	0 (0%)	4.90E-03	1.6E+04
17:1738924-1758733	7 (12%)	0 (0%)	4.90E-03	2.0E+04
17:4636312-4742067	7 (12%)	0 (0%)	4.90E-03	1.1E+05
18:6634394-6680637	7 (12%)	0 (0%)	4.90E-03	4.6E+04
18:6830148-6841557	7 (12%)	0 (0%)	4.90E-03	1.1E+04
18:70392209-70397061	7 (12%)	0 (0%)	4.90E-03	4.9E+03
18:70645532-70974303	7 (12%)	0 (0%)	4.90E-03	3.3E+05
18:72019892-72356083	7 (12%)	0 (0%)	4.90E-03	3.4E+05
19:56106962-56106962	7 (12%)	0 (0%)	4.90E-03	0.0E+00
20:7499995-7572747	7 (12%)	0 (0%)	4.90E-03	7.3E+04
10:2947286-2950317	16 (28%)	5 (9%)	5.40E-03	3.0E+03
7:152436388-152475034	9 (16%)	1 (2%)	5.80E-03	3.9E+04
13:21279294-21279294	9 (16%)	1 (2%)	5.80E-03	0.0E+00
15:28085029-28088527	9 (16%)	1 (2%)	5.80E-03	3.5E+03
17:5674078-5758986	9 (16%)	1 (2%)	5.80E-03	8.5E+04
10:2977064-2978202	14 (24%)	4 (7%)	6.90E-03	1.1E+03
8:8412246-8427477	12 (21%)	3 (5%)	8.50E-03	1.5E+04
10:2719235-2728665	12 (21%)	3 (5%)	8.50E-03	9.4E+03
13:21234739-21239466	12 (21%)	3 (5%)	8.50E-03	4.7E+03
21:46252267-46265743	12 (21%)	3 (5%)	8.50E-03	1.3E+04
1:54379745-54491077	6 (10%)	0 (0%)	9.00E-03	1.1E+05
1:145750325-149855359	6 (10%)	0 (0%)	9.00E-03	4.1E+06
1:163312158-163357909	6 (10%)	0 (0%)	9.00E-03	4.6E+04
2:80955775-83248796	6 (10%)	0 (0%)	9.00E-03	2.3E+06
2:221397282-221749310	6 (10%)	0 (0%)	9.00E-03	3.5E+05
3:195179804-195213553	6 (10%)	0 (0%)	9.00E-03	3.4E+04

4:6644018-6644018	6 (10%)	0 (0%)	9.00E-03	0.0E+00
4:6818237-7317757	6 (10%)	0 (0%)	9.00E-03	5.0E+05
4:7381577-7389520	6 (10%)	0 (0%)	9.00E-03	7.9E+03
4:7531322-7990094	6 (10%)	0 (0%)	9.00E-03	4.6E+05
6:19963912-20104787	6 (10%)	0 (0%)	9.00E-03	1.4E+05
6:21710724-23614469	6 (10%)	0 (0%)	9.00E-03	1.9E+06
6:51100102-53152338	6 (10%)	0 (0%)	9.00E-03	2.1E+06

Supplementary figure 1 – The homozygosity haplotype of a theoretical founder individual is reconstructed using the algorithm shown here. While the three genotyped subjects themselves are not homozygous the region shown here, by comparing across all subjects the haplotype of the founder can be reconstructed.
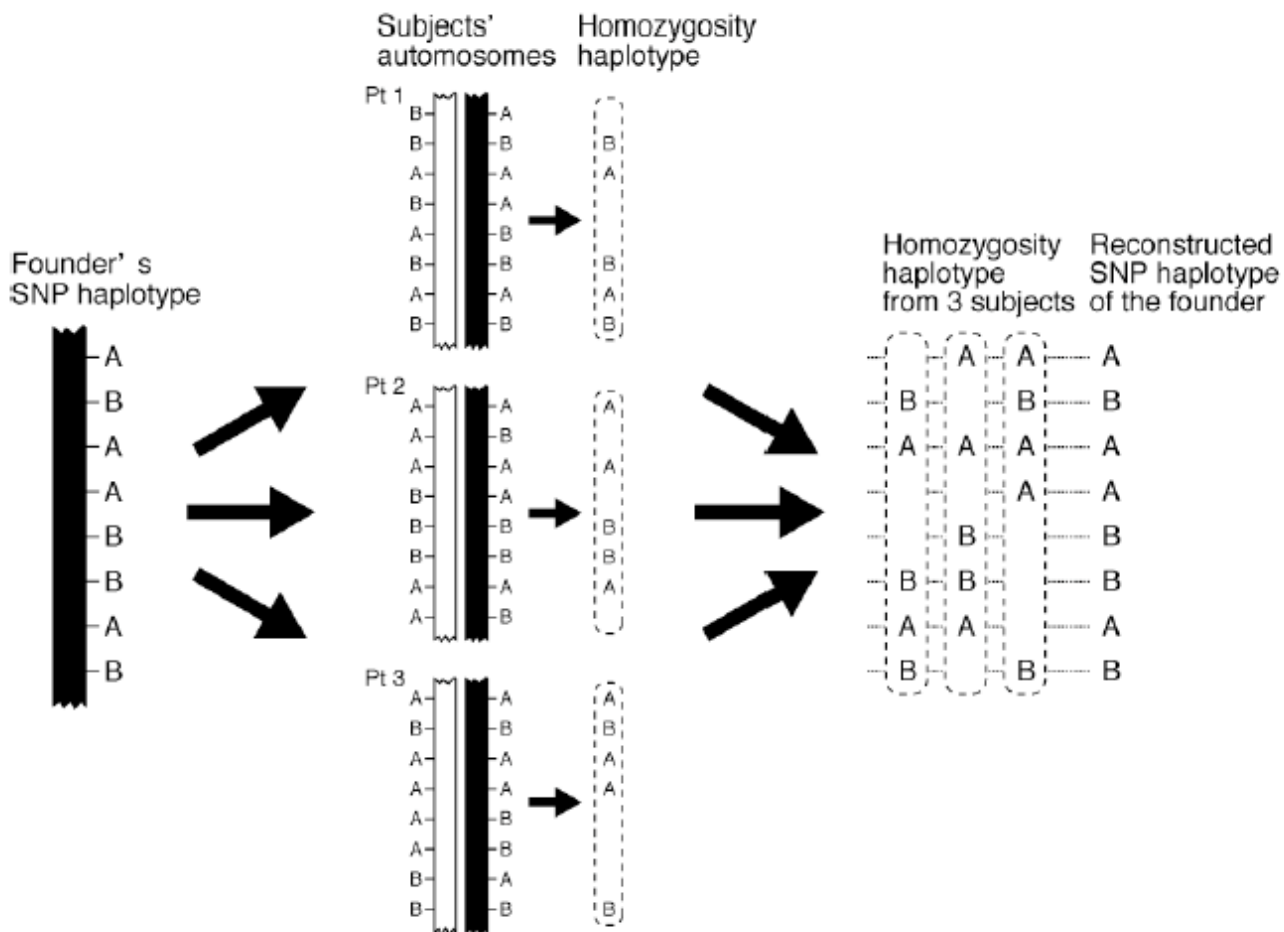
Supplementary figure 2 – The genetic length of a shared segment between any two individuals must be at least 3.0 cM, but the total length of the final RCHH may be much smaller due to differences in overlap. In this figure, the region derived from a common ancestor is shaded grey. To be identified by the algorithm, any two subjects must have an overlap of common ancestry that is at least 3.0 cM in length. However, the final reconstructed Region from a Common Ancestor using all subjects (indicated as the region between the two dashed lines) may be much smaller than 3.0 cM.
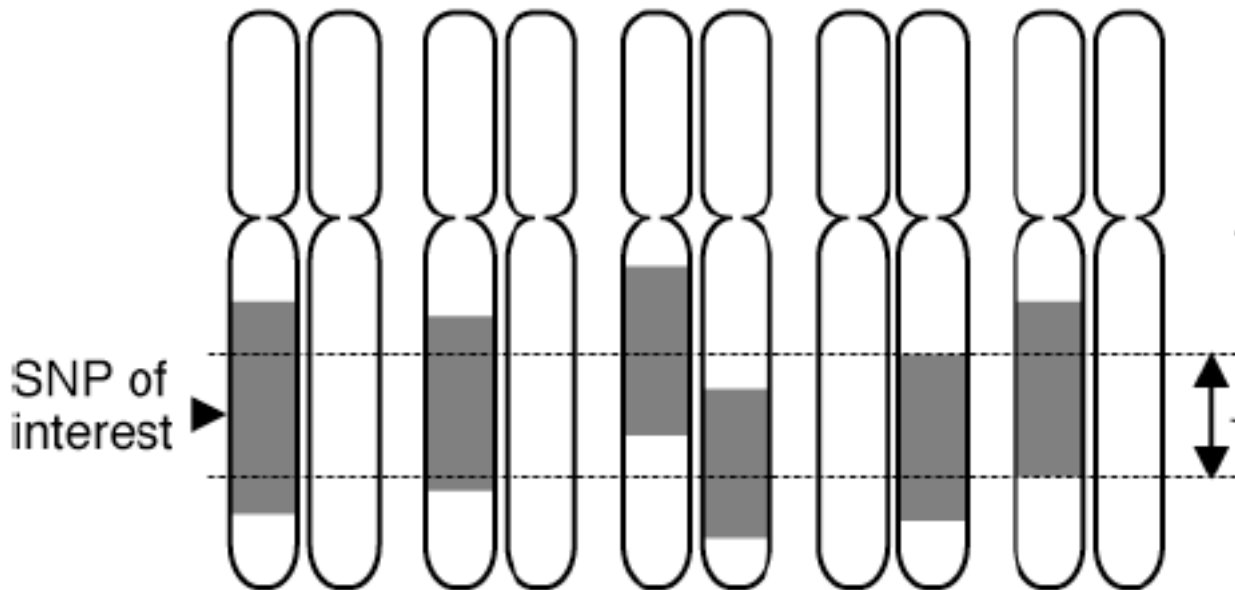
Supplementary figure 3 – In the case where more than one region from a common ancestor is identified, the region with the largest number of subjects in the subgroup is used. In the figure below, 5 subjects inherited the same region on at least one chromosome (dark grey), and two subjects inherited a different region (light grey). The HHAnalysis algorithm chooses the dark grey region as the ancestral segment.

A group (number of subjects = N)

SNP of interest ▶

SNP of interest ◀

a subgroup of M subjects (the largest subgroup)

a subgroup

Each pair of the subjects share an RCHH that contains SNP of interest

The subjects share an RCHH

▦ Regions derived from a common ancestor

▢ Regions derived from a different common ancestor