Multi-Study Fine Mapping of Chromosome 2q Identifies *XRCC5* as a COPD Susceptibility Gene

Craig P. Hersh, Sreekumar G. Pillai, Guohua Zhu, David A. Lomas, Per Bakke, Amund Gulsvik, Dawn L. DeMeo, Barbara J. Klanderman, Ross Lazarus, Augusto A. Litonjua, David Sparrow, John J. Reilly, Alvar Agusti, Peter M. A. Calverley, Claudio F. Donner, Robert D. Levy, Barry J. Make, Peter D. Paré, Stephen I. Rennard, Jørgen Vestbo, Emiel F. M. Wouters, Mary Beth Scholand, Hilary Coon, John Hoidal, Edwin K. Silverman

On-line data supplement

Methods

Study Populations

Subject enrollment and phenotype determination in the National Emphysema Treatment Trial (NETT)-Normative Aging Study (NAS) case-control study, the Boston Early-Onset COPD Study, the Norway case-control study, and the International COPD Genetics Network (ICGN) have been described previously (1-3). NETT Genetics Ancillary Study subjects consisted of 309 non-Hispanic white COPD cases without α1-antitrypsin (AAT) deficiency from 16 of the 17 NETT centers across the U.S. Subjects enrolled in NETT had $FEV_1 \leq 45\%$ predicted, hyperinflated lung volumes, and emphysema on chest CT scans (4). 330 control smokers (at least 10 pack-years) with normal lung function ($FEV_1 > 80\%$ predicted with $FEV_1/FVC > 90\%$ predicted) were derived from the NAS, a longitudinal study of aging in initially healthy men conducted by the Boston VA (5). The Boston Early-Onset COPD Study included 949 individuals in 127 extended pedigrees recruited through a proband with physician-diagnosed COPD, pre-bronchodilator $FEV_1 < 40\%$ predicted, age<53 years, without severe α1-antitrypsin deficiency.

Subjects in the Norway case-control study were recruited from Bergen, Norway. This analysis was restricted to 806 Caucasian COPD cases with $FEV_1 < 70\%$ predicted and $FEV_1/FVC < 0.7$ and 779 control smokers with $FEV_1 > 85\%$ predicted and $FEV_1/FVC \geq 0.7$. Both COPD cases and control subjects had at least 2.5 pack-years of smoking. We included 1839 white individuals from 603 families in the ICGN study, recruited at 10 centers in the U.S. and Europe. ICGN probands had $FEV_1 < 60\%$ predicted with $FEV_1/VC < 90\%$ predicted at age 45-65, with at least 5 pack-years smoking history and at least one sibling with a smoking history (≥5

pack-years). The highest value of slow or forced vital capacity (FVC) was used as a measure of vital capacity (VC) in ICGN.

The Lung Health Study (LHS) enrolled smokers ages 35-60 with COPD, defined by $FEV_1/FVC \leq 0.7$ and FEV_1 55-90% predicted, from 10 centers across North America (6). This analysis included baseline spirometry values in 5606 white subjects, as well as a comparison between subjects with slow vs. fast decline in FEV_1 across the study, based on published thresholds (7).

Studies were approved by the relevant institutional review boards, and subjects provided written informed consent.

SNP Genotyping

The Boston Early-Onset COPD Study had previously demonstrated linkage to chromosome 2q (8, 9), with a 1.5 LOD-drop support interval (similar to a 95% confidence interval (10)) ranging from 214.5-234.1 Mb, based on the May 2004 human genome reference sequence (NCBI Build 35). In the NETT-NAS study, 2484 single nucleotide polymorphisms (SNPs) were genotyped at Brigham and Women's Hospital (BWH) using Illumina (San Diego, CA) GoldenGate custom SNP panels. We selected SNPs to ensure linkage disequilibrium coverage ($r^2>0.8$, MAF>0.1) (11) across the region based on Phase I data from Caucasians in the International HapMap project (12). Additional SNPs were selected in regions of significant association. In the ICGN Study, preliminary linkage analysis identified a region on chromosome 2q to be linked to quantitative CT emphysema traits (13). A set of 2843 SNPs across this linkage interval (207.36-229.56 Mb) was genotyped by GlaxoSmithKline in both the ICGN and Norway studies, using Illumina GoldenGate custom panels. The overlapping region between NETT-NAS

and Norway encompassed 15 Mb on chromosome 2q: 214,552,727- 229,554,328 (NCBI Build 35). SNPs in the Boston Early-Onset COPD Study were genotyped using Sequenom (San Diego, CA) or TaqMan (Applied Biosystems, Foster City, CA) assays at BWH. LHS subjects were genotyped using TaqMan at the University of Utah. All genotyping platforms had high completion rates and few Mendelian inconsistencies in the two family-based studies.

In the NETT-NAS case-control study, an additional set of 195 common intergenic SNPs throughout the genome, excluding regions linked to COPD, had been previously genotyped using the Illumina GoldenGate assay and showed no evidence of population stratification ($\chi^2_{195\ df}$ =209.3, p=0.23) (14).

Statistical Analysis

All SNPs were tested for deviations from Hardy-Weinberg equilibrium in the control subjects or in the family members, as appropriate. In both the NETT-NAS and Norway case-control studies, SNPs were analyzed initially using the Cochran-Armitage trend test, which corresponds to an additive genetic model without covariate adjustment, using SAS/Genetics (SAS Institute, Cary, NC). Secondary analyses were adjusted for pack-years of smoking. In the Boston Early-Onset COPD and ICGN studies, SNPs were tested for association with COPD status, defined as GOLD 2 or greater (15), without covariate adjustment (16), and for association with post-bronchodilator spirometry measurements, adjusted for age, sex, height, and pack-years of smoking in both study samples, plus ever-smoking status in the Boston Early-Onset COPD Study only. SNPs were analyzed under additive models, using PBAT version 3.3 in the Boston Early-Onset COPD Study and Golden Helix (Bozeman, MT) PBAT in the ICGN (17). Fisher's method was used to combine p-values of overlapping SNPs across studies (18). Two-sided p-

values were used in the screening phase in the two case-control studies, and one-sided p-values were used in the replication analyses in the two family-based studies, to guard against SNP associations in the opposite direction producing deceptively small p-values(19). Secondary analyses of COPD were adjusted for pack-years of smoking, using logistic regression in the two case-control samples, and in the PBAT analyses of the two family-based studies. Additional analyses were stratified by age (<65 vs. ≥65) in NETT-NAS and Norway. Analysis of severe COPD, defined as GOLD 3 or greater, was performed in all four study populations.

After the initial association analysis was completed, genotype imputation was performed in the region encompassing genes *MREG* (melanoregulin), *PECR* (peroxisomal trans-2-enoyl-CoA reductase), *TMEM169* (transmembrane protein 169), and *XRCC5* (X-ray repair complementing defective repair in Chinese hamster cells 5). SNP genotypes were imputed in the NETT-NAS and Norway case-control studies using MACH1.0 software (20), implemented on the HapMap website (www.hapmap.org), using SNP genotype data from HapMap Phase II Caucasians as a reference. Imputed genotypes were analyzed as above. Haplotype blocks were defined by the method of Gabriel et al. (21). Linkage disequilibrium was calculated in Haploview (22).

DNA Sequencing

The exons, intron-exon boundaries, and promoter region of *XRCC5* and the neighboring gene *PECR* were resequenced in 23 probands from the Boston Early-Onset COPD Study and 1 CEPH control subject using dye-labeled dideoxy sequencing reactions and an ABI 3100 DNA sequencing machine (Applied Biosystems, Foster City, CA). Primer sequences are available on request. Sequence tracings were analyzed with Phred, Phrap, and Consed (23, 24), and

polymorphisms were identified using PolyPhred (25) and by manual review. SNPs with >5% minor allele frequency identified by sequencing were subsequently genotyped in the NETT-NAS and the Boston Early-Onset COPD Studies using TaqMan (Applied Biosystems, Foster City, CA) or Sequenom (San Diego, CA) assays.

References

E1. Hersh CP, Demeo DL, Lange C, Litonjua AA, Reilly JJ, Kwiatkowski D, Laird N, Sylvia JS, Sparrow D, Speizer FE, et al. Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am J Respir Cell Mol Biol* 2005;33:71-78.

E2. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, Anderson WH, Lomas DA, Silverman EK, Pillai SG. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am J Respir Crit Care Med* 2007;176:167-173.

E3. Patel BD, Coxson HO, Pillai SG, Agusti AG, Calverley PM, Donner CF, Make BJ, Muller NL, Rennard SI, Vestbo J, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2008;178:500-505.

E4. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, Weinmann G, Wood DE. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348:2059-2073.

E5. Bell B, Rose CL, Damon H. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* 1972;3:5-17.

E6. Anthonisen NR, Connett JE, Kiley JP, Altose MD, Bailey WC, Buist AS, Conway WA, Jr., Enright PL, Kanner RE, O'Hara P, et al. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA* 1994;272:1497-1505.

E7. Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, Pare PD. Susceptibility genes for rapid decline of lung function in the Lung Health Study. *Am J Respir Crit Care Med* 2001;163:469-473.

E8. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, et al. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *Am J Hum Genet* 2002;70:1229-1239.

E9. DeMeo DL, Celedon JC, Lange C, Reilly JJ, Chapman HA, Sylvia JS, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004;170:1294-1301.

E10. Dupuis J, Siegmund D. Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* 1999;151:373-386.

E11. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106-120.

E12. International HapMap Consortium. The International HapMap Project. *Nature* 2003;426:789-796.

E13. Silverman EK, Lake SL, Lomas DA, Rennard SI, Pare PD, Calverley PM, Vestbo J, Wouters EF, Agusti AG, Donner CF, et al. Linkage analysis of quantitative emphysema phenotypes in chronic obstructive pulmonary disease [abstract]. *Proc Am Thorac Soc* 2006;3:A620.

E14. Hersh CP, Hansel NN, Barnes KC, Lomas DA, Pillai SG, Coxson HO, Mathias RA, Rafaels NM, Wise RA, Connett JE, et al. Transforming growth factor-beta receptor-3 is associated with pulmonary emphysema. *Am J Respir Cell Mol Biol* 2009;41:324-331.

E15. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med* 2001;163:1256-1276.

E16. Poisson LM, Rybicki BA, Coon SW, Barnholtz-Sloan JS, Chase GA. Susceptibility scoring in family-based association testing. *BMC Genet* 2003;4 Suppl 1:S49.

E17. Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: tools for family-based association studies. *Am J Hum Genet* 2004;74:367-369.

E18. Fisher RA. Statistical Methods for Research Workers. London: Oliver and Boyd, Ltd.; 1925.

E19. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB, Willis-Owen SA, Klanderman B, Lasky-Su J, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet* 2009;84:581-593.

E20. Li Y, Ding J, Abecasis GR. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference[abstract]. *Am J Hum Genet* 2006;79:S2290.

E21. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-2229.

E22. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-265.

E23. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998;8:175-185.

E24. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998;8:195-202.

E25. Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 1997;25:2745-2751.

Figure Legends

Figure E1. Linkage disequilibrium (LD) plot in the NETT-NAS study of genotyped and imputed single nucleotide polymorphisms (SNPs) in the 285kb interval encompassing genes *MREG*, *PECR*, *TMEM169*, and *XRCC5*. LD is plotted as D' values, with darker red color denoting higher LD. SNP rs3821104 is marked by the arrow.

Figure E2. Linkage disequilibrium (LD) plot in the Norway study of genotyped and imputed single nucleotide polymorphisms (SNPs) in the 285kb interval encompassing genes *MREG*, *PECR*, *TMEM169*, and *XRCC5*. LD is plotted as D' values, with darker red color denoting higher LD. SNP rs3821104 is marked by the arrow.

Table E1: Genetic association results for overlapping single nucleotide polymorphisms (SNPs) in the *SERPINE2* gene in the National Emphysema Treatment Trial (NETT)-Normative Aging Study (NAS) and Norway case-control studies.

SNP	Location*	Norway p-value	NETT-NAS p-value	Combined p-value
rs6754561	224665201	0.91	0.22	0.53
rs16865390	224671520	0.66	0.07	0.19
rs975278	224673212	0.49	0.94	0.82
rs7583463	224674614	0.38	0.42	0.45
rs6748795	224676228	0.47	0.42	0.51
rs3795880	224676812	0.64	0.38	0.59
rs6721140	224684148	0.42	0.16	0.25
rs11695803	224686666	0.02	0.48	0.06
rs3795879	224688326	0.23	0.09	0.10
rs6747096	224688347	0.29	0.11	0.14
rs10196778	224689101	0.21	0.10	0.10
rs13392412	224689945	0.29	0.10	0.13
rs1530020	224696875	0.46	0.91	0.79
rs4674839	224704712	0.11	0.90	0.33
rs920251	224718450	0.30	0.14	0.18

* Chromosomal locations based on NCBI build 35.

Table E2: Genetic association results for SNP rs3821104 with severe COPD (GOLD 3 or greater)

SNP	Gene	NETT-NAS		Norway		ICGN		EOCOPD		Combined
		OR	p-value	OR	p-value	COPD risk	p-value*	COPD risk	p-value*	p-value
rs3821104	*XRCC5*	1.41	0.0036	1.16	0.097	increased	0.05	increased	0.00017	4.3×10^{-6}

*1-sided p-values in follow-up family-based studies

Figure E1

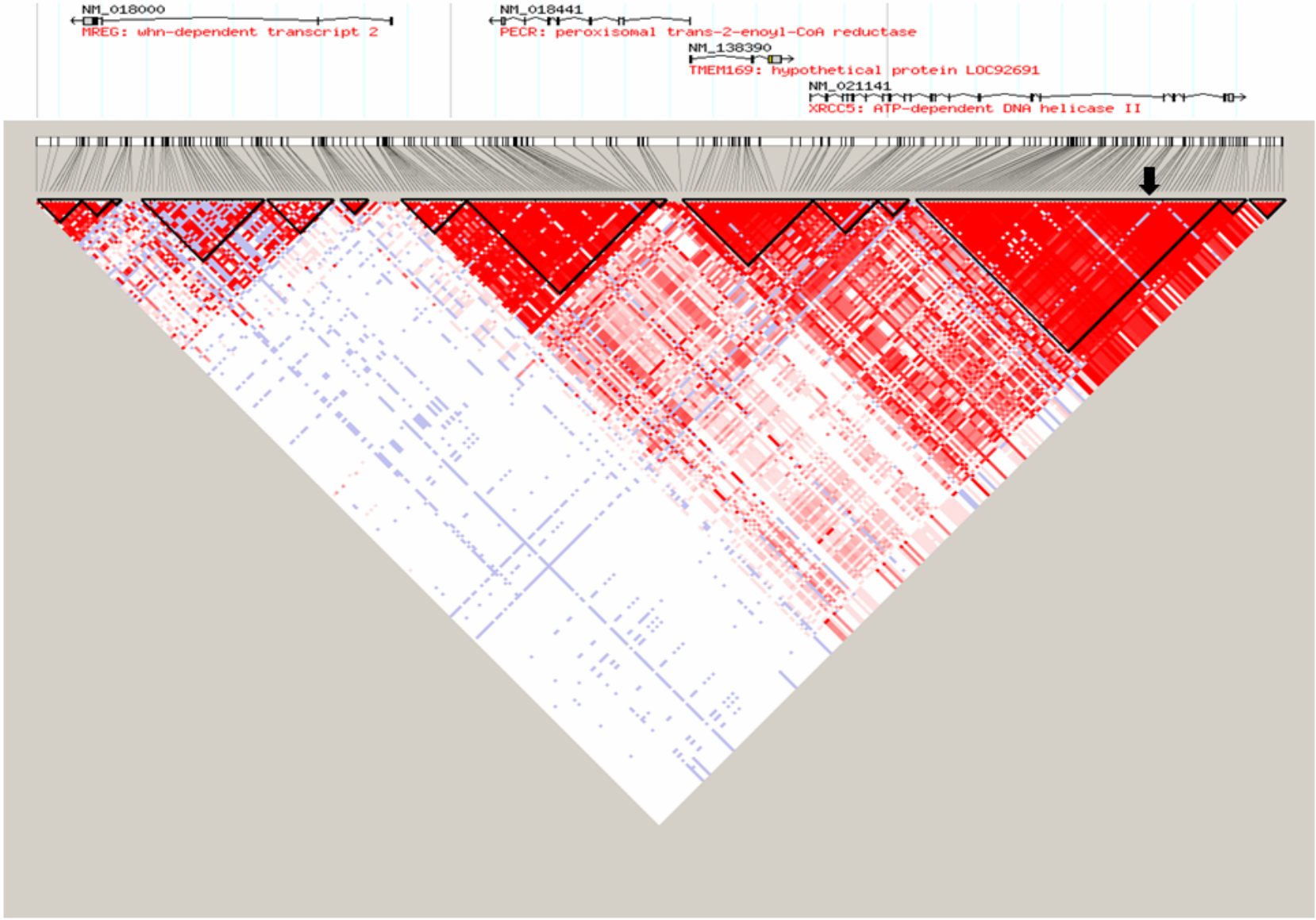
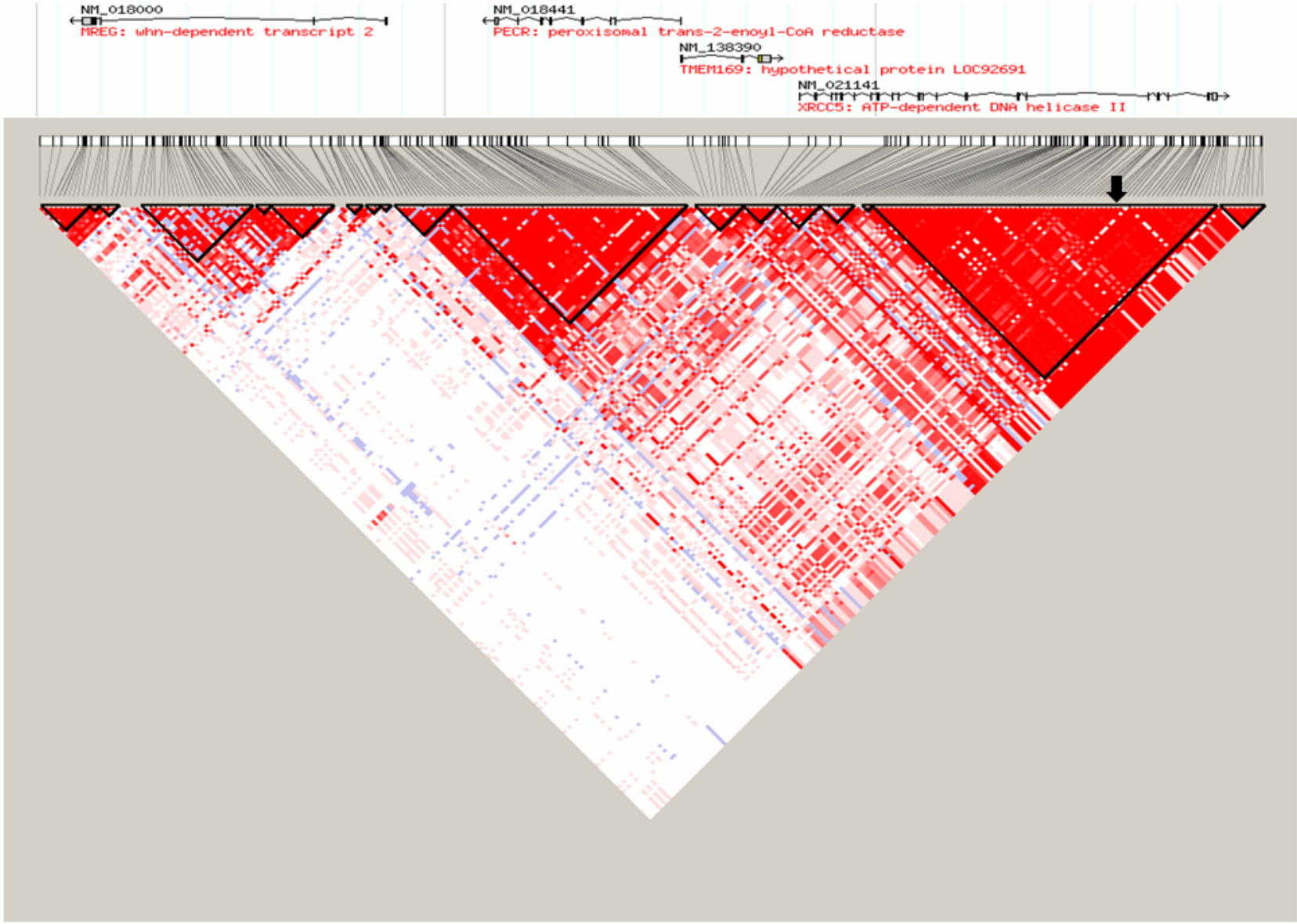

NM_018000
MREG: whn-dependent transcript 2
NM_018441
PECR: peroxisomal trans-2-enoyl-CoA reductase
NM_138390
TMEM169: hypothetical protein LOC92691
NM_021141
XRCC5: ATP-dependent DNA helicase II

Figure E2


NM_018000
MREG: whn-dependent transcript 2
NM_018441
PECR: peroxisomal trans-2-enoyl-CoA reductase
NM_138390
TMEM169: hypothetical protein LOC92691
NM_021141
XRCC5: ATP-dependent DNA helicase II