Discovery and Fine Mapping of Serum Protein Loci through Transethnic Meta-analysis

Nora Franceschiní,1,6,8,* Frank J.A. van Rooij,2,3,6,8 Bram P. Prins,4,6,8 Mary F. Feitosa,5,6,8 Mahir Karakas,5,6,8 John H. Eckfeldt,7 Aaron R. Folsom,8 Jeffrey Kopp,9 Ahmad Vaez,4 Jeanette S. Andrews,10 Jens Baumert,11 Vesna Boraska,12 Linda Broer,2,3 Caroline Hayward,13 Julius S. Ngwa,14 Yukinori Okada,15,16 Ozren Polasek,17 Harm-Jan Westra,18 Ying A. Wang,14,19 Fabiola Del Greco M.,20 Nicole L. Glazer,21 Karen Kapur,22,34 Ido P. Kema,23 Lorna M. Lopez,24,25 Arne Schillert,26 Albert V. Smith,27,28 Cheryl A. Winkler,29 Lina Zgaga,30,31 The LifeLines Cohort Study,32 Stefania Banninelli,33 Sven Bergmann,22,34 Mdaden Boban,35 Murielle Bochud,36 Y.D. Chen,37 Gail Davies,25 Abbas Dehghan,2,3 Jingzhong Ding,38 Angela Doering,11 J. Peter Durda,59 Luigi Ferrucci,40 Oscar H. Franco,2,3 Lude Franke,18 Grog Gunjaca,33 Albert Hofman,2,3 Fang-Chi Hsu,10 Ivana Kolcic,17 Aldi Kraja,5 Michiaki Kubo,41 Karl J. Lacker,42 Lenore Launer,43 Laura R. Loebh,1 Guo Li,44 Christa Meisinger,11 Yusuke Nakamura,45 Christine Schwienbacher,20,26 John M. Starr,24,47 Atsushi Takahashi,15 Vesela Torlak,48 André G. Uitterlinden,2,3 Veronique Vitart,15 Melanie Waldenberger,49 Philipp S. Wild,50 Mirna Kirin,51 Tanja Zeller,51 Tatijana Zemunik,12 Qingyun Zhang,3 Andreas Ziegler,26 Stefan Blankenberg,51 Eric Boerwinkle,52 Ingrid B. Borecki,5 Harry Campbell,30 Ian J. Deary,24,25 Timothy M. Frayling,53 Christian Gieger,11 Tamara B. Harris,43 Andrew A. Hicks,20 Wolfgang Koenig,6 Christopher J. O’Donnell,54,55 Caroline S. Fox,54,56 Peter P. Pramstaller,20,57 Bruce M. Psaty,58,59 Alex P. Reiner,60 Jerome I. Rotter,61 Igor Rudan,30 Harold Snieder,4 Toshio Tanaka,62 Cornelia M. van Duijn,2,3 Peter Vollenweider,63 Gerard Waeber,63

1Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599, USA; 2ErasmusAGE and Department of Epidemiology, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands; 3Netherlands Consortium for Healthy Aging, Netherlands Genomics Initiative, P.O. Box 9600, 2300 RC Leiden, the Netherlands; 4Department of Epidemiology, University Medical Center Groningen, University of Groningen, P.O. Box 30006, 9700 RB Groningen, the Netherlands; 5Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108, USA; 6Department of Internal Medicine II—Cardiology, University of Ulm Medical Center, Albert-Einstein-Allee 23, 89081 Ulm, Germany; 7Department of Laboratory Medicine and Pathology, University of Minnesota, 420 Delaware Street SE, Minneapolis, MN 55455, USA; 8Division of Epidemiology and Community Health, University of Minnesota, West Bank Office Building, 1300 S Second Street, Suite 300, Minneapolis, MN 55454-1015 MN, USA; 9National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 31 Center Drive, MSC 2560, Bethesda, MD 20892-2560, USA; 10Department of Biostatistical Sciences, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157-1063, USA; 11Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany; 12Department of Medical Biometry, University of Split School of Medicine, Soltanska 2, 21000 Split, Croatia; 13Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK; 14Department of Biostatistics, Boston University School of Public Health, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA; 15Laboratory for Statistical Analysis, Center for Genomic Medicine (CGM), RIKEN, 1-7-22 Suehiro-cho, Tsukiji-ku, Yokohama, Kanagawa 220-0045, Japan; 16Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; 17Department of Laboratory Medicine, University of Groningen, P.O. Box 30001, 9700 BB Groningen, the Netherlands; 18Department of Laboratory Medicine, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9700 RB Groningen, the Netherlands; 19Department of Public Health, University of Split School of Medicine, Soltanska 2, 21000 Split, Croatia; 20Department of Biostatistics and Computational Biology, University of North Carolina, Chapel Hill, NC 27599, USA; 21Center for Biomedicine, European Academy Bozen/Bolzano (EURAC, affiliated institute of the University of Lübeck, Ratzeburger Allee 160, Viale Druso, 1 / Drususallee 1, 39100 Bolzano/Bozen, Italy; 22Department of Medical Biometry, University of Split School of Medicine, 801 Massachusetts Avenue, Cambridge, MA 02139, USA; 23Center for Biomedical, European Academy Bozen/Bolzano (EURAC, affiliated institute of the University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; 24Department of Laboratory Medicine, University of Minnesota, 420 Delaware Street SE, Minneapolis, MN 55455, USA; 25Centre of Preventive Medicine and Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK; 26Department of Psychology, The University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK; 27Institute of Medical Biometry and Statistics, University Hospital Schleswig-Holstein, Ratzeburger Allee 160, Haus 4, 23538 Lübeck, Germany; 28Faculty of Medicine, University of Iceland, Menntavegi 1, 101 Reykjavík, Iceland; 29Laboratory for Molecular Genetics, National Institute of Health Sciences and Institute of Molecular Medicine, Medical School of Zagreb, 10 000 ZAGREB, Salata 3, Croatia; 30University Medical Center Groningen, University of Groningen, P.O. Box 30006, 9700 RB Groningen, the Netherlands; 31Geriatric Unit, Azienda Sanitaria di Firenze, 50125 Florence, Italy; 32Swiss Institute of Bioinformatics, Quarter Sorge, Batiment Genopole, 1015 Lausanne, Switzerland; 33Department of Pharmacology, University of Split School of Medicine, Soltanska 2, 21000 Split, Croatia; 34Community Prevention Unit, University Institute of Social and Preventive Medicine, Rue du Bugnon 17, 1005 Lausanne, Switzerland; 35Department of Medical Biology, University of Split School of Medicine, 801 Massachusetts Avenue, Suite 470, Boston, MA 02118, USA; 36Department of Laboratory Medicine, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9700 BB Groningen, the Netherlands; 37Department of Laboratory Medicine, University Medical Center Groningen, University of Groningen, P.O. Box 30006, 9700 RB Groningen, the Netherlands; 38Laboratory for Clinical Biochemistry Research, Department of Pathology, University of Vermont College of Medicine, 89 Beaumont Avenue, Courtyard at Given S269, Burlington, VT 05405, USA; 39Laboratory of Epidemiology, Demography, and Biometry, National Institutes of Aging, Gateway Building, 7201 Wisconsin Avenue, Bethesda, MD 20892-9205, USA; 40Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, 3001 Hanover Street, Baltimore, MD 21225, USA; 41Laboratory for Genotyping Development, CGM, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan; 42Institute for Clinical Chemistry and Laboratory Medicine, University Medical Center Mainz, Langenbeckstraße 1, 55131 Mainz, Germany; 43Laboratory of Epidemiology, Demography, and Biometry, National Institutes of Aging, Gateway Building, 7201 Wisconsin Avenue, Bethesda, MD 20892-9205, USA; 44Cardiovascular Health Research Unit, University of Washington, Metropolitan Park East Tower, 1730 Minor
Many disorders are associated with altered serum protein concentrations, including malnutrition, cancer, and cardiovascular, kidney, and inflammatory diseases. Although these protein concentrations are highly heritable, relatively little is known about their underlying genetic determinants. Through transethnic meta-analysis of European-ancestry and Japanese genome-wide association studies, we identified six loci at genome-wide significance ($p < 5 \times 10^{-8}$) for serum albumin (HPN-SCN1B, GCKR-FNDC4, SERPINF2-WDR81, TNFRSF11A-ZCCHC2, FRMD5-DWR76, and RPS11-FCGRT, in up to 53,190 European-ancestry and 9,380 Japanese individuals) and three loci for total protein (TNFRS13B, 6q21.3, and ELL2, in up to 25,539 European-ancestry and 10,168 Japanese individuals). We observed little evidence of heterogeneity in allele effects at these loci between groups of European and Japanese ancestry but obtained substantial improvements in the resolution of fine mapping of potential causal variants by leveraging transethnic differences in the distribution of linkage disequilibrium. We demonstrated a functional role for the most strongly associated serum albumin locus, HPN, for which Hpn knockout mice manifest low plasma albumin concentrations. Other loci associated with serum albumin harbor genes related to ribosome function, protein translation, and proteasomal degradation, whereas those associated with serum total protein include genes related to immune function. Our results highlight the advantages of transethnic meta-analysis for the discovery and fine mapping of complex trait loci and have provided initial insights into the underlying genetic architecture of serum protein concentrations and their association with human disease.
the European meta-analysis with data from a GWAS of Japanese ancestry with the aim of (1) identifying additional loci through increased sample size, (2) assessing the evidence of heterogeneity in allelic effects between ethnic groups, and (3) improving the resolution of fine mapping in associated regions by leveraging the expected differences in the structure of linkage disequilibrium (LD) between diverse populations.

The European-ancestry meta-analysis consisted of 53,190 individuals (from 20 GWASs) for serum albumin and 25,539 individuals (from six GWASs) for total protein (Tables S1–S3 and Supplemental Data available online). The procedures followed were approved by the institutional review board committees and are in accordance with the ethical standards of the institutional committees on human experimentation. All participants have given informed consent. Sample and SNP quality control (QC) were undertaken within each study. Each GWAS was then imputed at up to 2.5M autosomal SNPs with the use of CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection) samples from phase II of the International HapMap Project. Each SNP with minor allele frequency (MAF) >1% that passed QC was tested for association with serum albumin and total protein under an additive model after adjustment for study-specific covariates. The results of each GWAS were corrected for residual population structure using the genomic control inflation factor and were combined via fixed-effect inverse-variance-weighted meta-analysis. The results of the meta-analysis were subsequently corrected by a second round of genomic control ($\lambda_{GC} = 1.04$ for serum albumin and $\lambda_{GC} = 1.02$ for total protein) to allow for population differences between studies.

The European-ancestry meta-analysis identified six genome-wide significant loci ($p < 5 \times 10^{-8}$) for serum albumin and two for total protein (Table 1, Figure S1). These included association signals for serum albumin at *HPN* (MIM 142440)-*SCN1B* (MIM 600235) ($p = 3.3 \times 10^{-15}$), *SERPINF2* (MIM 613168)-*WDR81* (MIM 614218) ($p = 6.8 \times 10^{-13}$), *TNFRSF11A* (MIM 603499)-*ZCCHC2* ($p = 3.9 \times 10^{-9}$), and *FRMD5-WDR76* for total protein on chromosome 6q21.3 ($p = 3.4 \times 10^{-9}$). We also confirmed the associations previously reported in eastern Asian populations at *GCKR-FNDC4* ($p = 2.9 \times 10^{-14}$) and *RPS11-FCGRT* ($p = 3.2 \times 10^{-9}$) for serum albumin and *TNFRSF13B* ($p = 1.3 \times 10^{-10}$) for total protein. Inspection of the *HPN-SCN1B* locus (Figure 1) provided evidence for two independent association signals for serum albumin (rs4806073, $p = 3.3 \times 10^{-15}$; rs11671010, $p = 1.9 \times 10^{-13}$; CEU $\hat{r}^2 = 0.02, 4.3$ kb apart).

To perform conditional analyses at this locus, we applied genome-wide complex trait analysis (GCTA) to the results of the European-ancestry meta-analysis and individual-level genotype data from the Atherosclerosis Risk in Communities Study (8,127 European American individuals, Table S1) and confirmed that both SNPs remained genome-wide significant after adjustment for the effect of the other (rs4806073, $p = 1.6 \times 10^{-12}$; rs11671010, $p = 1.5 \times 10^{-11}$).

The results of the European meta-analysis were then combined with a Japanese-ancestry GWAS (BioBank Japan Project) consisting of 9,380 individuals for serum albumin and 10,168 individuals for total protein (Tables S1–S3, Supplemental Data). Sample and SNP QC were undertaken within the BioBank Japan Project. The GWAS was imputed at up to 2.5 M autosomal SNPs with the use of Han Chinese in Beijing and Japanese in Tokyo (CHB+JPT) samples from phase II of the International HapMap Project. Each SNP with MAF >1% that passed QC was then tested for association with serum albumin and total protein in the BioBank Japan Project under an additive model after adjustment for age and sex, and the results were corrected for residual population structure with the genomic control inflation factor ($\lambda_{GC} = 0.98$ for serum albumin and $\lambda_{GC} = 1.08$ for total protein).
The two panels present signal plots from the fixed-effect meta-analyses of European-ancestry individuals showing evidence for two independent associations in the region only: (A) SNPs tagged by rs4806073 and (B) SNPs tagged by rs11671010. In each panel, the lead SNP is represented by the purple circle. Each point represents a SNP plotted with their p (on a log_{10} scale) as a function of genomic position (build 36). The color coding of all other SNPs indicates LD with the lead SNP (estimated by CEU r^2 from phase II HapMap): red, r^2 ≥ 0.8; gold, 0.6 ≤ r^2 < 0.8; green, 0.4 ≤ r^2 < 0.6; cyan, 0.2 ≤ r^2 < 0.4; blue, r^2 < 0.2; and gray, r^2 unknown. Recombination rates are estimated from the International HapMap Project, and gene annotations are taken from the University of California Santa Cruz genome browser.

Figure 1. Signal Plot of the HPN-SCN1B Locus for Serum Albumin
The two panels present signal plots from the fixed-effect meta-analyses of European-ancestry individuals showing evidence for two independent associations in the region only: (A) SNPs tagged by rs4806073 and (B) SNPs tagged by rs11671010. In each panel, the lead SNP is represented by the purple circle. Each point represents a SNP plotted with their p (on a log_{10} scale) as a function of genomic position (build 36). The color coding of all other SNPs indicates LD with the lead SNP (estimated by CEU r^2 from phase II HapMap): red, r^2 ≥ 0.8; gold, 0.6 ≤ r^2 < 0.8; green, 0.4 ≤ r^2 < 0.6; cyan, 0.2 ≤ r^2 < 0.4; blue, r^2 < 0.2; and gray, r^2 unknown. Recombination rates are estimated from the International HapMap Project, and gene annotations are taken from the University of California Santa Cruz genome browser.
clusters according to a Bayesian partition model of similarity in terms of their allele frequency profile. Studies assigned to the same cluster have the same allelic effect. However, each cluster can have different allelic effects. Fixed-effect meta-analysis is thus equivalent to a Bayesian partition model with a single cluster of studies.

We observed strong evidence of association, as defined by a log_{10} Bayes factor (BF) of 5 (equivalent to prior odds of association of any SNP with either trait of 1:100,000), at all identified loci for both traits (Table 2, Figure S2). These loci included RPS11-FCGRT for total protein, which was previously observed in eastern Asian populations but not at genome-wide significance in our European-ancestry meta-analysis. The only exception was at the FRMD5-WDR76 locus (log_{10}BF = 4.79), where the lead SNP from the European meta-analysis (rs16948098) was not observed in the Japanese GWAS and is monomorphic in eastern Asian (CHB and JPT) HapMap populations. Using the threshold of log_{10}BF > 5 for strong evidence of association, we identified two additional “potential” loci for serum albumin and four for total protein (Table 2).

MANTRA revealed little evidence of heterogeneity in allelic effects between European-ancestry and Japanese studies at the majority of the serum albumin and total protein loci (Table 2). The extent of heterogeneity was assessed through comparison of association BF under a Bayesian partition model wherein the number of clusters of studies is unrestricted to that wherein there is a single cluster, the latter corresponding to homogeneous allelic effects across all ancestry groups. Subsequent fixed-effect inverse-variance-weighted meta-analysis across groups of European and Japanese ancestry (Table S4) revealed one additional signal for total-protein mapping to ELL2 (MIM 601874, p = 1.1 × 10^{-8}), although none of the other potential MANTRA loci showed genome-wide significance (p < 5 × 10^{-8}). Among these potential loci, however, there was strong evidence of heterogeneity at ARID5B (MIM 608538) for total protein (MANTRA log_{10}BF in favor of heterogeneity of 6.79). The lead SNP at this locus was previously observed in eastern Asian populations but not at genome-wide significance in our European-ancestry GWASs, and there is little evidence of variation in LD structure between CEU and CHB + JPT reference haplotypes from the 1000 Genomes Project (Figure S3). Although intrastudy phenotypic variation in total protein concentrations (such as Ig, which is not available for analyses here) might contribute to these apparent transethnic differences in allelic effects, further investigation is required to fully elucidate the source of heterogeneity between ancestry groups.

To assess the improvement in fine-mapping resolution due to transethnic meta-analysis in serum albumin and total protein loci, we defined “credible sets” of SNPs (J.B. Maller, personal communication) with the strongest signals of association and, hence, most likely to be causal (or tagging an unobserved causal variant), on the basis of European-ancestry GWASs only and then after inclusion of the Japanese study. At each locus, defined by the genomic region 500 kb up and downstream of the lead SNP, we calculated the posterior probability that the jth SNP is “causal” (or tags an unobserved causal variant) by

\[
\phi_j = \frac{BF_j}{\sum_k BF_k}
\]

In this expression, BF_j denotes the BF in favor of association of the jth SNP from the transethnic MANTRA analysis, and the summation in the denominator is over all SNPs passing QC across the locus (J.B. Maller, personal communication). A 100% credible set at the locus was then constructed through (1) ranking all SNPs according to their BF and (2) combining ranked SNPs until their cumulative posterior probability exceeded 0.9. Using this definition, we observed improved resolution, in terms of the number of SNPs and the genomic interval covered by the credible set, at HPN-SCN1B, TNFRSF11A-ZCCHC2, and RPS11-FCGRT for serum albumin and at TNFRSF13B and the 6q21.3 locus for total protein (Figure 2, Table S5). The most striking improvements in resolution were observed at the 6q21.3 locus for total protein, wherein the 99% credible set was reduced from 14 SNPs (covering 346 kb), to just three (covering 37 kb). Furthermore, after transethnic meta-analysis, the posterior probability that the lead SNP was causal (or tagged an unobserved causal variant) was more than 95% at GCKR-FNDC4 and SERPINF2-WDR81 for serum albumin and at TNFRSF13B and the 6q21.3 locus for total protein.

Two of the serum albumin loci, HPN-SCN1B and RPS11-FCGRT, can be validated by existing mouse models. The lead SNP at the HPN-SCN1B locus maps to an intron of HPN, a gene encoding hepsin, a membrane-bound serine protease that has substrate specificity for basic amino acids similar to that of proalbumin processing, suggesting a physiologic role of hepsin in the cleavage of proalbumin to albumin. We compared serum protein concentrations between hepsin knockout (KO) mice and wild-type littermates (Figure 3) using blood samples collected from the inferior vena cava and analyzed by Consolidated Veterinary Diagnostics (West Sacramento, CA, USA). In hepsin^{-/-} mice, we observed overwhelming evidence of reduced serum albumin (p = 9.1 × 10^{-12}) and, to a lesser extent, reduced total protein (p = 1.5 × 10^{-5}), but not Ig. At the RPS11-FCGRT locus, KO Fcgrt mice have been previously demonstrated to manifest low serum albumin and low serum gamma Ig concentrations. Furthermore, in humans, two siblings with genetic deficiency of FcRn due to lack of the β2 microglobulin component have manifested reduced serum albumin and gamma Ig concentrations. FCGRT encodes the heavy alpha chain of the FcRn, which prevents lysosomal degradation of...
Table 2. Loci with Strong Evidence of Association with Serum Albumin and Total Protein after Transethnic MANTRA Analysis of European-Ancestry and Japanese GWASs

<table>
<thead>
<tr>
<th>Lead SNP</th>
<th>Position (Build 36)</th>
<th>Alleles</th>
<th>European Ancestry GWAS Meta-analysis</th>
<th>Japanese GWAS</th>
<th>MANTRA Transethnic Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>EAF</td>
<td>Beta</td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sample Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Serum Albumin Established Loci</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1260326</td>
<td>2</td>
<td>27,584,444</td>
<td>T</td>
<td>0.41</td>
<td>0.0124</td>
</tr>
<tr>
<td>rs4806073</td>
<td>19</td>
<td>40,247,030</td>
<td>C</td>
<td>0.93</td>
<td>0.0257</td>
</tr>
<tr>
<td>rs11078597</td>
<td>17</td>
<td>1,655,113</td>
<td>C</td>
<td>0.18</td>
<td>0.0205</td>
</tr>
<tr>
<td>rs694419</td>
<td>18</td>
<td>58,277,092</td>
<td>T</td>
<td>0.52</td>
<td>0.0093</td>
</tr>
<tr>
<td>rs2280401</td>
<td>19</td>
<td>54,691,821</td>
<td>A</td>
<td>0.17</td>
<td>0.0121</td>
</tr>
<tr>
<td>rs16948098</td>
<td>15</td>
<td>42,006,899</td>
<td>A</td>
<td>0.06</td>
<td>0.0229</td>
</tr>
<tr>
<td><strong>Serum Albumin Additional Potential Loci</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2293579</td>
<td>11</td>
<td>47,397,334</td>
<td>A</td>
<td>0.40</td>
<td>0.0093</td>
</tr>
<tr>
<td>rs12914385</td>
<td>15</td>
<td>76,685,778</td>
<td>C</td>
<td>0.61</td>
<td>0.0064</td>
</tr>
<tr>
<td><strong>Total Protein Established Loci</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4561508</td>
<td>17</td>
<td>16,789,475</td>
<td>T</td>
<td>0.11</td>
<td>0.0360</td>
</tr>
<tr>
<td>rs204999</td>
<td>6</td>
<td>32,217,957</td>
<td>A</td>
<td>0.74</td>
<td>0.0250</td>
</tr>
<tr>
<td>rs2280401</td>
<td>19</td>
<td>54,691,821</td>
<td>A</td>
<td>0.16</td>
<td>0.0120</td>
</tr>
<tr>
<td><strong>Total Protein Additional Potential Loci</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3777200</td>
<td>5</td>
<td>95,260,547</td>
<td>T</td>
<td>0.27</td>
<td>0.0180</td>
</tr>
<tr>
<td>rs1260326</td>
<td>2</td>
<td>27,584,444</td>
<td>T</td>
<td>0.44</td>
<td>0.0150</td>
</tr>
<tr>
<td>rs2675609</td>
<td>10</td>
<td>63,306,537</td>
<td>T</td>
<td>0.39</td>
<td>0.0090</td>
</tr>
<tr>
<td>rs10097731</td>
<td>8</td>
<td>82,190,227</td>
<td>G</td>
<td>0.15</td>
<td>0.0230</td>
</tr>
</tbody>
</table>

Strong evidence is defined as log10 BF > 5. The following abbreviations are used: Chr, chromosome; GWAS, genome-wide association study; MANTRA, meta-analysis of transethnic association studies; EAF, effect allele frequency; and BF, Bayes factor.
Figure 2. Fine Mapping of the 6q21.3 Locus for Total Protein
The two panels present signal plots for the MANTRA association signal (A) after transethnic meta-analysis of European-ancestry and Japanese GWASs and (B) after meta-analysis of the European ancestry GWAS only. Each point represents a SNP passing QC in our MANTRA analysis, plotted with their BF on a log10 scale as a function of genomic position (build 36). In each panel, the lead SNP is represented by the purple circle. The color coding of all other SNPs indicates LD with the lead SNP (estimated by CEU $r^2$ from phase...
albumin and Ig in lysosomes and thereby extends their serum half-life.\textsuperscript{17}

To gain insights into the possible functional role of other serum albumin and total protein loci, we began by performing expression quantitative trait locus (eQTL) mapping using data derived from 1,469 whole blood samples.\textsuperscript{20} Gene expression levels were measured from peripheral blood and assayed in 1,240 individuals with Illumina HT12 v3 and in 229 individuals with Illumina H8 v2 BeadChip arrays. Both sets were independently quantile-normalized after log\textsubscript{2} transformation and subsequently corrected for 50 principal components obtained from the gene expression probe covariance matrix. To integrate both data sets, genotype data were imputed up to 2.5M autosomal SNPs with the use of CEU samples from phase II of the International HapMap Project.\textsuperscript{10} SNPs of low frequency (MAF < 5\%) or with deviation from Hardy-Weinberg equilibrium (p < 10\textsuperscript{-4}) were excluded from the subsequent analysis. Cis-eQTL effects (within 1 Mb of the probe) were determined with Spearman’s ranked correlation, and meta-analysis between the two data sets was performed with the use of weighted Z scores. The false discovery rate (FDR) was then assessed by permutation. Using this approach, we mapped lead SNPs at four of the identified loci (GCKR-FNDC4, SERPINF2-WDR81, and RPS11-FCGRT) for serum albumin; 6q21.3 for total protein) to cis expression levels of 18 genes (Table S6). The strongest associations were observed for expression of NOSIP (p = 2.4 \times 10^{-17} at RPS11-FCGRT and HLA-DQA1 (MIM 146880)/HLA-DQA2 (MIM 613503, p = 9.1 \times 10^{-30} at 6q21.3). NOSIP (nitric oxide synthase interacting protein) inhibits endothelial nitric oxide synthesis, whereas HLA-DQA1/2 is a human leukocyte antigen (HLA) class II antigen with an immune system role related to processing and presentation of antigen peptides.

We noticed that the lead SNP at the GCKR-FNDC4 locus (rs1260326, c.1337T>C [p. Leu446Pro]; RefSeq NM_001486.3) is a missense mutation with moderate predicted functional impact by snpEff, and has been previously associated with several metabolic traits, as well as kidney, liver, and hematologic phenotypes (Table S7). We used data from the 1000 Genomes Project\textsuperscript{15} to search for additional coding variants with predicted function in strong LD (\rho^2 > 0.5 in 283 individuals of European ancestry) with lead SNPs at our identified serum albumin and total protein loci (Table S8). Across serum albumin loci, two nonsynonymous SNPs mapped to SERPINF2, a gene encoding an inhibitor of plasmin which degrades plasma fibrin and other proteins, and one nonsynonymous SNP mapped to CHRNA5 (MIM 118505), a nicotinic acetylcholine receptor gene associated with smoking behavior\textsuperscript{21,22} and lung cancer.\textsuperscript{23} For total protein, nonsynonymous SNPs mapped to TNFRSF13B and ELL2, and to PPT2 (MIM 603298) and EGFL8 (MIM 609897) at the 6q21.3 locus. Mutations in TNFRSF13B cause immunodeficiency common variable type 2 (MIM 240500), characterized by hypogammaglobulinemia and recurrent bacterial infections due to failure of B cell differentiation and impaired production of Ig. They also cause selective IgA deficiency 2 (MIM 609529), the most common primary immunodeficiency, affecting 1 in 600 individuals in the western world. ELL2 product directs Ig secretion in plasma cells, and the 6q21.3 major histocompatibility complex class III region encompasses a number of genes involved in autoimmunity, inflammation, and complement proteins. Interrogation of the National Human Genome Research Institute (NHGRI) GWAS catalog\textsuperscript{24} highlighted that lead SNPs at ten of the identified loci have themselves been reported or are in LD (\rho^2 > 0.5 in 283 individuals of European ancestry) with those disclosed, for a diverse range of human complex traits (Table S7) but are enriched for metabolic phenotypes that are associated with, or are direct products of, protein metabolism.

Finally, we used the human interactome database (Cytoscape) to construct an interaction network consisting of...
250 proteins that directly interact with genes in the identified serum albumin and total protein loci reported in Table 2. For identifying molecular complexes within this first-degree interaction network, cluster analyses were performed with the FAG-EC algorithm, implemented in the ClusterViz plug-in, with standard settings applied. In total, 16 distinct clusters were identified, including three large complexes (Figure S4) that were carried forward for further analysis. Functional enrichment analyses within these clusters were performed following defined pathways from BioCarta, KEGG, PANTHER, and Reactome via the Database for Annotation, Visualization, and Integrated Discovery (DAVID). The most significantly enriched clusters from protein interaction network analyses incorporated ribosomal functioning and protein translation, proteasomal protein degradation, and immune-response signaling (Table S9). As a complementary approach, we applied an implementation of gene-set enrichment analysis (MAGENTA25) to identify whether defined biological pathways from BioCarta, Gene Ontology, Ingenuity, KEGG, PANTHER, and Reactome were enriched in the leading-edge fraction of the meta-analysis. In brief, gene association p values were calculated on the basis of the meta-analysis summary statistics for SNPs within a 110-kb-upstream and 40-kb-downstream window. These gene scores were then corrected for gene size, number of SNPs, and the LD between them, and subsequently ranked by p value. Enrichment in the 75th and 95th percentiles was assessed for significance by comparison with 10,000 randomly generated pathways. Using an FDR threshold of 5%, we observed significant overrepresentation of genes assigned to three pathways: RNA- (FDR = 0.044), sensory-perception- (FDR = 0.027), and protein-trafficking-related pathways (FDR = 0.043-0.044) (Table S10).

In conclusion, we have identified six loci for serum albumin concentration and three for total protein at genome-wide significance. These loci harbor genes that fall across a diverse range of biological pathways, including those involved in biomarkers, immune regulation, and disease, but are enriched for those relevant to the synthetic and degradation of serum protein. By combining GWAS data from European and Japanese populations, we observed some evidence of heterogeneity in allelic effects between ancestry groups and have demonstrated substantial improvements in the localization of potential causal variants. Taken together, our results highlight the advantages of transethnic meta-analysis for the discovery and fine mapping of complex trait loci and provide initial insights into the underlying genetic architecture of serum protein concentrations and their association with human disease.

Supplemental Data

Supplemental Data include four figures, ten tables, Supplemental Acknowledgments, study characteristics, samples, and assays and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

A.P.M acknowledges financial support from the Wellcome Trust, grant numbers WT081682, WT064890, WT098017, and WT090532. N.F. acknowledges financial support from AHA 0675001N, R01HL089651, and U01HG004803.

Received: May 11, 2012
Revised: June 18, 2012
Accepted: August 23, 2012
Published online: September 27, 2012

Web Resources

The URLs for data presented herein are as follows:
1000 Genomes Project, http://www.1000genomes.org/
GCTA, http://www.complextraitgenomics/software/gcta/
NHGRI Catalog of Published Genome-wide Association Studies, http://www.genome.gov/gwasstudies/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
SNIPPER, http://csg.sph.umich.edu/boehnke/snipper/

References


