Published in final edited form as:

Thromb Haemost. 2016 November 30; 116(6): 1041–1049. doi:10.1160/TH16-02-0151.

Genome-wide association reveals that common genetic variation in the kallikrein-kinin system is associated with serum L-arginine levels

Weihua Zhang^{1,2,*}, Fredrik Jernerén³, Benjamin C Lehne¹, Ming-Huei Chen⁴, Robert N Luben⁵, Carole Johnston³, Amany Elshorbagy^{3,6}, Ruben N Eppinga^{7,8}, William R Scott^{1,9}, Elizabeth Adeyeye^{2,10}, James Scott⁹, Rainer H. Böger¹¹, Kay-Tee Khaw⁵, Pim van der Harst^{7,8}, Nicholas J Wareham¹², Ramachandran S Vasan^{13,14}, John C Chambers^{1,2,15,*}, Helga Refsum^{3,16,*}, and Jaspal S Kooner^{2,9,15,*}

¹Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK ²Department of Cardiology, Ealing Hospital NHS Trust, Uxbridge Road, Southall, Middlesex UB1 3HW, UK ³Department of Pharmacology, University of Oxford, Oxford, UK ⁴Population Sciences Branch, National Heart, Lung, Blood Institute, National Institutes of Health, Framingham, MA 01702, USA ⁵Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK ⁶Department of Physiology, Faculty of Medicine, University of Alexandria, Alexandria, Egypt ⁷Department of Cardiology, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, the Netherlands ⁸Department of Genetics, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands ⁹National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus, London, UK ¹⁰Faculty of Medicine, Imperial College London, South Kensington Campus, London, SW7 2AZ ¹¹Department of Clinical Pharmacology and Toxicology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany ¹²Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK ¹³Section of Preventive Medicine and Epidemiology and Cardiology, Boston University School of Medicine, Boston, MA 02118, USA ¹⁴Boston University and National Heart, Lung & Blood Institute's Framingham Heart Study, Framingham, MA 01702, USA ¹⁵Imperial College Healthcare NHS Trust, London, UK ¹⁶Department of Nutrition, Institute of Basic Medical Science, Faculty of Medicine, University of Oslo, Oslo, Norway

Summary

L-arginine is the essential precursor of nitric oxide, and is involved in multiple key physiological processes, including vascular and immune function. The genetic regulation of blood L-arginine levels is largely unknown. We performed a genome-wide association study (GWAS) to identify genetic factors determining serum L-arginine levels, amongst 901 Europeans and 1,394 Indian Asians. We show that common genetic variations at the *KLKB1* and *F12* loci are strongly

Corresponding author: Weihua Zhang, Ph.D, Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK, Tel: +44 (0)20 8242 5926, Fax: +44 (0)20 8967 5007, weihua.zhang@imperial.ac.uk.

*Equal joint authorship

associated with serum L-arginine levels. The G allele of SNP rs71640036 (T/G) in *KLKB1* is associated with lower serum L-arginine concentrations (10 $\mu\text{mol/L}$ per allele copy, $p=1\times 10^{-24}$), while allele T of rs2545801 (T/C) near the *F12* gene is associated with lower serum L-arginine levels (7 $\mu\text{mol/L}$ per allele copy, $p=7\times 10^{-12}$). Together these two loci explain 7% of the total variance in serum L-arginine concentrations. The associations at both loci were replicated in independent cohorts with plasma L-arginine measurements ($p<0.004$). The two sentinel SNPs are in nearly complete LD with the nonsynonymous SNP rs3733402 at *KLKB1* and the 5'-UTR SNP rs1801020 at *F12*, respectively. SNPs at both loci are associated with blood pressure. Our findings provide new insight into the genetic regulation of L-arginine and its potential relationship with cardiovascular risk.

Keywords

Serum L-arginine concentration; genome-wide association; coagulation; kallikrein-kinin system

Introduction

L-arginine is a naturally occurring semi-essential amino acid, which plays a key biological role in cardiovascular physiology, immune function and hormonal signalling. L-arginine is the substrate for the synthesis of the potent vasodilator nitric oxide (NO) by the endothelial NO synthase (eNOS) in blood vessels. NO is crucial in maintaining vascular homeostasis, and its depletion may contribute to the development of hypertension, atherosclerosis, endothelial dysfunction and vascular inflammation (1–5). L-arginine bioavailability is a rate limiting factor in the synthesis of NO (3, 4) and low levels of circulating L-arginine are reported in a number of diseases and conditions (6). Oral administration or infusion of L-arginine has shown the beneficial effect of improved endothelia function and reduced blood pressure (7, 8).

Recent studies suggest that L-arginine concentrations are heritable, and associated with genetic variation in *ARG1* (9). In contrast, several studies using metabolome wide profiling, with semi-quantitative measurement of L-arginine, fail to identify genetic variation influencing L-arginine concentrations (10–16). The genetic factors regulating L-arginine levels are not well understood.

We performed a GWAS of serum L-arginine levels using targeted quantification of L-arginine by mass spectrometry (MS), to identify the genetic factors influencing L-arginine and investigate the potential relationships between L-arginine metabolism and cardiovascular traits.

Materials and methods

Samples and phenotypes at the discovery stage

We investigated 2,487 participants selected at random from the London Life Sciences Prospective Population (LOLIPOP) Study (17, 18), comprising 998 Europeans and 1,489 Indian Asians. The LOLIPOP study is approved by the local Research Ethics Committee, and all participants provided written consent for genetic studies.

Participants were interviewed by trained research nurses according to a standardized protocol. An interviewer-administered questionnaire was used to collect data on demographic factors, tobacco and alcohol consumption, as well as personal and family medical history. Indian Asians had all four grandparents born on the Indian subcontinent (India, Pakistan, Sri Lanka, or Bangladesh). Europeans were of self-reported white ancestry. Anthropometric (height, weight, waist and hip circumference) and blood pressure measurements were recorded. Blood samples were collected after an 8 hour fast for measurement of fasting glucose, hemoglobin A1c (HbA1c), lipid profile, creatinine, and high-sensitivity C-reactive protein (hsCRP). Aliquots of whole blood and serum were stored at -80°C.

Serum L-arginine measurement at the discovery stage

Serum concentrations of L-arginine were measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a Prominence LC-20AD_{XR} binary pump (Shimadzu, Kyoto, Japan) coupled to a QTRAP 5500 hybrid triple quadrupole mass spectrometer (AB Sciex, Framingham, MA, US). Serum (25 µL) was mixed with equal volumes of distilled water, 50 mM ammonium hydroxide, and a solution containing isotopically labelled internal standard, d₇-Arg (Cambridge Isotopes Laboratories, Andover, MA, US). Dithioerythritol (25 µL, 100 mM) was added and the samples were left for 15 min followed by protein precipitation using 25 µL perchloric acid (3.3 M) and centrifugation at 14'000 rpm for 10 min. Supernatant (5 µL) was then added to 195 µL NH₄⁺ (10 mM), and 7 µL was subsequently injected onto a Kinetex PFP column (100 × 2.10 mm, 2.6 µm; Phenomenex, Torrance, CA, US) kept at 35°C. Gradient elution at 300 µL/min was employed with mobile phases A (0.05 % formic acid in water, v/v) and B (0.1 % formic acid in methanol, v/v). 0% B was initially held at for 1.5 min, followed by a linear gradient reaching 10% B at 4.0 min, and 100% B at 4.3 min, and held there for 1.5 min, followed by equilibration at 0% B for 2 min. Gas 1 and 2 were set at 50, collision gas at high, and the curtain gas, ion spray voltage, and source temperature were set at 35, 4000 V, and 700°C, respectively. The following ion transitions were used for quantitation in multiple reaction monitoring (MRM) mode: m/z 175 → m/z 70 for L-arginine, and m/z 182 → m/z 77 for d₇-Arg. Optimisation of MRM transitions, data acquisition, and analysis were performed with Analyst 1.6.1 (AB Sciex, Framingham, MA, US). L-arginine was quantified by comparing the area ratio against the internal standard with calibration curves constructed in control plasma. For each batch, a number of randomly chosen samples, as well as all identified outliers, were identified and repeated to ensure accurate quantitation. Inter- and intra-assay coefficients of variation for L-arginine assessed at three QC levels were ~7%, well below the accepted limit recommended by the European Medicines Agency (see ^{URL}).

URLs

Guideline on bioanalytical method validation, European Medicines Agency, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/WC500109686.pdf
R-project, <https://www.r-project.org/>
EIGENSOFT, http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html
IMPUTE2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html
SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snpTest/snpTest.html
METAL, http://genome.sph.umich.edu/wiki/METAL_Documentation
LOCUSROOM, <http://locuszoom.sph.umich.edu/locuszoom/>
NHGRI GWAS Catalog, <http://www.ebi.ac.uk/gwas/>

Genotyping and imputation at the discovery stage

DNA was extracted from whole blood. Genome-wide genotyping was carried out using the Illumina Human610 array, as previously described (17), in 901 Europeans and 1,394 Indian Asians from amongst the 2,487 LOLIPOP participants with L-arginine measurements. SNPs with minor allele frequencies <1%, p-values for Hardy-Weinberg equilibrium (p-HWE) <1x10⁻⁶, or missing rates >5% were removed. Exclusion criteria for samples were duplicates, ethnic outliers, cryptic relatedness, gender discrepancy, extreme heterogeneity, missing rate >5%, or incomplete phenotype data.

Imputation of un-genotyped variants was performed in European (544,620 directly genotyped SNPs) and Indian Asian (541,497 SNPs) samples separately with IMPUTE2 v2.3.0 (19), using the cosmopolitan 1000 Genomes haplotypes (phase 1, March 2012 version, genomic build 37) as reference (20). The reference haplotypes most relevant to the study population by allelic consistency are chosen automatically and it has shown that the cosmopolitan reference haplotypes produce high imputation accuracy (20).

Replication studies

The lead SNPs identified at the discovery stage were tested in two independent cohorts with plasma L-arginine measurements. All participants in the replication studies provided written informed consent for genetic studies, and the study protocols were approved by local Institutional review boards and ethical committees.

The European Prospective Investigation into Cancer, Norfolk (EPIC-Norfolk) is a prospective population cohort of men and women of European origin aged 40-79 years living in Norfolk, UK, recruited from general practices between 1993 and 1998 (21). A nested case-control study of incident stroke was performed with an average follow-up of eleven years (N = 1,000). Controls were matched to cases on age and gender. Clinical and lifestyle data and blood samples were collected at baseline when all the cases and controls were stroke free. Baseline plasma L-arginine levels were measured with a modified approach as in our discovery study. Genotyping was performed on the Affymetrix 500K arrays, followed by quality control and imputation to Phase 1 1000 Genomes reference panel.

The Framingham Heart Study (FHS) is a longitudinal observational, community-based cohort initiated in 1948 in Framingham, MA, USA, to investigate risk factors prospectively for cardiovascular diseases (22). In this study, 2,981 participants of the Framingham Offspring Cohort (FOC), who had complete genotypic (Affymetrix 500K and 50K) and plasma L-arginine measurements were analysed after exclusions. The FOC consists of children (and children's spouses) of the original Framingham cohort participants. Details for plasma L-arginine measurement, genotyping and imputation were described previously (9).

Statistical analysis

In the discovery cohorts, outliers with serum L-arginine levels exceeding 3 standard deviations from the means in males or females separately were removed from further analysis. Continuous and discrete traits were compared between Europeans and Indian Asians using one-way ANOVA and Chi-square tests respectively with R (version 2.15.0, Table 1). Linear regression was used to investigate the associations between serum L-arginine levels and other phenotypic traits, with adjustment for age, gender and ethnicity (Table S1).

Principal components (PCs) of the genotype data were obtained with smartpca in EIGENSOFT using ~100K linkage disequilibrium (LD)-pruned genome-wide SNPs. Linear regression with additive genetic effect for association and conditional analysis were performed with SNPTEST (v2.4.1), adjusted for age, gender, and PCs 1-5 (23). No data transformation was applied to serum L-arginine as it conformed approximately to a normal distribution. Markers with minor allele frequency (MAF) <0.01, p-HWE <1×10⁻⁶, or imputation quality score <0.4 were excluded at the discovery stage. A total of 28.7M markers were obtained before quality control, in which about 8M SNPs retained after filtering for further analysis. The association of SNPs with L-arginine concentrations was tested in Europeans and Indian Asians separately, and association test results corrected for the genomic inflation factors. We then carried out inverse variance meta-analyses across the two ethnic groups using a fixed effects model as implemented in METAL (24). Regional plots for the associated loci were generated using LOCUSZOOM using genome build GRCh37/hg19 and 1000 Genome Nov 2014 European (EUR) population for linkage disequilibrium (LD) (25). Trait variance explained by the sentinel (lead) SNP(s) was calculated in Europeans or Indian Asians separately, using linear regression model with covariates adjusted as above. The r^2 from the regression model was taken as the estimates of variance explained.

In replication studies, association analyses of the sentinel or proxy SNPs and plasma L-arginine were carried out in FHS, using a linear mixed effects model with relationship coefficient matrix to account for familial correlation, adjusted for sex, age, diabetes mellitus, systolic and diastolic blood pressure, smoking, body mass index, and serum creatinine levels. In EPIC-Norfolk, association analyses were performed in cases (N = 500) and controls (N = 500) separately, under the additive genetic model with adjustment for age and gender. In meta-analyses of combined discovery and replication studies, we used the weighted Z-score method in METAL along with a random-effects model in METASOFT to account for heterogeneity between discovery and replication studies (26).

Expression quantitative trait loci (eQTLs) analysis

We analysed the relationship between the sentinel SNPs influencing L-arginine concentrations and the expression of genes within 1Mb either side of the sentinel SNPs in peripheral blood leucocytes amongst 816 Indian Asian participants of the LOLIPOP study. Bonferroni correction was applied for testing multiple genes. In brief, transcriptome-wide gene expression in peripheral blood leucocytes was measured using the Illumina Human HT-12 v4 expression BeadChip according to manufacturer's protocol in two LOLIPOP

cohorts of Indian Asian participants. Negative control-probes were used to calculate detection p-value threshold ($p < 10^{-9}$, corresponding to $p < 0.05$ adjusted for the number of tests). Transcripts with call-rates $< 10\%$ were removed. Expression intensities were read and processed using the R packages Illuminaio and Limma. Gene expression values were subjected to background correction, quantile normalisation and $\log_2$ transformation. Linear regression was applied to the processed expression data adjusted for age, gender, RNA integrity number, RNA conversion batch, scanning batch, and array position. Residuals from the regression were obtained for the following genetic association analysis.

Genotyping of the two cohorts was carried out using Illumina OmniExpress (N=612) or Illumina OmniExpressExome (N=204) whole-genome arrays. Standard quality control of GWAS was applied and genotype imputation was performed using the 1000 Genomes cosmopolitan reference panel as described above. Association between the residuals of the gene expression and the sentinel SNPs (rs71640036 vs *KLKB1* and rs2545801 vs *F12*) was carried out with Quicktest separately in the two datasets, adjusted for age, gender, and PCs 1-5. METAL was used to combine the effects of both cohorts by inverse-variance meta-analysis. Association results were corrected for genomic inflation factors before and after meta-analysis.

Finally we examined the association of our sentinel or proxy SNPs with gene expression using published data for liver and other tissues (27–29), using the Genotype-Tissue Expression (GTEx) data (30).

Association with cardiovascular and metabolic traits

We examined the relationship of the sentinel SNPs influencing L-arginine concentrations with blood pressure, cardiovascular disease and metabolic traits using public databases and the NHGRI GWAS Catalog (31). Where the sentinel SNP was not available in the respective dataset, we used the best available proxy SNP, selected based on LD ($r^2 > 0.9$).

Results

Characteristics of participants

The characteristics of the discovery and replication study samples are described in Tables 1 and S2, respectively. Serum L-arginine levels are higher among Indian Asians than Europeans ($p=0.01$). Indian Asians also have higher rates of coronary heart disease and type 2 diabetes, and higher measurements of waist-hip ratio and glycemic traits (fasting glucose, fasting insulin, and HbA1c), and lower levels of total and HDL cholesterol, compared to Europeans ($p < 0.05$, Table 1).

Genome-wide association in the discovery study

The LOLIPOP GWAS genomic inflation factors were 1.008 for Europeans, 1.004 for Indian Asians, and 1.002 for meta-analysis in the discovery study. Quantile-quantile (QQ) plots show evidence for association between genetic variation and serum L-arginine levels in both ethnic groups and in combined analysis (Figure S1). Twenty-four SNPs in the European samples and fifty-nine SNPs in the Indian Asian samples reach genome-wide significance

for association with serum L-arginine levels ($p < 5 \times 10^{-8}$, Figure S2), all on chromosome 4. In combined analysis, 82 SNPs are associated with L-arginine levels, comprising 77 SNPs at the *KLKB1* locus (kallikrein B, plasma (Fletcher Factor) 1) on chromosome 4, and 5 SNPs at the *F12* locus (Coagulation Factor XII (Hageman Factor)) on chromosome 5 (Table S3, Figure S2).

KLKB1 locus

The SNP with the lowest p-value (rs71640036, $p = 9.7 \times 10^{-25}$, the sentinel SNP) is an intron variant in the *KLKB1* gene (Tables 2 and S3). The G allele of rs71640036 is associated with decreased serum L-arginine levels at a rate 10 $\mu\text{mol/L}$ per allele copy (Table 3). The sentinel SNP is highly significantly associated with serum L-arginine levels in both ethnic groups and there is no significant heterogeneity between the two groups (Tables 2 and S3).

A missense non-synonymous SNP rs3733402 (Asn143Ser), which is 3,086 base pairs upstream of the sentinel SNP is also highly significantly associated with serum L-arginine levels ($p = 1.0 \times 10^{-23}$, Table S3). SNP rs3733402 is in high LD with the sentinel SNP ($r^2 = 0.964$ in Europeans and 0.981 in Indian Asians). The G allele of rs3733402 is associated with decreased serum L-arginine levels. Conditional analyses on either the sentinel SNP rs71640036 or rs3733402 showed that none of the other SNPs are significantly associated with serum L-arginine levels ($p > 0.05$ after Bonferroni correction) in this region (± 500 kb from the sentinel SNP, covering the *KLKB1*, *CYP4V2*, *F11* and other nearby genes, data not shown).

F12 locus

The five SNPs significantly associated with serum L-arginine levels on chromosome 5 are in or near the *F12* gene. The sentinel SNP rs2545801 ($p = 6.8 \times 10^{-12}$) is intergenic, but is in high LD with rs1801020 ($r^2 = 0.993$ in Europeans and 0.996 in Indian Asians) in the highly conserved 5' untranslated (promoter) region of *F12* gene, itself strongly associated with L-arginine levels ($p = 1.2 \times 10^{-11}$, Table S3). The minor allele T of the sentinel SNP is associated with decreased serum L-arginine levels (7 $\mu\text{mol/L}$ per allele copy, Table 3). There is no significant heterogeneity between the two ethnic groups. Conditional analyses on the sentinel SNP rs2545801 showed that there are no secondary associations with serum L-arginine levels in this region (± 500 kb from the sentinel SNP, data not shown). Together the sentinel SNPs from both *KLKB1* and *F12* loci explain approximately 7% of the total variance of serum L-arginine after adjusting for age, gender, and PCs 1-5 (Table 2).

Replication studies

Combined analyses of the replication cohorts showed significant associations at both the *KLKB1* and *F12* loci ($p < 0.004$, Table 2). Meta-analysis combining discovery and replication studies also showed strong associations with circulating L-arginine levels (Table 2).

Suggestive associations and other analyses

In combined analysis of the European and Indian Asian samples in the discovery study, 129 SNPs showed p-values $< 1 \times 10^{-5}$ but $> 5 \times 10^{-8}$ with consistent direction of effects in both ethnic groups (Table S4). Two intron SNPs in *ARG1* gene and 7 SNPs in the nearby *MED23*

gene on chromosome 6 are associated with serum L-arginine levels (lowest $p=5 \times 10^{-7}$, Table S4). Other regions with suggestive associations are in non-gene regions of chromosomes 2 and 11, *SOX6* (11p15.3), *PRH1-PRR4* (12p), *GPR133* (12q24.33), *HS3ST4* (16p11.2), *RP11-638* (18q22.1), and *TSHZ2* (20q13.2) (Table S4).

We separately performed association analyses with dominant or recessive models; no additional loci reached genome-wide significance. Gene-gene interaction analysis did not show any synergistic effects for the SNPs between the *KLKB1* and *F12* genes beyond their additive effect (data not shown).

Gene expression studies

SNP rs2545801 is associated with expression of *F12* ($p=0.04$ after Bonferroni correction), and SNP rs71640036 is associated with expression of *CYP4V2* ($p=0.01$, underlying mechanisms unknown) in our peripheral blood leucocyte samples (Table S5). In contrast there was no significant association of rs71640036 with expression of *KLKB1* ($p=0.53$). The sentinel SNPs were not significantly associated with expression levels of any other genes within the regions ($\pm 1\text{Mb}$). Both *KLKB1* and *F12* genes are predominantly expressed in liver (30). SNP rs2731672, which is in high LD ($r^2 = 0.994$ in Europeans and 0.987 in Indian Asians) with the sentinel SNP rs2545801 at the *F12* locus is strongly associated with *F12* gene expression in liver ($p=1.2 \times 10^{-23}$) (27).

Association of sentinel SNPs with phenotypic traits

Published studies show that the sentinel SNPs at the *KLKB1* and *F12* loci are also strongly associated with other phenotypic traits including levels of phenylalanine (32, 33), des-arg(9) bradykinin (12, 15), active plasma renin (34), midregional-proadrenomedullin (MR-pro-ADM) and C-terminal-pro-endothelin-1 (35) (Table S6). In addition, the *KLKB1* SNPs are associated with histidine (32), leucylphenylalanine (14), HXGXA (14), plasma B-Type Natriuretic Peptide (36) and ratios of serum amino acid levels (32), while *F12* SNPs are highly associated with blood FXII antigen levels (37) and prolonged activated partial thromboplastin time (aPTT) (38) and plasma renin activity (39) (Table S6).

Association of sentinel SNPs with cardiovascular disease

SNP rs71640036 at the *KLKB1* locus is associated with both systolic and diastolic blood pressure amongst 134,773 participants of UK Biobank ($p=9 \times 10^{-5}$ and $p=8 \times 10^{-3}$ respectively, Table 4). SNP rs2545801 at the *F12* locus is also associated with systolic blood pressure ($p=8 \times 10^{-3}$) but not with diastolic blood pressure. There were no associations of rs71640036 or rs2545801 with coronary heart disease or hsCRP, after Bonferroni correction for multiple testing (Tables 4 and S6).

Discussion

We show that common genetic variation at the *KLKB1* and *F12* loci is an important determinant of serum L-arginine concentrations. *KLKB1* and *F12* both encode components of the kinin system. *KLKB1* (4q35.2) encodes prekallikrein, which is a precursor of the serine protease plasma kallikrein. *F12* (5q35.3) encodes the coagulation Factor XII (FXII), a

serine protease involved in coagulation activation through the intrinsic pathway. Kallikrein and FXII are closely linked biologically and activate each other in a reciprocal fashion.

We found that nonsynonymous SNP rs3733402 in the *KLKB1* gene, which is in high LD with the sentinel SNP rs71640036, is associated with decreased serum L-arginine levels. This SNP leads to the replacement of asparagine with serine at codon 143 in exon 5 of *KLKB1* gene, located in apple domain A2 of plasma kallikrein. Previous studies show that this non-synonymous SNP reduces the binding activity of apple domain A2 to circulating high molecular weight kininogen (HMWK), and is associated with low levels of plasma prekallikrein and bradykinin metabolites (40).

The minor A allele of SNP rs1801020 in the promoter region of the *F12* gene, which was nearly in complete LD with the sentinel SNP rs2545801 and rs2731672, is associated with reduced serum L-arginine levels in our study, decreased plasma FXII concentrations (37, 41, 42) and prolonged aPTT (38). The SNP is located 4 bases upstream of the translation initiation codon of factor XII. Experiments using reporter constructs suggest that the altered upstream open reading frame (uORF) through the minor T allele is responsible for >50% reduction of FXII protein levels and <30% reduction of mRNA levels (43). The eQTL data showed that these genetic variants are associated with reduced *F12* gene expression in liver and blood leucocytes.

The mechanisms linking genetic variants in *KLKB1* and *F12* to serum L-arginine concentrations remain to be determined. We note that both *KLKB1* and *F12* are involved in bradykinin release from HMWK. Bradykinin is a nonapeptide with the amino acid sequence Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg and is rapidly degraded once it has been generated in cardiovascular system. Aminopeptidase P (APP) cleaves positions 1-2, while carboxypeptidase N (CPN) cleaves positions 8-9, both enzymatic reactions lead to the release of L-arginine (44–46). These observations raise the possibility that genetic variation at *KLKB1* and *F12* may influence blood L-arginine levels through an effect on bradykinin release and degradation.

Although the role of nitric oxide in human physiology is well described, the relationship of circulating L-arginine metabolism with human disease is uncertain. We found that SNP rs71640036 in the *KLKB1* gene, which is strongly associated with lower L-arginine levels, is also associated with lower blood pressure and the *F12* variant is associated with higher systolic blood pressure. These findings raise the possibility that L-arginine may be involved in blood pressure regulation. However, since plasma kallikrein and FXII participate in multiple key signalling pathways (35, 37, 47), the association between their genetic variants and blood pressure through L-arginine needs further investigation.

There is increasing evidence showing associations between variants in the *KLKB1* and *F12* loci and cardiovascular biomarkers (35, 36). The sentinel SNPs, however, showed only borderline or non-significant associations with CAD in the recent large-scale CARDIoGRAMplusC4D study ($p=0.02$ for the *KLKB1* variant) (48). This reflects the complex and multifactorial aetiology of CAD and these loci may contribute to the risk of CAD through intermediate traits, such as blood pressure and thrombosis. The same applies

to hsCRP, a marker of systemic inflammation. Given that hsCRP is highly associated with serum L-arginine concentrations in our data and the *KLKB1/F12* genes are involved in inflammation, further studies are required to explore the relation between L-arginine and hsCRP.

The kinin system also plays an important role in the contact activation of the intrinsic pathway in coagulation and fibrinolysis. A recent study indicated that the intrinsic pathway-dependent thrombin generation may be a risk factor for ischemic stroke and the *F12* variant is associated with thrombin generation (49). Factor XI (FXI), encoded by the *F11* gene, is a key factor in the intrinsic pathway and consistently associated with thrombosis risk. The *KLKB1* and *F11* genes are closely located but with moderate LD (r^2 up to 0.3). Our study revealed no association of *F11* variants with serum L-arginine levels after exclusion of the effects from the *KLKB1* locus. How the L-arginine metabolism affects thrombosis risk remains unclear.

In conclusion, we have identified genetic variation in the *KLKB1* and *F12* genes as important determinants of serum levels of L-arginine, and provide potential evidence for their role in blood pressure regulation. Our findings provide new insight into the pathways influencing metabolism of L-arginine, the essential substrate in the synthesis of NO, and may help resolve the understanding of the pathways linking L-arginine with endothelial function and cardiovascular health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z), the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143), and Action on Hearing Loss (G51). Fredrik Jerneén is supported by the Norwegian research Council. This work was also supported by the Norwegian research Council (to Helga Refsum) to fund the experiments. We thank the participants and research staff who made the study possible, and in particular the staff at University of Oxford (Cynthia Prendergast, Carol Johnston, and Maria Valdivia-Garcia) for performing the amino acid measurements. The Framingham Heart Study was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix Inc. for genotyping services (Contract No. N02-HL-6-4278). EPIC-Norfolk is supported by the Medical Research Council programme grants (G0401527, G1000143) and Cancer Research UK programme grant (C864/A8257). The authors would like to acknowledge the contribution of the staff and participants of the Framingham Offspring Cohort Study and the EPIC-Norfolk Study. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. We thank Dr Abbas Dehghan and the CHARGE Inflammation working group for providing the hsCRP and SNP association results. This research has been conducted using the UK Biobank Resource. The views expressed are those of the author(s) and not necessarily those of the Imperial College Healthcare NHS Trust, the NIHR or the Department of Health.

Financial support: the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931, G0401527, G1000143), the Wellcome Trust (084723/Z/08/Z), the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143), Action on Hearing Loss (G51), Cancer Research UK (C864/A8257), the Norwegian research Council, and the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix Inc. for genotyping services (Contract No. N02-HL-6-4278).

References

1. Popolo A, Adesso S, Pinto A, et al. L-Arginine and its metabolites in kidney and cardiovascular disease. *Amino Acids*. 2014; 46:2271–2286. [PubMed: 25161088]
2. Boger RH. The pharmacodynamics of L-arginine. *Altern Ther Health Med*. 2014; 20:48–54. [PubMed: 24755570]
3. Lorin J, Zeller M, Guillard JC, et al. Arginine and nitric oxide synthase: regulatory mechanisms and cardiovascular aspects. *Mol Nutr Food Res*. 2014; 58:101–116. [PubMed: 23740826]
4. Luiking YC, Ten Have GA, Wolfe RR, et al. Arginine de novo and nitric oxide production in disease states. *Am J Physiol Endocrinol Metab*. 2012; 303:E1177–1189. [PubMed: 23011059]
5. Tousoulis D, Kampoli AM, Tentolouris C, et al. The role of nitric oxide on endothelial function. *Curr Vasc Pharmacol*. 2012; 10:4–18. [PubMed: 22112350]
6. Morris SM Jr. Arginases and arginine deficiency syndromes. *Curr Opin Clin Nutr Metab Care*. 2012; 15:64–70. [PubMed: 22037011]
7. Gokce N. L-arginine and hypertension. *J Nutr*. 2004; 134:2807S–2811S. discussion 2818S–2819S. [PubMed: 15465790]
8. Dong JY, Qin LQ, Zhang Z, et al. Effect of oral L-arginine supplementation on blood pressure: a meta-analysis of randomized, double-blind, placebo-controlled trials. *Am Heart J*. 2011; 162:959–965. [PubMed: 22137067]
9. Luneburg N, Lieb W, Zeller T, et al. Genome-wide association study of L-arginine and dimethylarginines reveals novel metabolic pathway for symmetric dimethylarginine. *Circ Cardiovasc Genet*. 2014; 7:864–872. [PubMed: 25245031]
10. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet*. 2008; 4:e1000282. [PubMed: 19043545]
11. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet*. 2010; 42:137–141. [PubMed: 20037589]
12. Suhre K, Shin SY, Petersen AK, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*. 2011; 477:54–60. [PubMed: 21886157]
13. Rhee EP, Ho JE, Chen MH, et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab*. 2013; 18:130–143. [PubMed: 23823483]
14. Yu B, Zheng Y, Alexander D, et al. Genetic determinants influencing human serum metabolome among African Americans. *PLoS Genet*. 2014; 10:e1004212. [PubMed: 24625756]
15. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014; 46:543–550. [PubMed: 24816252]
16. Draisma HH, Pool R, Kobl M, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat Commun*. 2015; 6:7208. [PubMed: 26068415]
17. Chambers JC, Zhang W, Zabaneh D, et al. Common genetic variation near melatonin receptor MTNR1B contributes to raised plasma glucose and increased risk of type 2 diabetes among Indian Asians and European Caucasians. *Diabetes*. 2009; 58:2703–2708. [PubMed: 19651812]
18. Chambers JC, Elliott P, Zabaneh D, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet*. 2008; 40:716–718. [PubMed: 18454146]
19. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007; 39:906–913. [PubMed: 17572673]
20. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011; 1:457–470. [PubMed: 22384356]
21. Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer*. *Br J Cancer*. 1999; 80(Suppl 1):95–103. [PubMed: 10466767]
22. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979; 110:281–290. [PubMed: 474565]

23. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 2010; 11:499–511. [PubMed: 20517342]
24. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26:2190–2191. [PubMed: 20616382]
25. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26:2336–2337. [PubMed: 20634204]
26. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet.* 2011; 88:586–598. [PubMed: 21565292]
27. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008; 6:e107. [PubMed: 18462017]
28. Innocenti F, Cooper GM, Stanaway IB, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* 2011; 7:e1002078. [PubMed: 21637794]
29. Greenawalt DM, Dobrin R, Chudin E, et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res.* 2011; 21:1008–1016. [PubMed: 21602305]
30. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013; 45:580–585. [PubMed: 23715323]
31. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014; 42:D1001–1006. [PubMed: 24316577]
32. Kettunen J, Tukiainen T, Sarin AP, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet.* 2012; 44:269–276. [PubMed: 22286219]
33. Inouye M, Ripatti S, Kettunen J, et al. Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet.* 2012; 8:e1002907. [PubMed: 22916037]
34. Biswas N, Maihofer AX, Mir SA, et al. Polymorphisms at the F12 and KLKB1 loci have significant trait association with activation of the renin-angiotensin system. *BMC Med Genet.* 2016; 17:21. [PubMed: 26969407]
35. Verweij N, Mahmud H, Mateo Leach I, et al. Genome-wide association study on plasma levels of midregional-proadrenomedullin and C-terminal-pro-endothelin-1. *Hypertension.* 2013; 61:602–608. [PubMed: 23381795]
36. Musani SK, Fox ER, Kraja A, et al. Genome-wide association analysis of plasma B-type natriuretic peptide in blacks: the Jackson Heart Study. *Circ Cardiovasc Genet.* 2015; 8:122–130. [PubMed: 25561047]
37. Williams FM, Carter AM, Hysi PG, et al. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol.* 2013; 73:16–31. [PubMed: 23381943]
38. Tang W, Schwienbacher C, Lopez LM, et al. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am J Hum Genet.* 2012; 91:152–162. [PubMed: 22703881]
39. Lieb W, Chen MH, Teumer A, et al. Genome-wide meta-analyses of plasma renin activity and concentration reveal association with the kininogen 1 and prekallikrein genes. *Circ Cardiovasc Genet.* 2015; 8:131–140. [PubMed: 25477429]
40. Katsuda I, Maruyama F, Ezaki K, et al. A new type of plasma prekallikrein deficiency associated with homozygosity for Gly104Arg and Asn124Ser in apple domain 2 of the heavy-chain region. *Eur J Haematol.* 2007; 79:59–68. [PubMed: 17598838]
41. Kanaji T, Okamura T, Osaki K, et al. A common genetic polymorphism (46 C to T substitution) in the 5'-untranslated region of the coagulation factor XII gene is associated with low translation efficiency and decrease in plasma factor XII level. *Blood.* 1998; 91:2010–2014. [PubMed: 9490684]
42. Calafell F, Almasy L, Sabater-Lleal M, et al. Sequence variation and genetic evolution at the human F12 locus: mapping quantitative trait nucleotides that influence FXII plasma levels. *Hum Mol Genet.* 2010; 19:517–525. [PubMed: 19933701]

43. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A*. 2009; 106:7507–7512. [PubMed: 19372376]
44. Dendorfer A, Wolfrum S, Wagemann M, et al. Pathways of bradykinin degradation in blood and plasma of normotensive and hypertensive rats. *Am J Physiol Heart Circ Physiol*. 2001; 280:H2182–2188. [PubMed: 11299220]
45. Kuoppala A, Lindstedt KA, Saarinen J, et al. Inactivation of bradykinin by angiotensin-converting enzyme and by carboxypeptidase N in human plasma. *Am J Physiol Heart Circ Physiol*. 2000; 278:H1069–1074. [PubMed: 10749699]
46. Sheikh IA, Kaplan AP. Mechanism of digestion of bradykinin and lysylbradykinin (kallidin) in human serum. Role of carboxypeptidase, angiotensin-converting enzyme and determination of final degradation products. *Biochem Pharmacol*. 1989; 38:993–1000. [PubMed: 2539165]
47. Bjorkqvist J, Jamsa A, Renne T. Plasma kallikrein: the bradykinin-producing enzyme. *Thromb Haemost*. 2013; 110:399–407. [PubMed: 23846131]
48. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015; 47:1121–1130. [PubMed: 26343387]
49. Olson NC, Butenas S, Lange LA, et al. Coagulation factor XII genetic variation, ex vivo thrombin generation, and stroke risk in the elderly: results from the Cardiovascular Health Study. *J Thromb Haemost*. 2015; 13:1867–1877. [PubMed: 26286125]

Table 1

Characteristics of participants for the discovery cohorts. Results are presented as mean (SD) or % amongst Europeans (n = 998) and Indian Asians (N = 1,489). P value is for the comparison between the two populations.

	Europeans	Indian Asians	P
Age (years)	55.4 (10.0)	54.4 (10.0)	1.8×10^{-2}
Female (%)	29.3	36.6	1.7×10^{-4}
Current smoking (%)	17.0	7.2	3.3×10^{-14}
Alcohol drinking (%)	73.0	31.6	$< 2 \times 10^{-16}$
Type 2 diabetes (%)	8.3	20.5	4.0×10^{-16}
Coronary heart disease (%)	2.4	6.4	8.3×10^{-6}
Hypertension (%)	43.5	50.5	6.9×10^{-4}
Body mass index (kg/m²)	27.6 (4.75)	27.2 (4.2)	3.9×10^{-2}
Waist hip ratio	0.92 (0.08)	0.95 (0.08)	$< 2 \times 10^{-16}$
Diastolic blood pressure (mmHg)	80.8 (10.2)	80.0 (10.3)	7.7×10^{-2}
Systolic blood pressure (mmHg)	132.5 (19.0)	131.1 (19.4)	7.3×10^{-2}
Total cholesterol (mmol/L)	5.6 (1.1)	5.3 (1.1)	1.9×10^{-11}
Triglycerides (mmol/L)	1.6 (1.1)	1.7 (1.2)	1.5×10^{-2}
LDL cholesterol (mmol/L)	3.5 (0.9)	3.3 (0.9)	3.9×10^{-8}
HDL cholesterol (mmol/L)	1.4 (0.4)	1.3 (0.3)	$< 2 \times 10^{-16}$
Glucose (mmol/L)	5.5 (1.5)	5.9 (1.9)	5.2×10^{-9}
Insulin (IU/L)	9.8 (9.1)	12.1 (12.4)	7.5×10^{-7}
HbA1c (%)	5.5 (0.9)	6.1 (1.2)	$< 2 \times 10^{-16}$
HOMA-IR	2.6 (3.2)	3.3 (4.3)	2.6×10^{-6}
hsCRP (mg/L)	3.7 (6.9)	4.0 (5.3)	2.1×10^{-1}
Creatinine (μmol/L)	92.0 (14.9)	90.3 (25.1)	4.8×10^{-2}
L-arginine (μmol/L)	131.7 (31.6)	135.0 (33.5)	1.3×10^{-2}

Footnote: Alcohol drinking is defined as any consumption of alcohol in the previous week. Coronary heart disease is defined as history of myocardial infarction, coronary revascularization, or angiographically severe disease. Hypertension is defined as history of hypertension on medical treatment. HOMA-IR, homeostatic model assessment-estimated insulin resistance index. hsCRP, high-sensitivity C-reactive protein.

Table 2
Association of sentinel SNPs identified with serum or plasma L-arginine levels in genome-wide association.

SNP (EA/AA) / Stage	Data	EAF	Effect (SE)	P	N	P _{het}
rs71640036 (G/T)						
Discovery	LOLIPOP European	0.50	-9.1 (1.5)	1.4×10^{-09}	901	-
	LOLIPOP Indian Asian	0.61	-11.2 (1.3)	4.7×10^{-17}	1,394	-
	Discovery combined	-	-10.3 (1.0)	9.7×10^{-25}	2,295	3.1×10^{-01}
Replication	EPIC-Norfolk-cases	0.47	-0.87 (1.2)	4.8×10^{-01}	450	-
	EPIC-Norfolk-controls	0.50	-1.51 (1.0)	1.4×10^{-01}	450	-
	FHS	0.50	-1.4 (0.5)	9.5×10^{-03}	2,981	-
	Replication combined	-	-1.4 (0.5)	2.4×10^{-03}	3,881	8.3×10^{-01}
All combined	WZ method	-	-	3.8×10^{-18}	6,176	7.9×10^{-09}
	RE method	-	-	1.6×10^{-22}	6,176	-
rs2545801 (T/C)						
Discovery	LOLIPOP European	0.24	-7.9 (1.8)	9.0×10^{-06}	901	-
	LOLIPOP Indian Asian	0.44	-6.8 (1.3)	1.3×10^{-07}	1,394	-
	Discovery combined	-	-7.1 (1.0)	6.8×10^{-12}	2,295	6.0×10^{-01}
Replication	EPIC-Norfolk-cases	0.28	-3.6 (1.2)	4.0×10^{-03}	457	-
	EPIC-Norfolk-controls	0.26	0.7 (1.2)	6.0×10^{-01}	457	-
	FHS	0.16	-4.3 (1.5)	3.4×10^{-03}	2,981	-
	Replication combined	-	-2.2 (0.7)	3.5×10^{-03}	3,895	1.3×10^{-01}
All combined	WZ method	-	-	6.3×10^{-12}	6,190	1.0×10^{-02}
	RE method	-	-	9.8×10^{-13}	6,190	-

Footnote: EA, effect allele; AA, alternate allele; EAF, effect allele frequency; Effect, the change of L-arginine concentration per copy of the effect allele (μmol/L); SE, standard error; R^2 , proportion of variance explained by the respective SNP; R^2 for combined effect of rs71640036 and rs2545801 was 0.075 for European and 0.066 for Indian; EPIC-Norfolk, the European Prospective Investigation into Cancer, Norfolk; FHS, Framingham Heart Study; WZ, weighted Z-score method for meta-analysis; RE, Random effects model for meta-analysis accounting for heterogeneity. Imputation quality scores for both sentinel SNPs in discovery and replication datasets were > 0.9 , except for rs2545801 in FHS (oevar_jmp = 0.25). Removing FHS from meta-analyses of rs2545801, there was a weak association in the replication cohort (EPIC-Norfolk study with adjustment for age, gender and stroke status, $p = 5.1 \times 10^{-2}$) and a significant association in combined discovery and replication analysis (WZ method, $p = 6.1 \times 10^{-12}$). In FHS, SNP rs3733402 was used as a proxy to rs71640036. Serum samples were used in the discovery studies and plasma samples were used in the replication studies.

Table 3

Serum L-arginine concentrations (mean [SD], $\mu\text{mol/L}$) in relation to the genotypes of sentinel SNPs.

SNP	Genotype	<u>Europeans</u>		<u>Indian Asians</u>	
		%	L-arginine	%	L-arginine
rs71640036	TT	0.24	140.1 (31.8)	0.16	149.6 (31.3)
	TG	0.50	134.3 (31.3)	0.47	137.7 (34.8)
	GG	0.26	121.2 (29.7)	0.38	126.8 (30.6)
rs2545801	CC	0.58	134.8 (31.8)	0.33	141.1 (34.1)
	TC	0.37	129.6 (30.9)	0.47	134.5 (33.3)
	TT	0.06	116.8 (23.7)	0.20	126.7 (31.1)

Table 4

Association of sentinel or proxy SNPs with cardiovascular disease.

SNP	EA/AA	Trait	N	Effect	P
<i><u>KLKB1</u></i>					
rs71640036	G/T	Coronary heart disease (48)	185,000	0.023	2×10 ⁻²
		Diastolic blood pressure [*]	134,773	-0.152	9×10 ⁻⁵
		Systolic blood pressure [*]	134,773	-0.180	8×10 ⁻³
<i><u>F12</u></i>					
rs2545801	T/C	Coronary heart disease (48)	185,000	0.017	1×10 ⁻¹
		Diastolic blood pressure [*]	134,773	0.067	1×10 ⁻¹
		Systolic blood pressure [*]	134,773	0.201	7×10 ⁻³

Footnote: EA, effect allele (minor allele, associated with reduced serum L-arginine levels); AA, alternate allele; Logistic and linear regressions (for coronary heart disease and blood pressure, respectively) under the genetic model of additive effect were applied. Effect was calculated as the logarithm of the odds ratio for coronary heart disease, or change of blood pressure (mmHg) per copy of the effect allele. *Data from UK Biobank. The critical p-value for significance after Bonferroni correction for six tests is 8×10^{-3} at the familywise error rate of 0.05.