LPAR1 and *ITGA4* regulate peripheral blood monocyte counts

SUPPORTING INFORMATION

Narelle Maugeri, Joseph Powell, Peter A.C. 't Hoen, Eco J. C. de Geus, Gonneke Willemsen, Mathijs Kattenberg, Anjali K. Henders, Leanne Wallace, Brenda Penninx, Jouke-Jan Hottenga, Sarah E. Medland, Viatcheslav Saviouk, Nicholas G. Martin, Peter M. Visscher, Gert-Jan B. van Ommen, Ian H. Frazer, Dorret I. Boomsma, Grant W. Montgomery, Manuel A.R. Ferreira

SUPP. METHODS

Gene-expression experiments

We used two approaches to quantify the expression levels of genes located near two SNPs shown to associate with variation in monocyte counts previously (rs7023923, Ferreira et al. 2009) or in the present study (rs6740847). The first approach quantified genome-wide RNA expression levels using Illumina expression arrays. Briefly, whole blood was collected directly into a PAXgene^{TM} tube (QIAGEN) for each twin of 48 pairs of monozygotic (MZ) twins (26 female and 22 male pairs, mean age 14) ascertained from the general population (Aitken et al. 1996; McGregor et al. 1999). Total RNA was extracted using the PBMC gene RNA purification kit, converted to cDNA, amplified and purified. Expression profiles were then generated by hybridising 750 ng of cDNA to Illumina HumanHT-12 v3.0 Beadchip according to the Illumina assay guide. Relative expression values were generated for each transcript using Illumina Genome Studio software and standardised as follows. First, background noise detected from negative control beads was subtracted from raw expression values for each transcript. Data were then filtered for transcripts which were present across 100% of samples at $p < 0.05$ according to the global-error threshold calculated by Genome Studio's cross-gene error model. Adjusted expression levels for each transcript were transformed using a quantile transformation to achieve a stabilized variance distribution across average expression levels. Further normalisation was obtained using a mixed linear model to account for differences in the mean expression level between individuals and variation in the data due to chip, gene and scanner differences. The residuals from this model were used in all further analyses. This conservative procedure results in normalised expression phenotypes that are comparable between individuals and across transcripts. To reduce the amount of variation in transcript levels due to non-genetic factors, we

calculated for each transcript the mean expression level for each MZ twin pair (ie. average across both twins) and used this for subsequent analyses.

We also used DeepSAGE analysis to independently validate in additional samples the gene-expression results from the Illumina array experiments. Briefly, RNA was isolated from total blood of 94 healthy controls (28 males, 66 females; mean age 44, range 19-64) using PAXgene™ tubes. DeepSAGE on the Illumina GAII was then used to measure gene expression levels as described in detail elsewhere (Hestand et al. 2010). The average number of SAGE tags with unique hits to the Human Reference Genome Build 19 was $15x10^6$ (range $4.1x10^6$ – $33x10^6$). Gene expression levels were then calculated by summing all SAGE tags in annotated transcripts for each gene. Next, after removing highly abundant hemoglobin transcripts, gene expression levels were scaled across samples using a robust normalization procedure implemented in edgeR (Robinson et al 2010) and a square root transformation applied to stabilize variance (Freeman and Tukey 1950).

Follow-up of ten loci previously suggested to associate with monocyte counts

The ten loci that showed strong but not experiment-wide level of association with monocyte counts in our previous study (Ferreira et al. 2009) were followed-up in 1,122 individuals randomly drawn from the NTR-Biobank study (Willemsen et al 2010). Participants (mean age 49, SD = 14; 60.8% female) gave informed consent to participate and all studies were approved by appropriate ethics committees. Biological samples were taken at the respondents' home between 0700 and 1000 h. Full blood counts were then performed within 5h of blood collection using a Coulter instrument.

Genome-wide SNP-genotyping was performed by hybridizing 200 ng of DNA to the Illumina™ Human660W-Quad DNA Analysis BeadChips, containing 657,366 markers per

sample. Experimental procedure was performed according to the Infinium® HD Assay Super manual from Illumina (Illumina, San Diego, USA). Allele calls were determined using Illumina BeadStudio (Illumina, San Diego, USA). Complete genotyping procedure was performed in the Genomics platform (certified service provider (CSPro(R)) for Illumina Inc.) at the LIFE & BRAIN Center Bonn. Quality control excluded SNPs based on MAF < 0.01, missing genotype rate > 0.05 or a p-value < 1 x 10-5 in a test of Hardy-Weinberg equilibrium. After quality control, 515,781 SNPs were left (78%). Samples were excluded if they showed evidence for contamination by excessive allele sharing with multiple samples and excessive levels of heterozygosity (F < -0.10). Samples were also excluded if they had a higher than 90% genotype missing rate. Subsequently, genotypes of ~3.8 million SNPS were imputed with the IMPUTE program (Howie et al. 2009), using the HapMap CEU data (release 24, NCBI build 36), available from the IMPUTE website, as reference. Imputed SNPs were excluded if they had a minor allele frequency < 0.01 or a properinfo < 0.3, leaving 2,147,160 autosomal SNPs for analysis.

Individual SNPs were tested for association with monocyte counts using linear regression in PLINK (Purcell et al. 2007), after adjusting for sex, age and applying an inverse normal transformation. To estimate the proportion of variation in monocyte counts jointly explained by rs7023923 and rs6740847 we used a linear regression of monocyte counts on rs7023923 and rs6740847 allelic dosage, i.e. MONO ~ rs7023923_add + rs6740847_add. The regression r^2 corresponds to the variance in monocyte counts explained by this model. The combined analysis of the original GWAS and replication panels (*N*=5,347) for rs6740847 was performed by merging the actual genotype data from the two studies and applying a linear regression as described above.

SUPP. TABLES

Supp. Table S1. Location and DNA sequence of the Illumina probes analysed in the gene-expression array experiment

PROBE_ID	GENE	POSITION(hg18)	PROBE_SEQUENCE
ILMN_1701441	LPAR1	chr9:112675937-112675986	CTTTGTTTGAGGGACTCGTTATCCAGCTCTTGGTAGCCACACCTGCAATG
ILMN_1745801	OR2K2	chr9:113129999-113130048	GCCCTGCAGATACCCCTCTGTGGGAATCTCATCGATCACTTCACGTGTGA
ILMN_1847822	KIAA0368	chr9:113168382-113168625	GGGGTTAATGCTTTTGGAAGAAGAACATGCCGATCCTGAGGCTTTGGCTG
ILMN_1747052	ITGA4	chr2:182108850-182108899	TACCATATGTGCTTGCCTCAGTAAAATGAACCCCACTGGGTGGGCAGAGG
ILMN_1801091	CERKL	chr2:182110830-182110879	GGGGAGGAACTCAGTAATGCTCCATTGCTCTAACCTCCTGCTGTGTGGCC
ILMN_1740837	CERKL	chr2:182138415-182138464	GTTCGCAATACATTGGAGTGTACAGAGATTGTTCACTGGCATGCAAACGT
ILMN_2302692	CERKL	chr2:182111194-182112090	GGAAGTTGCATCAGAGGTCCATATTAGATTGCATCCAAGACTTATCAGTC
ILMN_1808810	CERKL	chr2:182117766-182120636	CCTTGCCTGTGTTCAGTGGCACCTAGAGGCTTGGCACCTAATACCAGATT

Supp. Table S2. Association results reported by Zeller et al. (2010) between the expression levels of *LPAR1* measured in monocytes extracted from 1,490 individuals and rs2150052, a SNP in strong LD (r^2 = 0.98) with rs7023623

rs2150052 genotype[a]	N	Mean Expression	SD	h^2(%)[b]	*P*-value
TT	383	5.34	0.37		
TA	744	5.81	0.30	44.6	$6.3x10^{-191}$
AA	359	6.14	0.28		

Abbreviations are as follows: N, sample size; SD, standard deviation.

[a] The rs2150052A allele that is associated with increased *LPAR1* expression in Zeller et al. (2010) is on the same haplotype as the rs7023623T allele that is associated with increased expression of *LPAR1* (cf. Table 1) and increased monocyte counts (Ferreira et al., 2010) in our study.

[b] h^2 represents the proportion of the between-individual variation in gene expression explained by the SNP.

Supp. Table S3. Replication results for ten loci previously reported to have strong ($P<10^{-5}$) but not experiment-wide ($P<5.5x10^{-9}$) significance in a recent GWAS of monocyte counts (Ferreira et al. 2009)

Chromosome, bp position	SNP, allele	Nearest gene, bp distance	Original GWAS			Replication panel[b]		
			Effect[a]	SE	*P*-value	Effect	SE	*P*-value
14,103087858	rs10141157,C	*BAG5*,4783	0.13	0.02	1.6E-08	-0.01	0.04	0.799
2,182016597	rs6740847,G	*ITGA4*,13266	0.13	0.02	3.6E-08	0.13	0.04	0.002
7,152402256	rs17173921,A	*ACTR3B*,218860	0.13	0.03	2.2E-06	0.02	0.05	0.643
16,84523155	rs524432*,T	*IRF8*,9443	-0.14	0.03	2.7E-06	-0.09	0.06	0.115
10,52236063	rs10821846,G	*A1CF*,267	0.11	0.02	2.9E-06	0.00	0.04	0.924
8,97408427	rs11782606*,G	*PTDSS1*,7523	-0.21	0.05	3.8E-06	0.01	0.09	0.952
7,25361712	rs1967297,T	*NPVF*,127082	-0.11	0.03	4.0E-06	-0.05	0.05	0.245
21,38113768	rs2835988,C	*KCNJ6*,96798	0.11	0.02	4.7E-06	-0.01	0.04	0.849
16,84537847	rs965773,A	*IRF8*,24135	0.11	0.02	7.4E-06	-0.04	0.04	0.412
6,31221031	rs3094225,A	*PSORS1C1*,5215	-0.11	0.03	7.4E-06	0.03	0.05	0.480

[a] Effect corresponds to standard deviation units for the transformed phenotype

[b] The replication panel consisted of 1,122 unrelated individuals ascertained from the general population as described above. All SNPs were directly genotyped or imputed with high confidence (properinfo score >0.9) using IMPUTE (Howie et al., 2009).

* SNPs rs524432 and rs11782606 were not available in the replication panel and so results are presented for two proxy (r^2=1) SNPs (rs411221 and rs11782664, respectively).

Supp. Table S4. Association between the expression levels of *ITGA4* measured in whole-blood by Illumina expression arrays[a] and rs6740847, a SNP originally identified through a recent GWAS of monocyte counts (Ferreira et al. 2009) and confirmed in the present study

Genotype	N	Mean Expression	SD	Effect[b]	SE	h^2(%)[c]	P-value
AA	11	69.6	32.3				
GA	24	60.3	24.9	-2.3	5.5	0.0	0.67
GG	13	64.5	25.0				

Abbreviations are as follows: N, sample size; SE, standard error.

[a] Expression of *ITGA4* was assessed by probe ILMN_1747052 that is present in the Illumina HumanHT-12 v3.0 Beadchip used.

[b] Effect corresponds to the slope from a linear regression of expression levels on the number of G alleles. Sex was not a significant predictor of the expression of the transcripts tested and so was not included as a covariate.

[c] h^2 represents the proportion of the between-individual variation in gene expression explained by the SNP.

Supp. Table S5. Association results reported by Zeller et al. (2010) between the expression levels of *ITGA4* measured in monocytes extracted from 1,490 individuals and rs2124440, a SNP in complete LD (r^2 = 1) with rs6740847

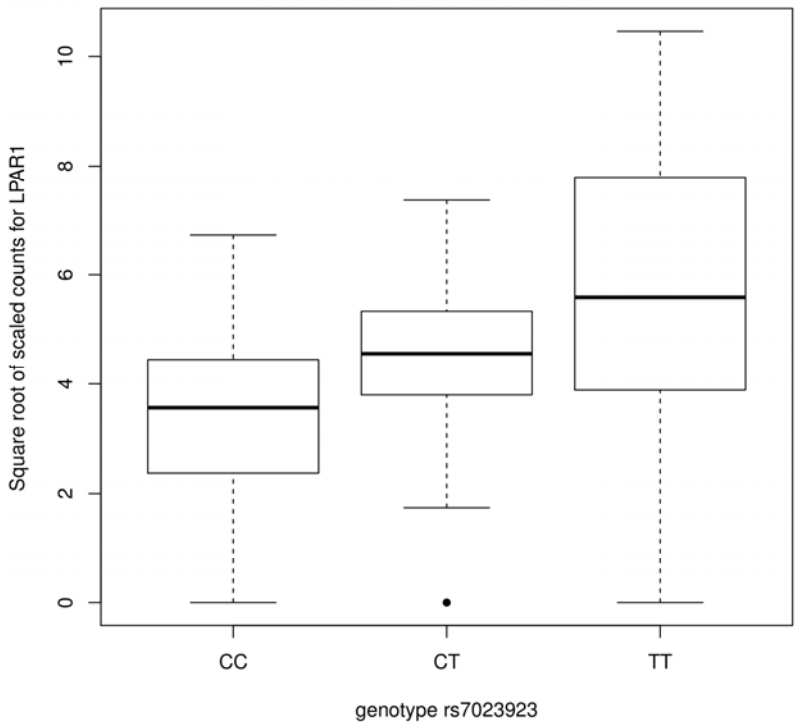
rs2124440 genotype[a]	N	Mean Expression	SD	h^2(%)[c]	*P*-value
GG	285	5.87	0.36		
GA	745	5.69	0.38	9.6	$2.2x10^{-33}$
AA	460	5.51	0.40		

Abbreviations are as follows: N, sample size; SD, standard deviation.

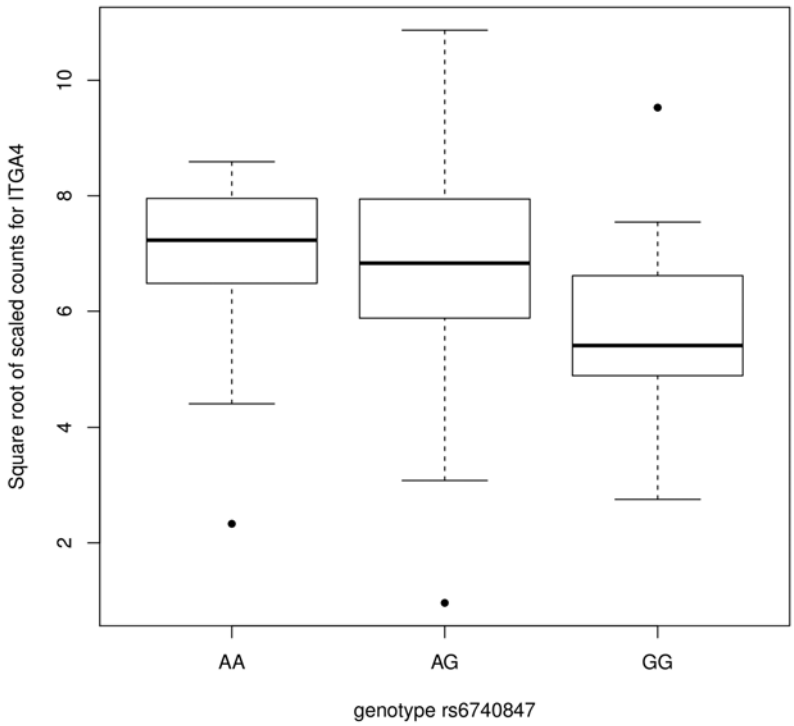
[a] The rs2124440A allele associated with decreased *ITGA4* expression in Zeller et al. (2010) is on the same haplotype as the rs6740847G allele that is associated with decreased *ITGA4* expression (cf. Supp. Figure S2) and increased monocyte counts (cf. Supp. Table S3) in our study.

[b] h^2 represents the proportion of the between-individual variation in gene expression explained by the SNP.

SUPP. FIGURES



Supp. Figure S1. Association between the expression levels of *LPAR1* measured in whole-blood with DeepSAGE and rs7023923 genotype, a SNP previously confirmed to associate with monocyte counts (Ferreira et al. 2009). Allelic dosage was significantly associated with *LPAR1* expression (P=5.4x10^{-5}), with the T allele associated with increased *LPAR1* expression. Details on the quantification methods used in the DeepSAGE experiments are provided in the Supp. Methods section above.



Supp. Figure S2. Association between the expression levels of *ITGA4* measured in whole-blood with DeepSAGE and rs6740847 genotype, a SNP confirmed in this study to associate with monocyte counts. Allelic dosage was significantly associated with *ITGA4* expression (P=0.02), with the G allele associated with decreased *ITGA4* expression. Details on the quantification methods used in the DeepSAGE experiments are provided in the Supp. Methods section above.

SUPP. REFERENCES

Aitken JF, Green AC, MacLennan R, Youl P, Martin NG..1996. The Queensland Familial Melanoma Project: study design and characteristics of participants. Melanoma Res 6:155-65.

Ferreira MA, Hottenga JJ, Warrington NM, Medland SE, Willemsen G, Lawrence RW, Gordon S, de Geus EJ, Henders AK, Smit JH, Campbell MJ, Wallace L, Evans DM, Wright MJ, Nyholt DR, James AL, Beilby JP, Penninx BW, Palmer LJ, Frazer IH, Montgomery GW, Martin NG, Boomsma DI. 2009. Sequence variants in three loci influence monocyte counts and erythrocyte volume. Am J Hum Genet 85:745-9.

Freeman MF and Tukey JW . 1950. Transformations related to the angular and the square root. Ann Math Statist 21:607-611.

Hestand MS, Klingenhoff A, Scherf M, Ariyurek Y, Ramos Y, van Workum W, Suzuki M, Werner T, van Ommen GJ, den Dunnen JT, Harbers M, 't Hoen PA .2010. Tissue-specific transcript annotation and expression profiling with complementary next-generation sequencing technologies. Nucleic Acids Res. 38(16):e165.

Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5(6):e1000529.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC .2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559-75.

Robinson MD, McCarthy DJ, Smyth GK . 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139-40.

Willemsen G, de Geus EJ, Bartels M, van Beijsterveldt CE, Brooks AI, Estourgie-van Burk GF, Fugman DA, Hoekstra C, Hottenga JJ, Kluft K, Meijer P, Montgomery GW, Rizzu P, Sondervan D, Smit AB, Spijker S, Suchiman HE, Tischfield JA, Lehner T, Slagboom PE, Boomsma DI. 2010. The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. Twin Res Hum Genet. 13:231-45.

Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, Maouche S, Germain M, Lackner K, Rossmann H, Eleftheriadis M, Sinning CR, Schnabel RB, Lubos E, Mennerich D, Rust W, Perret C, Proust C, Nicaud V, Loscalzo J, Hübner N, Tregouet D, Münzel T, Ziegler A, Tiret L, Blankenberg S, Cambien F. 2010. Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility. PLoS One 5(5): e10693.