

Ensemble genomic analysis in human lung tissue
identifies novel genes for chronic obstructive pulmonary disease

Jarrett D. Morrow, Michael H. Cho, John Platig, Xiaobo Zhou, Dawn L. DeMeo, Weiliang Qiu, Bartholome Celli, Nathaniel Marchetti, Gerard J. Criner, Raphael Bueno, George R. Washko, Kimberly Glass, John Quackenbush, Edwin K. Silverman, Craig P. Hersh

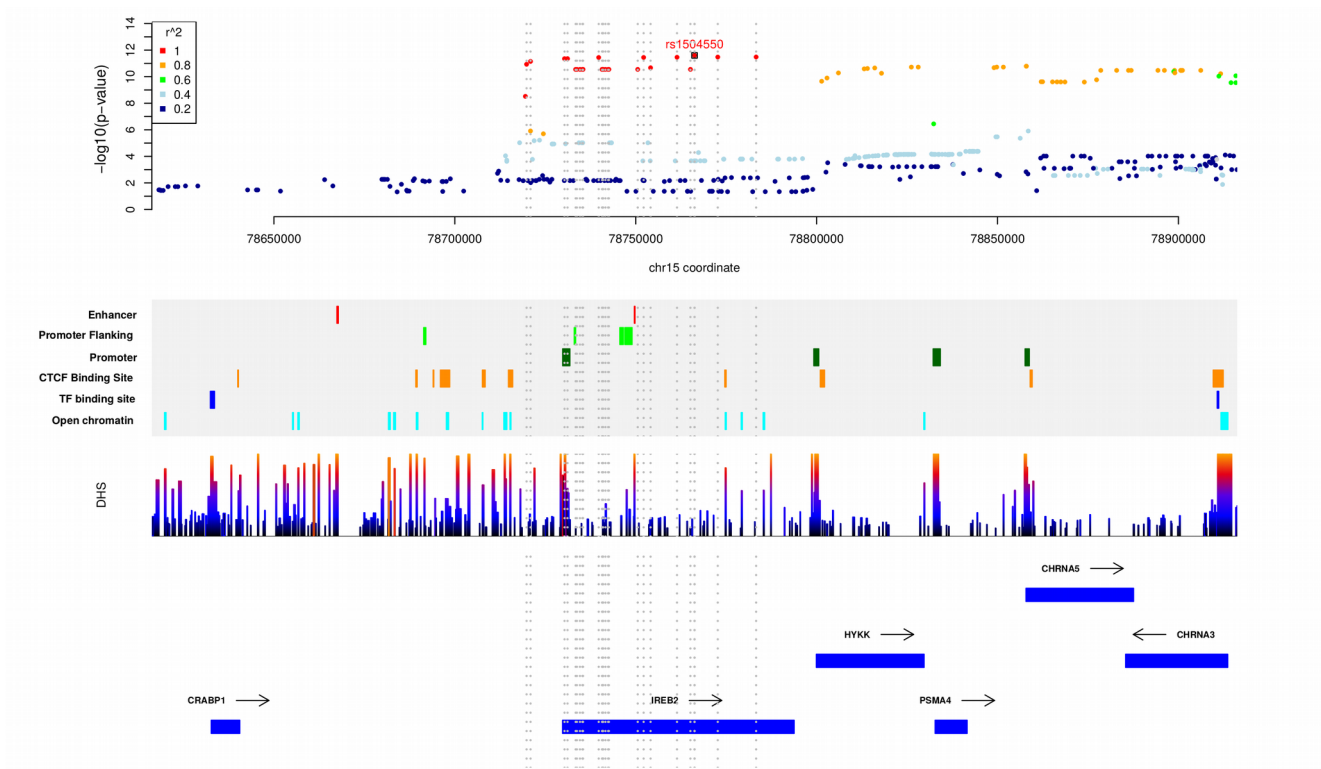


Figure S1. Regional genomic plot for rs1504550:*IREB2* locus with regulatory features and gene information; red and blue marker colors indicates higher and lower linkage disequilibrium (LD), respectively.

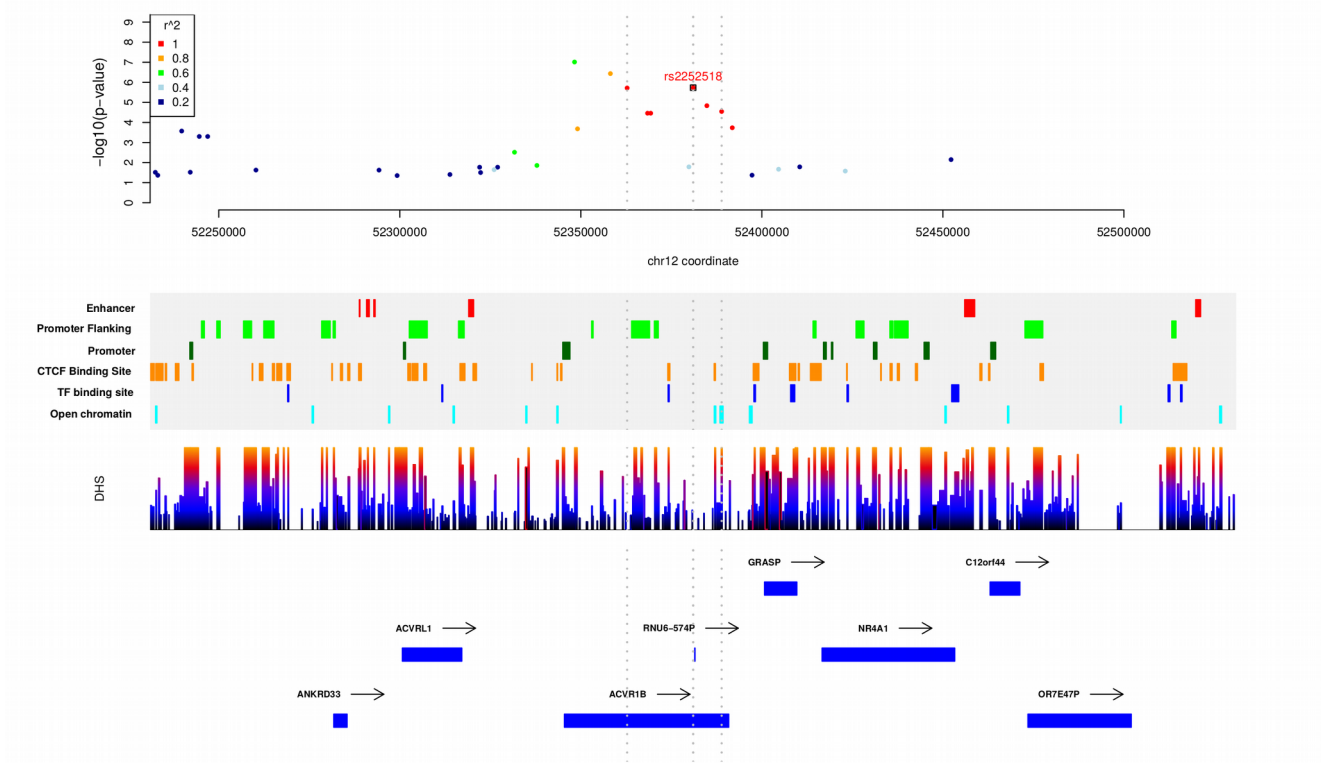


Figure S2. Regional genomic plot for rs2252518:*ACVR1B* locus with regulatory features and gene information; red and blue marker colors indicates higher and lower linkage disequilibrium (LD), respectively.

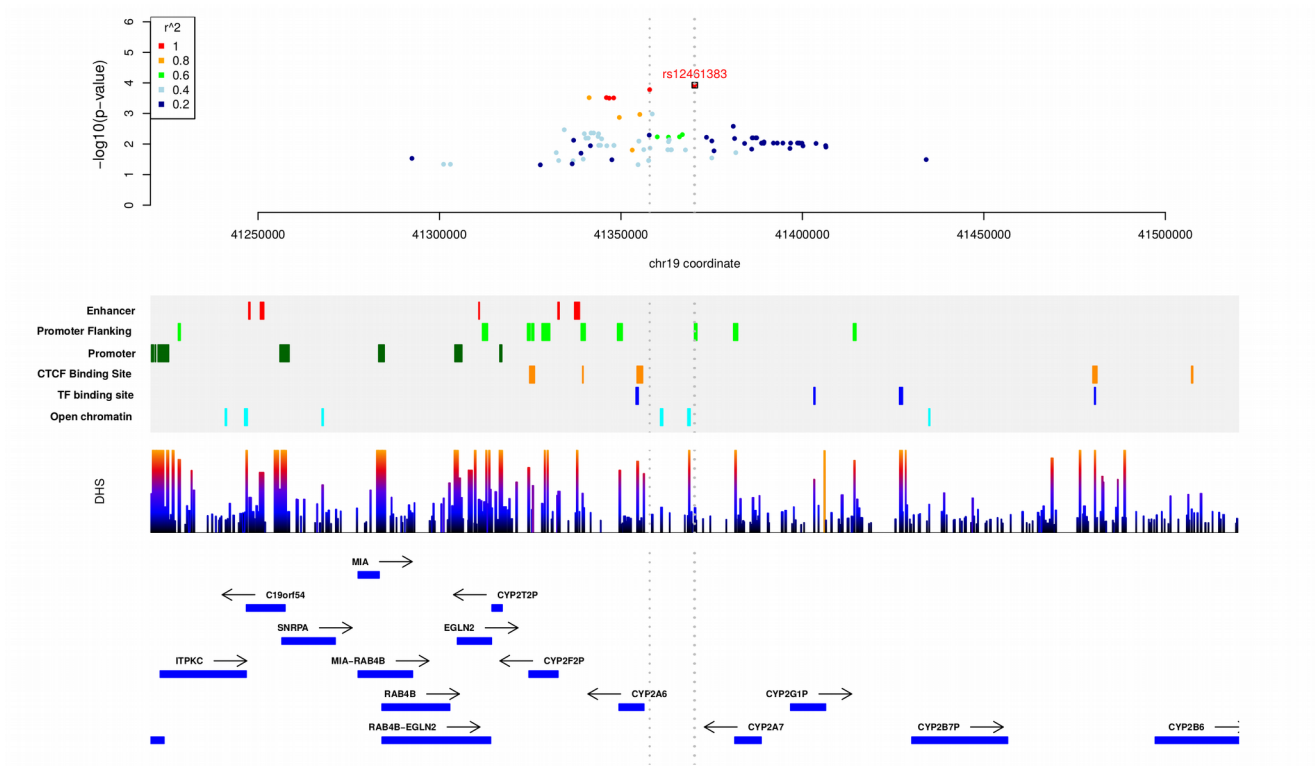


Figure S3. Regional genomic plot for rs12461383:*C19orf54* locus with regulatory features and gene information; red and blue marker colors indicates higher and lower linkage disequilibrium (LD), respectively.

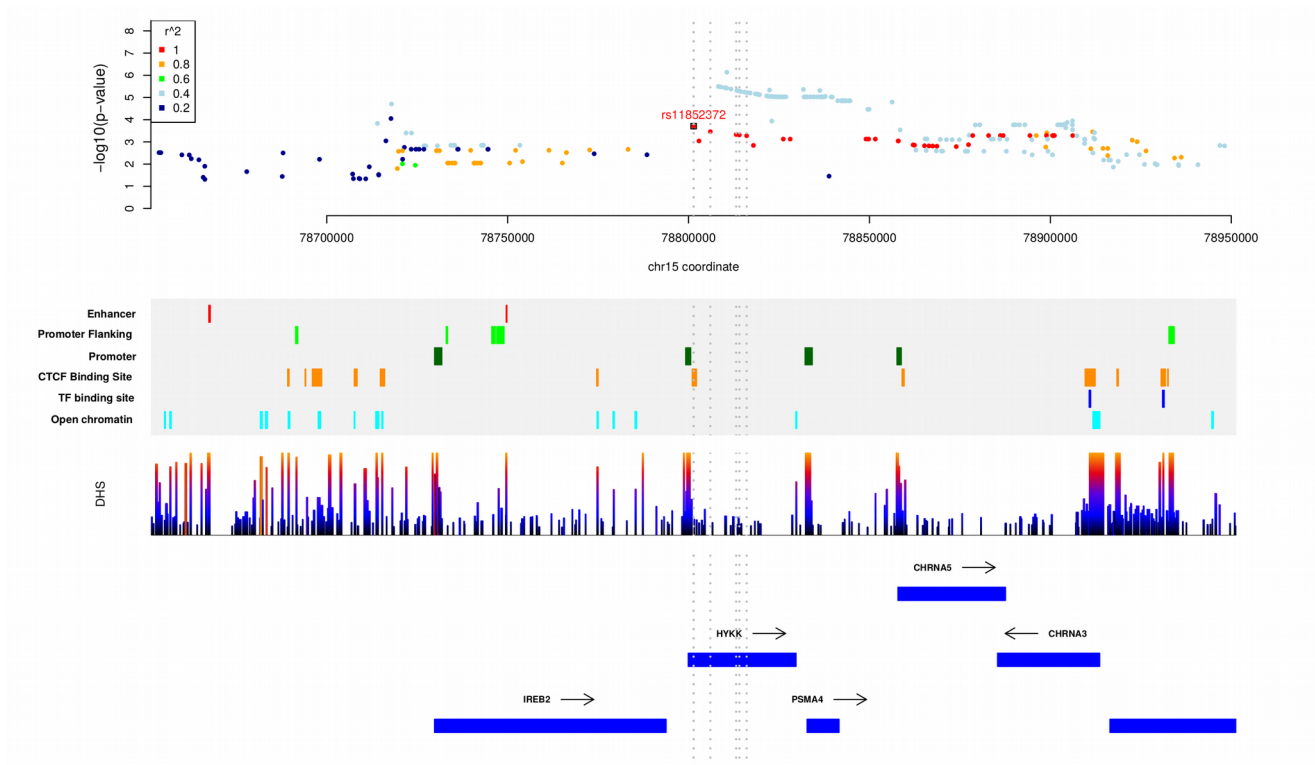


Figure S4. Regional genomic plot for rs11852372:*CHRNA5* locus with regulatory features and gene information; red and blue marker colors indicates higher and lower linkage disequilibrium (LD), respectively.

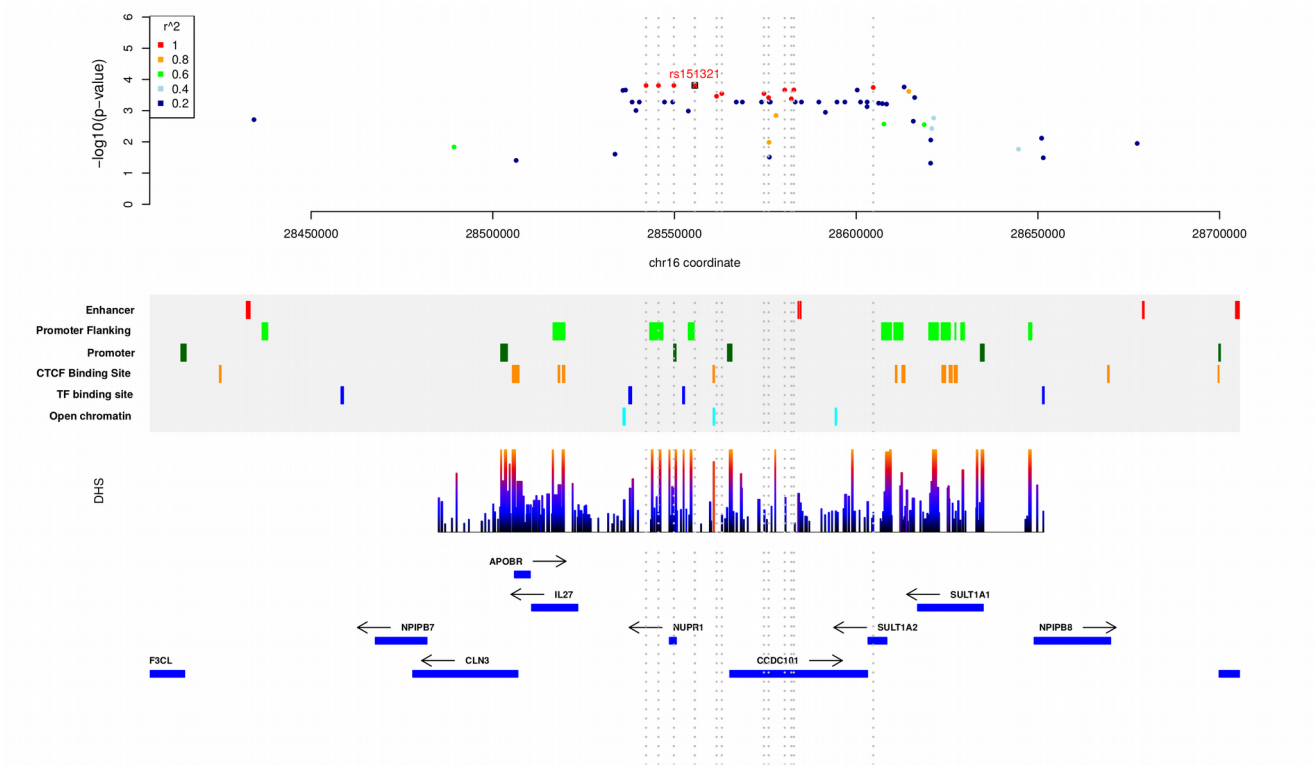


Figure S5. Regional genomic plot for rs151321:*SULT1A2* locus with regulatory features and gene information; red and blue marker colors indicates higher and lower linkage disequilibrium (LD), respectively.

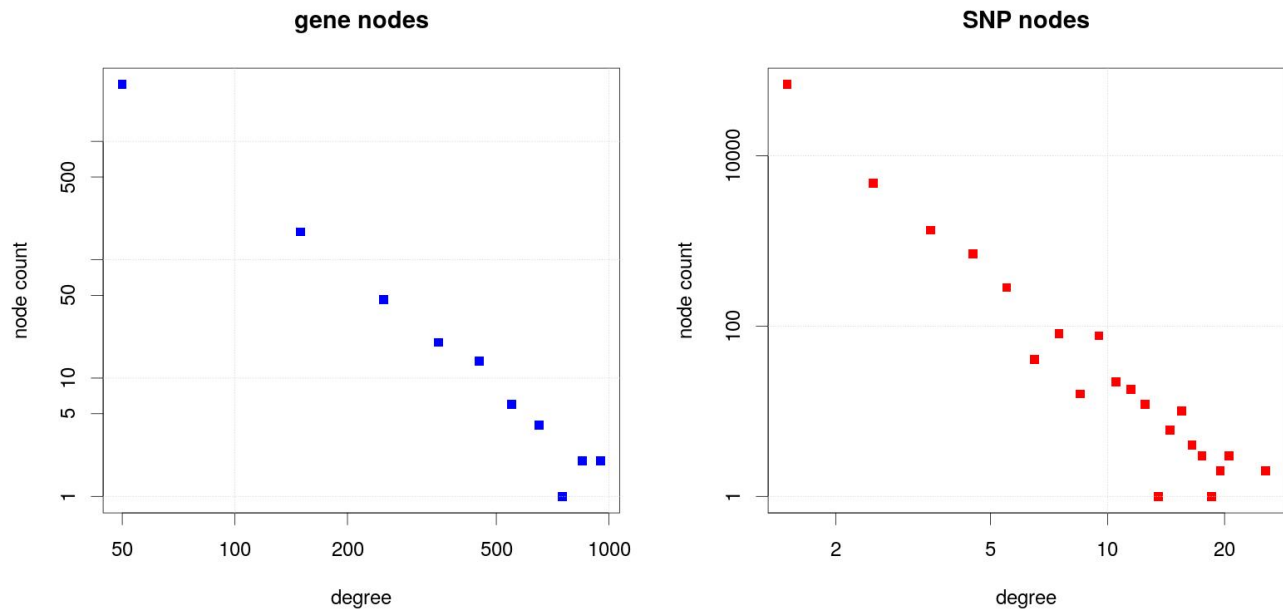


Figure S6. Degree distribution for the gene and SNP nodes

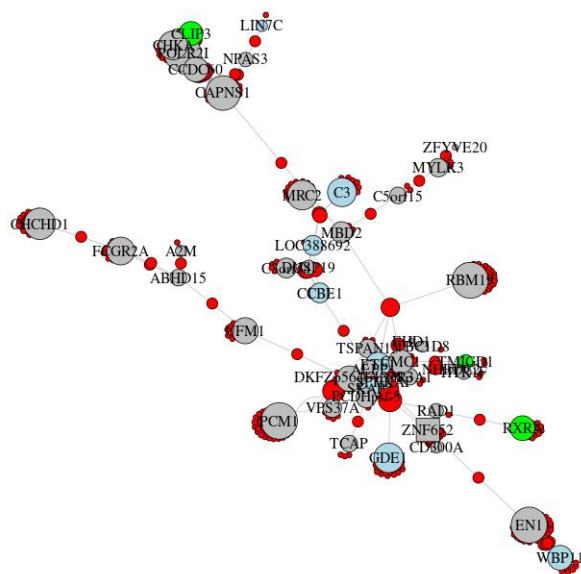


Figure S7. Community 218 containing the Sherlock gene *ZNF652*. Community genes are listed in Supplemental Table S6. (red = SNP, yellow = SNP with GWAS $p < 10^{-4}$; square = Sherlock gene, gray = gene, green = gene with differentially methylated site ($p < 0.05$ and effect $> 5\%$), light blue = gene with differentially expressed probe ($p < 0.05$), cyan = gene with differentially methylated site and differentially expressed probe)

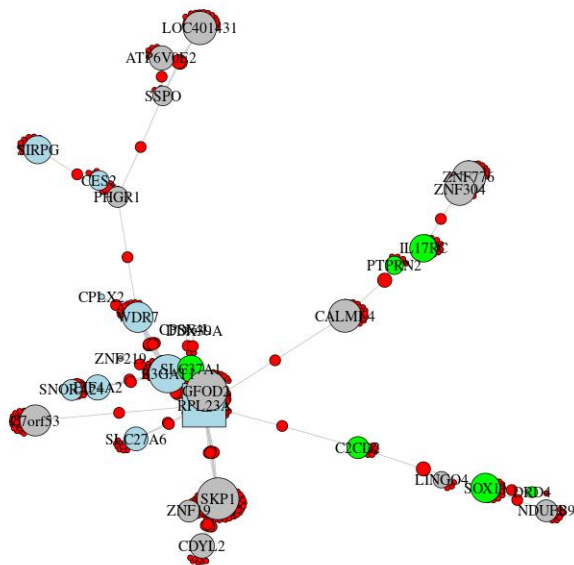


Figure S8. Community 223 containing the Sherlock gene *RPL23A* (sub-threshold SNP is hidden behind *RPL23A*). Community genes are listed in Supplemental Table S6. (red = SNP, yellow = SNP with GWAS $p < 10^{-4}$; square = Sherlock gene, gray = gene, green = gene with differentially methylated site ($p < 0.05$ and effect $> 5\%$), light blue = gene with differentially expressed probe ($p < 0.05$), cyan = gene with differentially methylated site and differentially expressed probe)

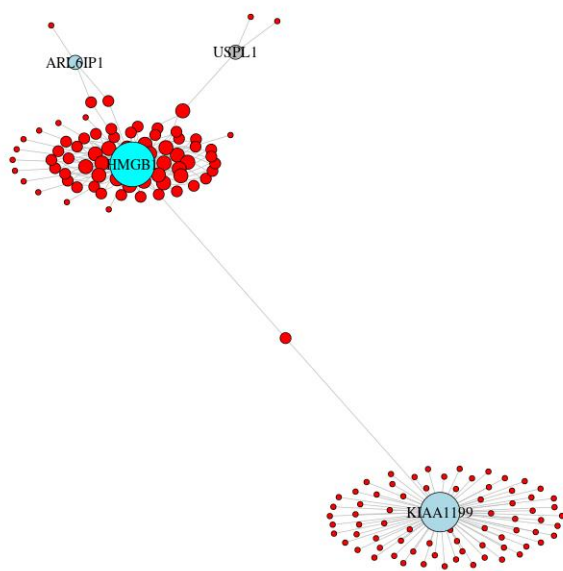


Figure S9. Community 98 containing the Sherlock gene *HMGB1*. Community genes are listed in Supplemental Table S6. (red = SNP, yellow = SNP with GWAS $p < 10^{-4}$; square = Sherlock gene, gray = gene, green = gene with differentially methylated site ($p < 0.05$ and effect $> 5\%$), light blue = gene with differentially expressed probe ($p < 0.05$), cyan = gene with differentially methylated site and differentially expressed probe)

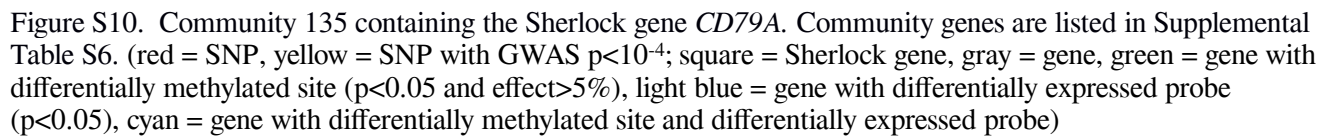


Table S1. Demographics of study subjects with expression and genotyping data

Demographics	COPD cases N=86	Control smokers N=31	P-value
Age, years	64.1 (±6.4)	65.4 (±9.9)	0.50
Female Sex	45 (52.3%)	18 (58.1%)	0.68
White Race	100%	100%	NA
Former Smokers	100%	100%	NA
Smoking History, pack-years	65.2 (± 26.9)	36.9 (± 22.3)	<0.0001
Time since quitting, months	72.0 (± 64.2)	176.1 (± 144.9)	0.0006
FEV 1 % predicted	26.4 (± 9.9)	97.1 (± 10.6)	<0.0001
FEV 1 /FVC	0.32 (± 0.11)	0.79 (± 0.05)	<0.0001
Body Mass Index	25.5(±4.6)	27.8 (±4.5)	0.02
Abbreviations: FEV1=forced expiratory volume in 1 sec; FVC= forced vital capacity			

Table S2. Top 20 cis-expression quantitative trait locus results. Most significant eSNP for each gene is shown.

probe ID	SNP	p-value	FDR q-value	Gene symbol	Chr	SNP to Gene distance (bases for hg19)
ILMN_3236498	rs1060817	5.31E-46	6.28E-39	<i>LOC253039</i>	chr9	27793
ILMN_3298167	rs8066902	3.59E-44	1.82E-37	<i>ZSWIM7</i>	chr17	8518
ILMN_1753164	rs11834524	2.62E-41	1.34E-35	<i>IPO8</i>	chr12	26468
ILMN_1726647	rs10876864	1.56E-40	7.48E-35	<i>RPS26</i>	chr12	35761
ILMN_1697499	rs72851099	3.48E-39	1.49E-33	<i>HLA-DRB1</i>	chr6	13928
ILMN_1798177	rs10129479	3.10E-37	4.53E-32	<i>FNTB</i>	chr14	16401
ILMN_1719064	rs6663	3.29E-37	4.76E-32	<i>KCTD10</i>	chr12	14301
ILMN_1747935	rs113512099	3.07E-36	3.06E-31	<i>GOLGB1</i>	chr3	20812
ILMN_2345908	rs4931424	2.11E-35	1.25E-30	<i>DDX11</i>	chr12	12935
ILMN_2214278	rs755898	6.31E-35	3.46E-30	<i>ANKRD32</i>	chr5	11387
ILMN_2165753	rs9260193	1.34E-34	6.83E-30	<i>HLA-A</i>	chr6	497
ILMN_3250243	rs2551943	1.67E-34	8.42E-30	<i>METTL21A</i>	chr2	14232
ILMN_1782178	rs11120049	1.92E-34	9.15E-30	<i>LQK1</i>	chr1	2929
ILMN_1678974	rs2295716	2.40E-34	1.10E-29	<i>MRPL43</i>	chr10	1789
ILMN_1778488	rs150478765	2.78E-34	1.26E-29	<i>WDR41</i>	chr5	41198
ILMN_3286411	rs78789402	7.33E-34	3.15E-29	<i>C22orf41</i>	chr22	5810
ILMN_1719170	rs4717819	7.73E-34	3.31E-29	<i>WBSCR27</i>	chr7	17926
ILMN_2312606	rs2272347	9.99E-34	4.22E-29	<i>IRF5</i>	chr7	35372
ILMN_3251217	rs8060185	1.45E-33	5.76E-29	<i>PDXDC2</i>	chr16	52368
ILMN_2120982	rs2277448	1.54E-33	6.06E-29	<i>ALG11</i>	chr13	9603

Table S3. Top 20 trans-expression quantitative trait locus results. Most significant eSNP for each gene is shown.

probe ID	SNP	p-value	FDR q-value	Gene symbol	SNP Chromosome	Probe Chromosome
ILMN_3235221	rs7612	1.85E-38	2.71E-27	LOC644936	chr7	chr5
ILMN_2393693	rs2532340	1.14E-32	3.77E-23	LRRC37A4	chr17	chr17
ILMN_1724634	rs143761582	2.95E-26	4.36E-18	FLJ00326	chr4	chr8
ILMN_2118663	rs146271029	5.20E-26	7.55E-18	ERV3	chr7	chr7
ILMN_1675387	rs9271225	1.80E-21	2.02E-13	LIMS1	chr6	chr2
ILMN_3238751	rs2286990	6.05E-20	4.88E-12	hPMS6	chr7	chr7
ILMN_1825186	rs10772397	7.30E-20	5.81E-12	LOC338817	chr12	chr12
ILMN_3248654	rs2668668	3.98E-19	2.93E-11	C17orf69	chr17	chr17
ILMN_3308808	rs10876864	1.11E-17	5.84E-10	MIR130A	chr12	chr11
ILMN_1781236	rs2596492	5.40E-17	2.56E-09	LOC554223	chr6	chr6
ILMN_1767509	rs2760975	6.47E-17	3.03E-09	DEF8	chr6	chr16
ILMN_2415926	rs202223551	1.59E-16	6.98E-09	THOC3	chr5	chr5
ILMN_3308158	rs35159999	2.65E-16	1.13E-08	MIR330	chr4	chr19
ILMN_3278245	rs10222496	3.13E-16	1.32E-08	ZBTB20	chr3	chr3
ILMN_1745116	rs1203329	1.58E-15	6.19E-08	ABHD12	chr20	chr20
ILMN_1667452	rs141815353	3.91E-15	1.50E-07	LOC441208	chr7	chr7
ILMN_1722426	rs116178429	4.50E-15	1.72E-07	OR7D2	chr19	chr19
ILMN_3235044	rs71399704	5.77E-15	2.18E-07	LOC100288615	chr15	chr15
ILMN_1676528	rs35789010	5.92E-15	2.23E-07	BTN3A2	chr6	chr6
ILMN_3251217	rs7190039	1.42E-14	5.24E-07	PDXDC2	chr16	chr16

Table S4. Cis-eQTL results (FDR<5%) intersected with sub-threshold GWAS ($p < 10^{-4}$). The most significant eSNP for each gene is shown.

probe ID	SNP	eQTL p-value	eQTL FDR q-value	Gene symbol	Differential expression p-value	GWAS p-value	Chr
ILMN_1726554	rs1504550	2.50E-12	3.27E-09	<i>IREB2</i>	0.119	6.35E-12	chr15
ILMN_1798308	rs1177287	1.60E-07	8.12E-05	<i>AHSA2</i>	0.499	1.51E-05	chr2
ILMN_1754501	rs1177292	1.54E-06	6.00E-04	<i>C2orf74</i>	0.292	7.83E-05	chr2
ILMN_2243308	rs2252518	1.85E-06	7.04E-04	<i>ACVR1B</i>	0.152	6.23E-05	chr12
ILMN_1738369	rs153106	2.30E-05	6.17E-03	<i>TUFM</i>	0.006	5.18E-05	chr16
ILMN_3238570	rs17707300	2.38E-05	6.34E-03	<i>EIF3CL</i>	0.644	5.16E-05	chr16
ILMN_2191192	rs2644899	2.79E-05	7.25E-03	<i>CYP2B7</i>	0.114	3.81E-05	chr19
ILMN_1767665	rs3111182	3.07E-05	7.85E-03	<i>GPX8</i>	0.116	8.00E-05	chr5
ILMN_1740319	rs61980585	8.13E-05	1.73E-02	<i>IFI27L2</i>	0.435	4.46E-05	chr14
ILMN_1729546	rs12461383	1.18E-04	2.34E-02	<i>C19orf54</i>	0.892	8.16E-06	chr19
ILMN_1798528	rs151321	1.55E-04	2.90E-02	<i>SULT1A2</i>	0.018	3.71E-05	chr16
ILMN_1657302	rs12447461	1.79E-04	3.23E-02	<i>SULT1A1</i>	0.563	9.08E-05	chr16
ILMN_1770044	rs11852372	1.96E-04	3.47E-02	<i>CHRNA5</i>	0.178	2.35E-11	chr15

Table S5. Top Sherlock results ($p < 10^{-3}$) including score for each expression quantitative trait locus

See additional file: Additional_File_2.pdf

Table S6. Genes within communities from Table 2

Comm ID	Sherlock or interactor gene(s)	Member genes
98	<i>HMGB1</i>	<i>ARL6IP1</i> , <i>HMGB1</i> , <i>KIAA1199</i> , <i>USPL1</i>
113	<i>CDH23</i>	<i>ALDH1A3</i> , <i>C14orf102</i> , <i>CDH23</i> , <i>CTDSPL2</i> , <i>FLJ34503</i> , <i>FOXRED2</i> , <i>GLO1</i> , <i>ID1</i> , <i>PSMC1</i> , <i>RBM43</i> , <i>ZCRB1</i> , <i>ZNF555</i>
135	<i>CD79A</i>	<i>ACN9</i> , <i>ACOT2</i> , <i>ACPI</i> , <i>ACTB</i> , <i>ACTG1</i> , <i>ADCY7</i> , <i>AGL</i> , <i>ANK3</i> , <i>ANP32B</i> , <i>ANX2P2</i> , <i>AQP4</i> , <i>ARF4</i> , <i>ARL6IP4</i> , <i>ATG4B</i> , <i>ATP2C1</i> , <i>ATP5A1</i> , <i>BAIAP2</i> , <i>BBX</i> , <i>BIRC2</i> , <i>BLVRB</i> , <i>BST2</i> , <i>C15orf44</i> , <i>C18orf45</i> , <i>C19orf6</i> , <i>C1orf64</i> , <i>C4orf41</i> , <i>C9orf7</i> , <i>CACYBP</i> , <i>CASC4</i> , <i>CCDC126</i> , <i>CCNC</i> , <i>CCS</i> , <i>CCT6B</i> , <i>CCT7</i> , <i>CD79A</i> , <i>CD79B</i> , <i>CDC42BPB</i> , <i>CHMP5</i> , <i>CLASP2</i> , <i>CNTD2</i> , <i>COG3</i> , <i>CPEB4</i> , <i>CRK</i> , <i>CSDE1</i> , <i>DISC1</i> , <i>DNAJA2</i> , <i>DOCK9</i> , <i>DPP7</i> , <i>DTX4</i> , <i>DUT</i> , <i>EAPP</i> , <i>EEF1A1</i> , <i>EFEMP1</i> , <i>EIF4G2</i> , <i>ERCC1</i> , <i>ERP29</i> , <i>FAM195A</i> , <i>FANCL</i> , <i>FLJ45079</i> , <i>FN1</i> , <i>GABPA</i> , <i>GALNT10</i> , <i>GATA3</i> , <i>GNG10</i> , <i>GOLGA2L1</i> , <i>GPD1</i> , <i>GPRI1</i> , <i>GRIK2</i> , <i>GTF2H2B</i> , <i>hCG_2028511</i> , <i>HIST1H3D</i> , <i>HMGXB3</i> , <i>HMP19</i> , <i>HSPA4</i> , <i>IDH1</i> , <i>IGDCC4</i> , <i>IGFN1</i> , <i>IL6ST</i> , <i>ILF3</i> , <i>ING3</i> , <i>ITIH1</i> , <i>JAK1</i> , <i>KBTBD6</i> , <i>KIAA1737</i> , <i>KIF21A</i> , <i>KRAS</i> , <i>KRT4</i> , <i>KRT80</i> , <i>LHX3</i> , <i>LOC100129794</i> , <i>LOC646576</i> , <i>LOC647979</i> , <i>LOC730101</i> , <i>LYRM2</i> , <i>LYSMD3</i> , <i>MAT2B</i> , <i>MCMBP</i> , <i>MED22</i> , <i>MGMT</i> , <i>MIR1908</i> , <i>MOCS2</i> , <i>MORC3</i> , <i>MTMR4</i> , <i>MUC13</i> , <i>NCOR1</i> , <i>NFYC</i> , <i>NLRP1</i> , <i>NLRX1</i> , <i>NR2C2</i> , <i>NUP35</i> , <i>NUP37</i> , <i>OIT3</i> , <i>PAPD5</i> , <i>PEMT</i> , <i>PPFIA1</i> , <i>PPP2R5B</i> , <i>PRDX2</i> , <i>PRDX3</i> , <i>PRKCDBP</i> , <i>PRNP</i> , <i>PSD</i> , <i>PSMA6</i> , <i>PSMC4</i> , <i>PWWP2B</i> , <i>RAB5C</i> , <i>RBBP4</i> , <i>RNF8</i> , <i>SBDSP</i> , <i>SERPINB10</i> , <i>SF1</i> , <i>SIRT3</i> , <i>SLC15A4</i> , <i>SLC38A10</i> , <i>SMPDL3A</i> , <i>SNAPC5</i> , <i>SPG21</i> , <i>SRP9</i> , <i>SRRM1</i> , <i>SSFA2</i> , <i>STOML2</i> , <i>STXBP3</i> , <i>SYPL1</i> , <i>SYT13</i> , <i>TAF1L</i> , <i>TALDO1</i> , <i>TCERG1</i> , <i>TIAL1</i> , <i>TMCO2</i> , <i>TMED10</i> , <i>TMEM43</i> , <i>TSPAN12</i> , <i>TSSC4</i> , <i>TUBB</i> , <i>UBA3</i> , <i>USP14</i> , <i>WASF2</i> , <i>YAP1</i> , <i>ZDHC2</i> , <i>ZNF32</i> , <i>ZNF384</i> , <i>ZNF442</i> , <i>ZNF572</i>
202	<i>CHRNA5</i> <i>HNRNPAB</i> <i>IREB2</i> <i>PCBP2</i>	<i>C11orf80</i> , <i>CHRNA5</i> , <i>FBXO16</i> , <i>GGTLC1</i> , <i>HNRNPAB</i> , <i>IREB2</i> , <i>LIN54</i> , <i>LOC100286979</i> , <i>LOC283392</i> , <i>MAP6</i> , <i>PAQR6</i> , <i>PCBP2</i> , <i>RBM14</i> , <i>TARBP2</i> , <i>TECPRI</i> , <i>TLK1</i> , <i>ZNF114</i>
218	<i>ZNF652</i>	<i>A2M</i> , <i>ABHD15</i> , <i>ALPP</i> , <i>C3</i> , <i>C3orf34</i> , <i>C5orf15</i> , <i>CAPNS1</i> , <i>CCBE1</i> , <i>CCDC60</i> , <i>CD300A</i> , <i>CHCHD1</i> , <i>CHKA</i> , <i>CLIP3</i> , <i>CMC1</i> , <i>DKFZp564F1378</i> , <i>DUSP19</i> , <i>EHD1</i> , <i>EN1</i> , <i>ETF1</i> , <i>FCGR2A</i> , <i>GDE1</i> , <i>HTR1F</i> , <i>LIN7C</i> , <i>LOC388692</i> , <i>MBD2</i> , <i>MRC2</i> , <i>MYLK3</i> , <i>NHEDC2</i> , <i>NPAS3</i> , <i>OR3A1</i> , <i>PCDHpsi-5</i> , <i>PCMI</i> , <i>POLR2I</i> , <i>RAD1</i> , <i>RBM19</i> , <i>RXRA</i> , <i>SKAP2</i> , <i>SPHKAP</i> , <i>TBC1D8</i> , <i>TCAP</i> , <i>TMIGD1</i> , <i>TSPAN13</i> , <i>UFM1</i> , <i>VPS37A</i> , <i>WBP11</i> , <i>ZFYVE20</i> , <i>ZNF652</i>
222	<i>ACVR1B</i>	<i>ACVR1B</i> , <i>ADCY5</i> , <i>ADSSL1</i> , <i>ANKS1B</i> , <i>AP3D1</i> , <i>AQPEP</i> , <i>C19orf28</i> , <i>CHD4</i> , <i>CLEC4GPI</i> , <i>CPNE4</i> , <i>CRISPLD1</i> , <i>CSRNP1</i> , <i>DCLK1</i> , <i>DGKQ</i> , <i>DKKL1</i> , <i>EIF3C</i> , <i>FSCB</i> , <i>GMDS</i> , <i>HAND2</i> , <i>HSPB7</i> , <i>IDUA</i> , <i>ILK</i> , <i>INVS</i> , <i>IQGAP1</i> , <i>KIAA1429</i> , <i>LOC221442</i> , <i>LOC399959</i> , <i>LYNX1</i> , <i>MED13L</i> , <i>MFGE8</i> , <i>MPPED2</i> , <i>MYL9</i> , <i>MYOZ2</i> , <i>NCSI</i> , <i>NPR2</i> , <i>NTN1</i> , <i>NTRK3</i> , <i>PACSI</i> , <i>PDP1</i> , <i>PRDM16</i> , <i>PRDM6</i> , <i>PTPLA</i> , <i>RNF115</i> , <i>RNF138</i> , <i>RTN1</i> , <i>SCMH1</i> , <i>SCUBE3</i> , <i>SEMA4B</i> , <i>SGK196</i> , <i>SHC4</i> , <i>SHKBP1</i> , <i>SHROOM3</i> , <i>SLC10A7</i> , <i>SLC16A7</i> , <i>SLC25A4</i> , <i>SLMAP</i> , <i>SNORA1</i> , <i>SNRNP70</i> , <i>SRSF11</i> , <i>TCEB2</i> , <i>TCF3</i> , <i>TMC7</i> , <i>TRIP6</i> , <i>TRPT1</i> , <i>UCP3</i> , <i>VCL</i> , <i>ZNF609</i>
223	<i>RPL23A</i>	<i>ATP6V0E2</i> , <i>B3GAT1</i> , <i>C2CD2</i> , <i>C7orf53</i> , <i>CALML4</i> , <i>CDYL2</i> , <i>CES2</i> , <i>CPLX2</i> , <i>CPSF4L</i> , <i>DDX39A</i> , <i>DRD4</i> , <i>EIF4A2</i> , <i>GFOD2</i> , <i>IL17RC</i> , <i>LINGO4</i> , <i>LOC401431</i> , <i>NDUFB9</i> , <i>PHGR1</i> , <i>PTPRN2</i> , <i>RPL23A</i> , <i>SIRPG</i> , <i>SKP1</i> , <i>SLC27A6</i> ,

		<i>SLC37A1, SNORA24, SOX11, SSPO, WDR7, ZNF19, ZNF219, ZNF304, ZNF776</i>
20	<i>WDR47</i>	<i>ACER2, AKD1, C14orf169, GPSM2, SMPD2, WDR47</i>
78	<i>CHIAP2</i>	<i>ATP5F1, C1orf88, CHI3L2, CHIA, DYRK3, FBN2, FOXM1, GABRB3, KCNN2, CHIAP2, LOC646324, LYN, PITX2, SGTA, SLC39A3, TICAM1, TIMM23, ZCWPW2</i>
131	<i>AHSA2</i> <i>C2orf74</i>	<i>AHSA2, C2orf74, SLC33A1, XPO1</i>
161	<i>SMG6</i>	<i>ABLIM1, C17orf109, C17orf79, ENPP4, FBXO46, FLJ43663, GOLGA6L1, LOC728643, LTBP3, PAPD4, RPL11, RPS3A, SIGMAR1, SMG6, SRR, TCFL5, TMEM44, WDR23</i>
179	<i>DSP</i>	<i>DSP, EVI2B, FEM1A, FOXD4L1, OSR2, SF3B4, SPATA22</i>
181	<i>FSTL5</i>	<i>ATP6V0A4, CEP57, CLEC10A, CYCSP52, DKFZp686K04236, DPY19L2P2, DPY30, FAM92B, FSTL5, HMHB1, LOC728175, LYPLA2, MAF, MEF2D, MGC16075, MTMR14, NLRP7, PDPR, RAX2, RNF180, SNORD114-31, ZBTB2, ZNF668</i>
187	<i>SNRPB</i>	<i>GNA12, GTF2H4, HYAL4, NPDC1, POLR2J2, RNF5, RRM2B, SEPHS1, SIRT7, SNAPC2, SNRPB, TENC1, TUBB1, WASH2P</i>
210	<i>CTSH</i>	<i>AIF1, C10orf57, CTSH, GATSL3, GOLGB1, HCLS1, ILDR1, IQCB1, KEL, LSM11, STARD3, WDR5B</i>
249	<i>TRIM4</i>	<i>AZGP1, HIP1R, INTS6, KDM3B, MEPCE, PILRB, PODXL2, SFT2D2, TRIM4, TSC22D4, ZKSCAN1</i>

Table S7. Significant pathway analysis results (FDR q-value < 0.05) from ConsensusPathDB for the communities of interest

Community ID (Sherlock or interactor gene)	Pathway name	Gene set size (number found among available expression data genes)	Overlap (percentage of available genes that overlap)	p-value	FDR q-value	pathway source
222 (<i>ACVR1B</i>)	cGMP-PKG signaling pathway	168(145)	4 (2.8%)	0.0035	0.037	KEGG
	Bacterial invasion of epithelial cells	78(71)	3 (4.2%)	0.0036	0.037	KEGG
	Focal adhesion	201(178)	4 (2.2%)	0.0072	0.051	KEGG
135 (<i>CD79A</i>)	Folding of actin by CCT/TriC	10(10)	3 (30.0%)	0.00019	0.032	Reactome
	Regulation of actin dynamics for phagocytic cup formation	139(48)	5 (10.4%)	0.00025	0.032	Reactome
	RHO GTPases Activate WASPs and WAVEs	37(31)	4 (12.9%)	0.00047	0.037	Reactome
	Innate Immune System	1309(1000)	24 (2.4%)	0.00058	0.037	Reactome
	Fcgamma receptor (FCGR) dependent phagocytosis	157(64)	5 (7.8%)	0.00095	0.048	Reactome
	Immune System	1950(1519)	31 (2.0%)	0.0013	0.0497	Reactome
	CD22 mediated BCR regulation	71(5)	2 (40.0%)	0.00137	0.0497	Reactome

Table S8. Summary of genotype data quality control process

Subjects		Markers	
Genotyped subjects after initial standard QC	183	Genotyped markers after initial standard QC	2,556,097
Missingness > 5%	-0	Markers with technical issues	-3415
Relatedness (sample contamination/swap)	-6	Missingness > 1%	-149,608
Gender mismatch	-1	HWE (767 total flagged markers)	-709 unique
Duplicate sample	-1	Mapping issues and monomorphic	-498,967
Discordant genetic ancestry	-2	Markers passing QC – available for imputation	1,903,398
Control samples	-4	Markers after imputation with MAF>5% and info>0.3	6,347,053
Subject data available	169	Duplicate markers	-4
Caucasian subjects with expression data	117	Markers with MAF≤5% or info≤0.5 for subset of 117 subjects	-357,198
Female: 45 cases, 18 controls Male: 41 cases, 13 controls		Total markers included in analysis	5,989,851