Hospitals, Geneva, Switzerland, **47** Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **48** Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom, **49** Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, **50** Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, **51** Department of Computer Science and Networking, Wentworth Institute of Technology, Boston, Massachusetts, United States of America, **52** Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, **53** Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London, United Kingdom, **54** Ulm University Medical Centre, Department of Internal Medicine I, Ulm University, Ulm, Germany, **55** LKC School of Medicine, Imperial College London and Nanyang Technological University, Singapore, Singapore, **56** Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States of America, **57** Division of Primary Care Medicine, Department of Community Medicine and Primary Care and Emergency Medicine , Geneva University Hospitals, Geneva, Switzerland, **58** Centre for Population Health Sciences, The University of Edinburgh Medical School, Edinburgh, Scotland, United Kingdom, **59** Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Ernst-Moritz-Arndt University Greifswald, Greifswald, Germany, **60** Department of Prosthetic Dentistry, Gerostomatology and Dental Materials, University Medicine Greifswald, Greifswald, Germany, **61** Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany, **62** Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom, **63** Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **64** Clinical Chemistry Laboratory, Lausanne University Hospital, Lausanne, Switzerland, **65** William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom, **66** Synlab Centre of Laboratory Diagnostics, Heidelberg, Germany, **67** Institute of Medical Sciences, Uppsala University Hospital, Uppsala, Sweden, **68** Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **69** Department of Medicine, Division of Nephrology, University of Washington, Seattle, Washington, United States of America, **70** Faculty of Medicine, National Heart & Lung Institute, Cardiovascular Science, Hammersmith Hospital, Hammersmith Campus, Imperial College London, London, United Kingdom, **71** Imperial College Healthcare NHS Trust, London, United Kingdom, **72** Welch Center for Prevention, Epidemiology and Clinical Research, John Hopkins University, Baltimore, Maryland, United States of America, **73** Service of Nephrology, Lausanne University Hospital, Lausanne, Switzerland, **74** Division of Endocrinology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Calcium is vital to the normal functioning of multiple organ systems and its serum concentration is tightly regulated. Apart from *CASR*, the genes associated with serum calcium are largely unknown. We conducted a genome-wide association meta-analysis of 39,400 individuals from 17 population-based cohorts and investigated the 14 most strongly associated loci in ≤21,679 additional individuals. Seven loci (six new regions) in association with serum calcium were identified and replicated. Rs1570669 near *CYP24A1* ($P = 9.1E-12$), rs10491003 upstream of *GATA3* ($P = 4.8E-09$) and rs7481584 in *CARS* ($P = 1.2E-10$) implicate regions involved in Mendelian calcemic disorders: Rs1550532 in *DGKD* ($P = 8.2E-11$), also associated with bone density, and rs7336933 near *DGKH/KIAA0564* ($P = 9.1E-10$) are near genes that encode distinct isoforms of diacylglycerol kinase. Rs780094 is in *GCKR*. We characterized the expression of these genes in gut, kidney, and bone, and demonstrate modulation of gene expression in bone in response to dietary calcium in mice. Our results shed new light on the genetics of calcium homeostasis.

Citation: O'Seaghdha CM, Wu H, Yang Q, Kapur K, Guessous I, et al. (2013) Meta-Analysis of Genome-Wide Association Studies Identifies Six New Loci for Serum Calcium Concentrations. PLoS Genet 9(9): e1003796. doi:10.1371/journal.pgen.1003796

Editor: Gonçalo R. Abecasis, University of Michigan, United States of America

Received February 8, 2013; **Accepted** July 29, 2013; **Published** September 19, 2013

Funding: All grant numbers for each human study and for the animal studies are listed in the Acknowledgement section of the Supplementary material. There is no specific funding for the GWAS meta-analysis itself. Grant numbers: N01-AG-1-2100. HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HSN268201100012C, R01HL087641, R01HL59367, R01HL086694, U01HG004402, HHSN268200625226C. KO3598/2-1. UL1RR025005. N01-HC- 85239, N01-HC-85079 through N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01- HC-75150, N01-HC-45133, HHSN268201200036C, HL080295, HL087652, HL105756, AG-023629, AG-15928, AG-20098, AG-027058, UL1RR033176, UL1TR000124, DK063491. 33CSCO- 122661, 3200BO-111361/2, 3100AO-116323/1,310000-112552. LSHG-CT-2006- 018947. 108- 1080315-0302. LSHG-CT-2006-018947, 108-1080315- 0302. 108-1080315- 0302. N01-HC-25195. N02-HL-6-4278. N01AG62101, N01AG62103, N01AG62106. 1R01AG032098-01A1. HHSN268200782096C. ICS110.1/RF97.71. 263 MD 9164 and 263 MD 821336. SP/04/002. G0700704/84698. (MIUR) no. 5571/DSPAR/2002 and (FIRB) D. M. no. 718/Ric/2005. LSHG-CT-2006-018947. 01ZZ9603, 01ZZ0103, 01ZZ0403, 03IS2061A, 03ZIK012. 175.010.2005.011, 911-03-012. 014- 93-015; RIDE2. 050-060-810. vici, 918-76-619. Vasoplus-037254. FIRB - RBIN064YAT. LSHM-CT-2004-503485. WT064890, WT081682. 01ZZ9603, 01ZZ0103, and 01ZZ 0403. 03IS2061A. 03ZIK012. HEALTH-F2-2008-201865-GEFOS. HEALTH-F4-2007-201413. QLG2-CT-2002-01254. G20234. 091746/Z/10/Z. G9521010D. PG/02/128. SP/04/002. SP/04/002. PP00P3-133648. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: foxca@nhlbi.nih.gov (CSF); Murielle.Bochud@chuv.ch (MB)

☯ These authors contributed equally to this work.

¶ Writing group.

Introduction

Normal calcium homeostasis is regulated by three major hormones acting on their corresponding receptors in gut, kidney, and bone: parathyroid hormone (PTH) release governed by the calcium-sensing receptor (CASR), calcitonin, and the active metabolite of vitamin D, $1,25(OH)_2$-D. Despite heritability estimates of 33–78%, the genetic determinants of serum calcium are poorly understood [1,2,3]. We have previously reported a variant in *CASR* associated with calcium concentrations in European-ancestry individuals [4,5]. To detect additional loci, we conducted a two-stage genome-wide association meta-analysis

Author Summary

Calcium is vital to many biological processes and its serum concentration is tightly regulated. Family studies have shown that serum calcium is under strong genetic control. Apart from *CASR*, the genes associated with serum calcium are largely unknown. We conducted a genome-wide association meta-analysis of 39,400 individuals from 17 population-based cohorts and investigated the 14 most strongly associated loci in ≤21,679 additional individuals. We identified seven loci (six new regions) as being robustly associated with serum calcium. Three loci implicate regions involved in rare monogenic diseases including disturbances of serum calcium levels. Several of the newly identified loci harbor genes linked to the hormonal control of serum calcium. In mice experiments, we characterized the expression of these genes in gut, kidney, and bone, and explored the influence of dietary calcium intake on the expression of these genes in these organs. Our results shed new light on the genetics of calcium homeostasis and suggest a role for dietary calcium intake in bone-specific gene expression.

of serum calcium and studied expression of identified genes in key calcium homeostatic organs in the mouse under various calcium diets.

Results

Genome-wide association meta-analysis in Europeans

The discovery analysis consisted of 39,400 individuals from 17 population-based cohorts of European descent (**Table 1** and **Table S1**). There was little evidence for population stratification at study level (median genomic inflation factor, $\lambda = 1.006$) or meta-analysis level ($\lambda = 1.03$), and we detected an excess of association signals beyond those expected by chance (**Figure S1**).

The *CASR* locus, previously identified in Europeans, was confirmed in our meta-analysis ($P = 6.5E\text{-}59$, **Figure S2**) [4,5]. In addition, SNPs from five independent regions reached genome-wide significance ($P<5E\text{-}08$) in the overall discovery meta-analysis (**Figure 1, Table 1, Table S2**): rs1550532 (in *DGKD*, $P = 4.60E\text{-}08$), rs780094 (in *GCKR*; $P = 3.69E\text{-}11$), rs17711722 (near *VKORC1L1*, $P = 2.78E\text{-}11$), rs7481584 (in *CARS*, $P = 9.21E\text{-}10$) and rs1570669 (near *CYP24A1*; $P = 3.98E\text{-}08$).

Fourteen SNPs from Stage 1 were sent for Stage 2 validation in ≤21,679 additional Europeans: the twelve independent (≥1 Mb apart) SNPs with lowest P values (6.5E-59 to 8.1E-06) in Europeans and two additional genome-wide significant loci (rs9447004 and rs10491003) from a combined sample including 8318 Indian-Asians (**Table 1**). Of the fourteen SNPs, seven were considered successfully replicated (i.e. were in the same direction of effect as the discovery meta-analysis, had a one-side replication $P<0.05$ and were genome-wide significant ($P<5E\text{-}8$) in combined meta-analysis of discovery and replication sets). These were rs1801725 in *CASR*, rs1550532 in *DGKD*, rs780094 in *GCKR*, rs7336933 near *KIAA0564* and *DGKH*, rs10491003 (closest gene *GATA3*), rs7481584 in *CARS* and rs1570669 near *CYP24A1* (**Table 1**). Regional association plots are presented in **Figure S3**. Details on the seven SNPs that did not replicate are presented in **Table S2**. Association results for serum calcium in Caucasians for all SNPs with P value<5*E-5 are listed in **Table S3**. In a secondary analysis, all SNPs identified in the primary analysis showed consistent and significant association with serum calcium adjusted for serum albumin (**Table S4, Figure S4**), as well as an excess of association signals beyond those expected by chance (**Figure S5**); no additional locus was identified using albumin-corrected serum calcium (**Table S5**).

Copy number variations (CNVs) and eQTL analyses

We found no significant association of the 7 replicated SNPs known to provide reliable tags for copy number variations (CNVs) in people of European-descent from the Hypergene dataset. For all the SNPs, the calculated correlation was below 0.002. We also explored a list of SNPs tagging CNVs from the GIANT consortium. Out the 7 SNPs tested, only the rs1570669 was in slight linkage disequilibrium ($r^2 = 0.54$) with one SNP of the WTCCC2 list (rs927651). The corresponding SNP tags the CNVR7875.1 CNV located 455b from the SNP of interest.

For each of the 7 replicated SNPs, we identified all proxy SNPs with $r^2>0.8$ in HapMap CEU (releases 21, 22, and HapMap 3 version 2) using the online SNAP database (http://www.broadinstitute.org/mpg/snap/). This led to the identification of 40 SNPs. We then queried each of these SNPs in the eQTL database of the University of Chicago (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/). Three of the seven SNPs are in strong linkage disequilibrium with an eQTL, as illustrated in **Table S6**.

Information on genes mapping into the replicated genomic regions

Proposed functions of the genes mapping into the associated intervals (±250 kb) are in **Box 1 and** in **Table S7** for the gene-rich *GCKR* region. We report in **Table S8** the mechanism and/or location of all available biological processes, cellular components and molecular functions related to the genes mapping into the associated intervals from the AmiGo 1.8 gene ontology database. We also queried the OMIM database for each genes located within ±250 kb of the replicated loci (**Table S9**)

Validation across ethnicities

In Indian-Asians, all 7 replicated SNPs had beta-coefficients that were direction-consistent with the primary analysis and 3 were statistically significant ($P<0.05$): rs1801725 (*CASR*, $P = 1.4E\text{-}31$), rs1550532 (*DGKD*, $P = 0.002$) and rs10491003 (*GATA3*, $P = 0.009$) (**Table S10**). In Japanese, 3 SNPs had betas that were direction-consistent with the primary analysis, but only rs1801725 (*CASR*) was associated with serum calcium ($P = 0.001$) (**Table S10**).

Associations with related phenotypic traits

We conducted analyses of related bone mineral and endocrine phenotypic traits for the 7 replicated loci (**Table 2**). Several SNPs were associated ($P<0.05$) with bone mineral density (BMD) in the GEFOS consortium [6]: rs1801725 at *CASR* ($P = 0.025$; previously reported [4,5]) and rs780094 (*GCKR*) at the lumbar spine ($P = 0.006$), rs1570669 at *CYP24A1* at the femoral neck ($P = 0.04$), and rs1550532 at *DGKD* at both the lumbar spine ($P = 0.003$) and the femoral neck ($P = 0.003$). For endocrine phenotypes, rs1570669 at *CYP24A1* was associated with higher PTH concentrations ($P = 0.0005$) and rs1801725 at *CASR* with higher serum PTH concentrations ($P = 0.028$) and lower serum phosphate concentrations, as previously reported [4,5]. No SNP was associated significantly with circulating 25-OH vitamin D concentrations (all $P>0.05$) in the SUNLIGHT consortium [7].

Animal studies

We selected biologically plausible gene(s) at each locus for *in vivo* studies in a mouse model as described in **Methods'** section. We

Table 1. Genome-wide significant and replicated loci for serum calcium in Europeans.

						Discovery analysis				Replication analysis				Meta-analysis			
Markers*	chr	Position	Nearby Genes	A1	A2	N	Freq A1	Effect A1	SE	*P value*	N	Freq A1	Effect A1	SE	*P value**	N	Freq A1	Effect A1	SE	*P value*
Known locus																				
rs1801725	3	123486447	*CASR*	t	g	39400	0.15	0.069	0.004	6.5E-59	21654	0.15	0.076	0.007	3.6E-30	61054	0.15	0.071	0.004	8.9E-86
Novel loci																				
rs1550532	2	233929587	*DGKD*	c	g	39400	0.31	0.018	0.003	4.6E-08	21598	0.31	0.019	0.005	0.0002	60998	0.31	0.018	0.003	8.2E-11
rs780094	2	27594741	*GCKR*	t	c	39400	0.41	0.020	0.003	3.7E-11	21558	0.42	0.008	0.005	0.049	60958	0.42	0.017	0.003	1.3E-10
rs10491003	10	9368657	*GATA3*	t	c	38361	0.09	0.027	0.006	1.6E-06	21679	0.10	0.028	0.008	0.0003	60040	0.09	0.027	0.005	4.8E-09
rs7481584	11	2985665	*CARS*	a	g	39400	0.29	−0.021	0.003	9.2E-10	21611	0.30	−0.013	0.005	0.008	61011	0.30	−0.018	0.003	1.2E-10
rs7336933	13	41457076	*DGKH; KIAA0564*	a	g	39400	0.15	−0.023	0.004	1.6E-07	21528	0.14	−0.022	0.007	0.0009	60928	0.15	−0.022	0.004	9.1E-10
rs1570669	20	52207834	*CYP24A1*	a	g	39400	0.66	−0.018	0.003	4.0 E-08	21566	0.66	−0.020	0.005	4.5E-05	60966	0.66	−0.018	0.003	9.1E-12

P values are corrected for inflation using genomic control. Replication criteria: overall genome-wide significance ($P<5E-8$) and one-sided replication $P<0.05$. I^2 was zero for rs1801725, rs1550532, rs10491003, rs7336933 and rs1570669 (I^2 $P>0.20$). For rs780094 and rs7481584, I^2 were 0.79 and 0.43 with I^2 P 0.03 and 0.19, respectively. For these latter SNPs, sample size weighted meta-analysis P values were 2.93E-10 and 2.03E-10, respectively.
Chr, chromosome. Effect A1 = beta regression coefficient for allele A1; SE, standard error.
*one-sided P values.
doi:10.1371/journal.pgen.1003796.t001

first analyzed gene expression in the three primary calcium-handling organs: duodenum, kidney and bone (tibia). *CASR* for the rs1801725 locus, *DGKD* for the rs1550532 locus, *GATA3* for the rs10491003 locus, *CARS*, *NAP1L4* and *CDKN1C* for the rs7481584 locus, *DGKH* and *KIAA0564* for the rs7336933 locus, were expressed in all organs, whereas *CYP24A1* (rs1570669 locus) was solely, and *PHLDA2* (rs7481584 locus) mainly, expressed in the kidney (**Figure 2**). No significant expression of *GCKR* (rs780094 locus) was observed in any organ tested, which is of interest considering the strong attenuation of the association of rs780094 with serum calcium after adjustment for albumin (**Table S4**). In micro-dissection of nephron segments [8,9], *DGKD*, *DGKH*, *CARS*, *KIAA0564* and *CYP24A1* were primarily transcribed in the proximal tubule, *CASR* in the thick ascending limb, and *GATA3* predominantly in the distal nephron and collecting duct (**Figure 3**).

In order to determine regulation of gene expression by calcium intake, we measured gene expression levels in mice fed low and high calcium diets (0.17% vs. 1.69% calcium) for one week, with normal diet as control (0.82%) (**Figure 4 and Table S11**). In the kidney, both *DGKD* and *DGKH* were upregulated in response to low calcium diet ($P \leq 0.05$; **Figure 4**). In the tibia, *CASR* was markedly upregulated in response to low calcium diet (2.5-fold increased expression), as were *GATA3*, *KIAA0564* and *CARS* ($P \leq 0.05$ for all; **Figure 4**), findings that suggest regulation by $1,25(OH)^2$-D. *DGKD* and *DGKH* were upregulated in the tibia in response to high *and* low calcium diet ($P \leq 0.05$ for all; **Figure 4**). The expression in duodenum of the majority of genes was not modified by dietary calcium, with the exception of *NAP1L4* and *CDKN1C*.

Discussion

We have identified and replicated one known and six new loci for serum calcium near genes linked to bone metabolism and endocrine control of calcium. Of these, 4 loci (*DGKD*, *GCKR*, *CASR*, and *CYP24A1*) were nominally associated with BMD in the general population. In supporting mouse studies, we demonstrate expression of several of these genes in tibia, and show regulation of gene expression in response to dietary calcium intake. We also demonstrate expression in nephron segments known to regulate calcium homeostasis. Taken together, these results shed new light on the genetics of calcium balance.

The vast majority of total body calcium is bound in the skeleton as hydroxyapatite and other calcium-phosphate complexes [10]. Apart from providing skeletal strength, bone serves as a calcium reservoir to maintain tightly controlled circulating concentrations vital to cellular signaling, muscle contraction and coagulation [10]. However, the genetic basis of the dynamic cross talk that occurs between these compartments is poorly understood. Our results advance our understanding in this area. Eight genes identified in the GWAS are constitutively expressed in bone and are regulated in response to dietary calcium, in particular low calcium diet, whereas no clear change was observed in kidney or duodenum. This bone reactivity in response to dietary calcium intake is consistent with what was recently reported for *CASR* [11]. Further, of the eight genes expressed in bone and regulated in response to dietary calcium, we show that rs1550532 (*DGKD*) and rs1801725 (*CASR*) are associated with BMD in humans, the primary determinant of fracture risk.

The A allele of rs1570669 (*CYP24A1* locus) was associated with reduced BMD at the femoral neck although *CYP24A1* was not found to be expressed in bone in mice experiment, which suggests an indirect role in bone mineralization. This may occur via its

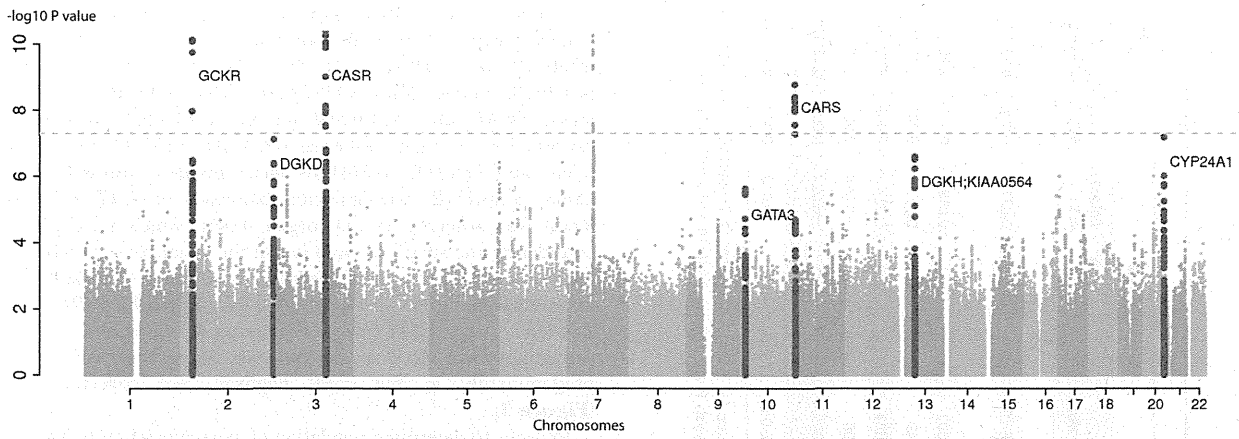

Figure 1. Genome-wide association for serum calcium in discovery analysis in Europeans. Manhattan plot showing $-\log_{10}$(P values) for all SNPs in the discovery GWAS for uncorrected serum calcium in Europeans (N = 39,400), ordered by chromosomal position. The plot is truncated at −log10 P values of 10 (truncated $-\log_{10}$P values for GCKR and CASR). The values correspond to the association of uncorrected serum calcium, including age and sex as covariates in the model as well as study-specific covariates if needed. The gene closest to the SNP with the lowest P value is listed at each locus. Six loci reached genome-wide significance (*P*<5E-08) at discovery analysis (*GCKR, DGKD, CASR, VKORC1L1* (in grey on chromosome 7), *CARS* and *CYP24A1*. The seven loci that reached genome-wide significance at the combined analysis following replication are highlighted in red (*GCKR, DGKD,* CASR, *GATA3, CARS, DGKH*-KIAA0564 and *CYP24A1*).
doi:10.1371/journal.pgen.1003796.g001

documented role in vitamin D metabolism, discussed below, and/or its association with higher PTH concentrations identified in the present analysis.

We observed specific expression patterns of several genes in the mouse nephron: *DGKD*, *DGKH*, *CARS*, *KIAA0564* and *CYP24A1* were primarily transcribed in the proximal tubule, *CASR* expression was mostly localized to the thick ascending limb, whereas *GATA3* was predominantly found in the distal part of the nephron and the collecting duct. This pattern of expression in segments known to be involved in calcium reabsorption suggests a role in renal calcium handling and is consistent with previous exploratory transcriptome analyses in humans and mice [12,13]. Both *DGKD* and *DGKH* were significantly upregulated in the kidney in response to low calcium diet, suggesting specific involvement of these genes in renal calcium handling.

Several of the newly identified loci harbor genes linked to the hormonal control of serum calcium. First, the association of *CASR* with PTH concentrations is consistent with its known role in PTH signaling. Second, several lines of evidence implicate rs1570669 (*CYP24A1*) in the vitamin D pathway: its association with serum calcium and PTH concentrations, its selective expression in the proximal tubule where 1,25(OH)2-D metabolism occurs, and that loss-of-function *CYP24A1* mutations cause vitamin D-induced hypercalcemia in children (idiopathic infantile hypercalcemia). Third, we identified variants linked to 2 chromosomally distinct isoforms of diacylglycerol kinase, part of the phosphoinositol second messenger system, that may interact with each other at the protein level [14,15].

Strengths of this study are the large sample size and consistent mouse studies to support the statistical associations and advance our knowledge of the biology at these loci. Human and mice largely share physiological processes linked to calcium metabolism, including tissue-specific gene expression. Limitations include the lack of a direct marker of bone remodeling and the potential for bias in gene selection for experimental follow-up. Mice may display subtle differences in the regulation of the genes tested compared to humans.

We have identified and replicated one known and six new loci for serum calcium near genes linked to bone metabolism and endocrine control of serum calcium. Supporting experimental mouse studies suggest a role for dietary calcium in bone-specific gene expression. Further work is needed to identify the causal variants and to understand how they influence calcium homeostasis.

Materials and Methods

Ethics statement

In each human study, the local institutional review board approved the study and participants signed written informed consent, including for DNA analyses. The experimental protocol in mice was approved by the local veterinarian authorities and fulfilled Swiss federal regulations for experiences with animals.

Participating studies (human data)

Discovery and replication cohorts. A list of all discovery and replication studies, their sample size, mean serum calcium levels, age and serum albumin as well as proportion of women can be found in **Table S1**. We replicated findings using *de novo* genotyping in the Bus Santé Study and *in silico* data in all other cohorts. In most studies, serum calcium was measured using a colorimetric assay. The size of discovery tables varied from 488 to 9,049 for a total of 39,400 participants. A detailed description of the characteristics of discovery and replication cohorts, including laboratory method for serum calcium measurement, can be found in **Table S12**.

Genotyping

Detailed information on the genotyping plateforms and data cleaning procedures for each discovery and replication cohort can be found in **Table S13**. *De novo* replication genotyping was perfomed in 4670 participants to the Bus Santé Study using KASPar v4.0 after whole genome amplification by primer extension pre-amplification (PEP) using thermostable DNA polymerases.

Table 2. Look-ups of serum calcium loci with related phenotypes: bone mineral density in the GEFOS dataset [6] and endocrine phenotypes from the SHIP, SHIP Trend and SUNLIGHT [7] datasets.

			Lumbar bone density			Femoral bone density			Serum phosphorus			25OH Vitamin D			Parathyroid hormone		
Markers	**Gene**	**A1**	**N**	**Effect A1**	**SE**	***P value***	**N**	**Effect A1**	**SE**	***P value***	**N**	**Effect A1**	**SE**	***P value***	**N**	**zscore**	***P value***	**N**	**Effect A1**	**SE**	***P value***
rs1801725	*CASR*	T	**31791**	**−0.029**	**0.013**	**0.03**	32948	−0.011	0.012	0.4	16190	**−0.038**	**0.008**	**3.4E-07**	22537	−0.426	0.7	4181	**0.031**	**0.014**	**0.03**
rs1550532	*DGKD*	C	**31681**	**−0.028**	**0.009**	**0.003**	**32845**	**−0.025**	**0.009**	**0.003**	16190	**−0.013**	**0.006**	**0.03**	20371	−0.888	0.4	4181	**0.032**	**0.010**	**0.002**
rs780094	*GCKR*	T	**31783**	**−0.024**	**0.009**	**0.006**	32946	−0.009	0.008	0.3	16190	**0.011**	**0.005**	**0.03**	22520	−0.699	0.5	4181	0.0002	0.010	1.0
rs10491003	*GATA3*	T	31797	0.007	0.016	0.6	32740	0.015	0.015	0.3	16190	−0.001	0.010	0.9	22543	−1.328	0.2	4181	0.018	0.018	0.3
rs7481584	*CARS*	A	31667	0.013	0.009	0.2	32948	0.006	0.009	0.5	16190	0.011	0.006	0.08	20366	−1.630	0.1	4181	−0.006	0.011	0.6
rs7336933	*DGKH*	A	30992	−0.006	0.013	0.7	32152	0.000	0.012	1.0	16190	0.0115	0.008	0.1	20437	0.648	0.5	4181	−0.010	0.013	0.4
rs1570669	*CYP24A1*	A	31739	−0.004	0.009	0.7	32900	**−0.017**	**0.009**	**0.04**	16190	0.0040	0.006	0.5	20385	0.144	0.9	4181	**0.035**	**0.010**	**0.0005**

NA, not available. P values<0.05 were considered as statistically significant. A1, effect allele. β, regression coefficient for allele A1, SE, standard error. P, two-sided P value. Zscore, z score.

doi:10.1371/journal.pgen.1003796.t002

Statistical analyses for the genome-wide association meta-analysis

In each discovery study, genotyping was performed using a genome-wide chip and nearly 2.5 million SNPs were genotyped or imputed using the HapMap CEU panels release 22 or 21 as the reference. Each study applied quality control before imputation. Detailed imputation information is provided in **Table S13**. Each SNP was modeled using an additive genetic effect (allele dosage for imputed SNPs), including age and sex as covariates in the model as well as study-specific covariates if needed (e.g. principal components, study center). The primary dependent variable in each discovery study was untransformed and uncorrected serum calcium expressed in mg/dL. Beta regression coefficients and standard errors were used with at least 5 decimal places. For secondary analyses, albumin-corrected serum calcium was computed using the following formula: ([4-plasma albumin in g/dL] ×0.8+serum calcium in mg/dL) and the same model as for the primary analyses was used. Each file of genome-wide summary statistics underwent extensive quality control prior to meta-analysis both for primary and secondary analyses, including (1) boxplots of all beta coefficients, as well as all standard errors multiplied by the square-root of the sample size, for each study separately; (2) the range of P values, MAF, imputation qualities, call rates and Hardy-Weinberg equilibrium P values and (3) QQ plots. In addition, we checked the direction and magnitude of effect at the previously reported rs1801725 *CASR* variant. Genome-wide meta-analyses were conducted in duplicate by two independent analysts. For each SNP, we used a fixed effect meta-analysis using inverse-variance weights as implemented in the meta-analysis utility Metal [16]. Results were confirmed by a z-score based meta-analysis. Data were available for 2,612,817 genotyped or imputed autosomal SNPs for the primary and secondary analyses. After the meta-analysis, genomic control correction was applied (λGC was 1.03 for both uncorrected and corrected serum calcium). Our pre-specified criterion to declare genome-wide significance was P value<5E-8 to account for 1 million independent tests according to the Bonferroni correction. We choose to move forward for replication all SNPs with discovery P value<1E-7 in the European sample or genome-wide significant SNP in the overall sample that included Indian Asians. To choose a single SNP per genome-wide associated region for replication, we merged all SNPs within 1 Mb region and selected the lowest P value for each region. Altogether, fourteen SNPs were moved forward for replication. Up to 17,205 participants contributed information to the replication analyses *in silico* and 4,670 participants provided data for *de novo* genotyping. We used fixed-effects inverse-variance weighted meta-analysis to combine discovery and replication meta-analysis results. Replication was considered as present whenever a combined P value<5E-8 together with an effect-concordant one-sided replication P value<0.05 were obtained.

Data for look-ups of serum calcium loci with related phenotypes

We conducted look-ups for femoral and lumbar bone density in the GEnetic Factors of OSteoporosis (GEFOS) dataset [17]. Bone mineral density (BMD) is used in clinical practice for the diagnosis of osteoporosis and bone density at different skeletal sites is predictive of fracture risk. BMD was measured in all cohorts at the lumbar spine (either at L1–L4 or L2–L4) and femoral neck using dual-energy X-ray absorptiometry following standard manufacturer protocols [17]. Serum phosphorus was looked up from a

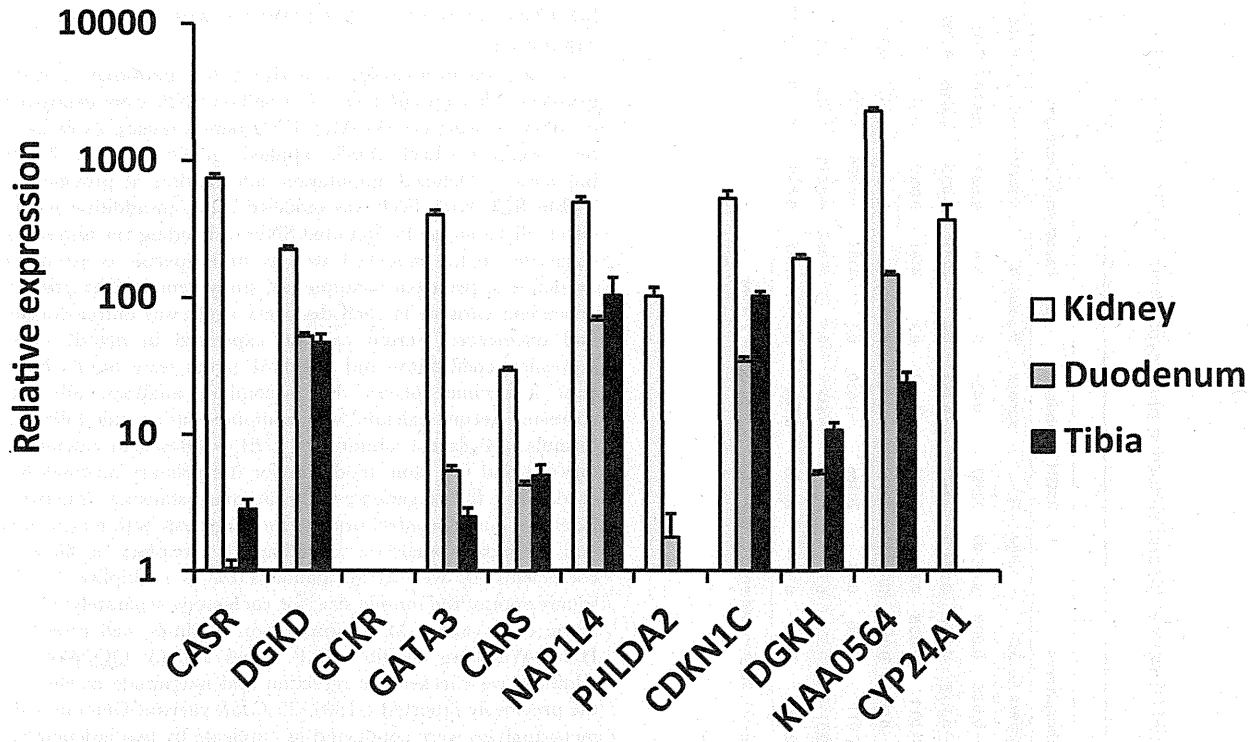

Figure 2. Relative mRNA expression of replicated genes in three calcium-transporting tissues (kidney, duodenum, tibia). The expression (based on delta CT [cycle threshold] normalized to actin) of the selected genes is compared to the expression of the *CASR* gene in the duodenum, thereby providing a relative expression. Cut-off was set at delta CT≤15. Data are means ± standard error of the mean (SEM) of values obtained from 5 mice fed a normal diet. *GCKR* was not expressed.
doi:10.1371/journal.pgen.1003796.g002

previously published GWAS meta-analysis, including 16,264 participants of European ancestry [18]. Serum phosphorus concentrations were quantified using an automated platform in which inorganic phosphorus reacts with ammonium molybdate in an acidic solution to form a colored phosphomolybdate complex [18]. The 25-hydroxyvitamin D was looked-up in the SUNLIGHT consortium [7], which includes data from 33,996 individuals of European descent from 15 cohorts. 25-hydroxyvitamin D concentrations were measured by radioimmunoassay, chemiluminescent assay, ELISA, or mass spectrometry

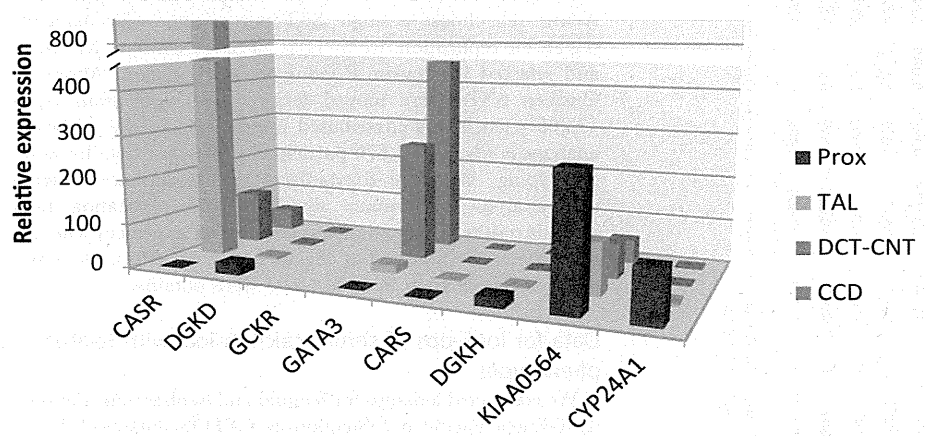

Figure 3. Relative mRNA expression of identified genes in kidney tubule segments. The renal tubular segments analyzed were the proximal tubule (PROX), the thick ascending limb of the loop of Henle (TAL), the distal convoluted tubule and connecting tubule (DCT-CNT), and the cortical collecting duct (CCD). The expression (based on the delta CT [cycle threshold]) of the selected genes is compared to the expression of the *CASR* gene in the PROX, thereby providing a relative expression. Data are means of values obtained from 3 mice fed a normal diet. *GCKR* was not expressed.
doi:10.1371/journal.pgen.1003796.g003

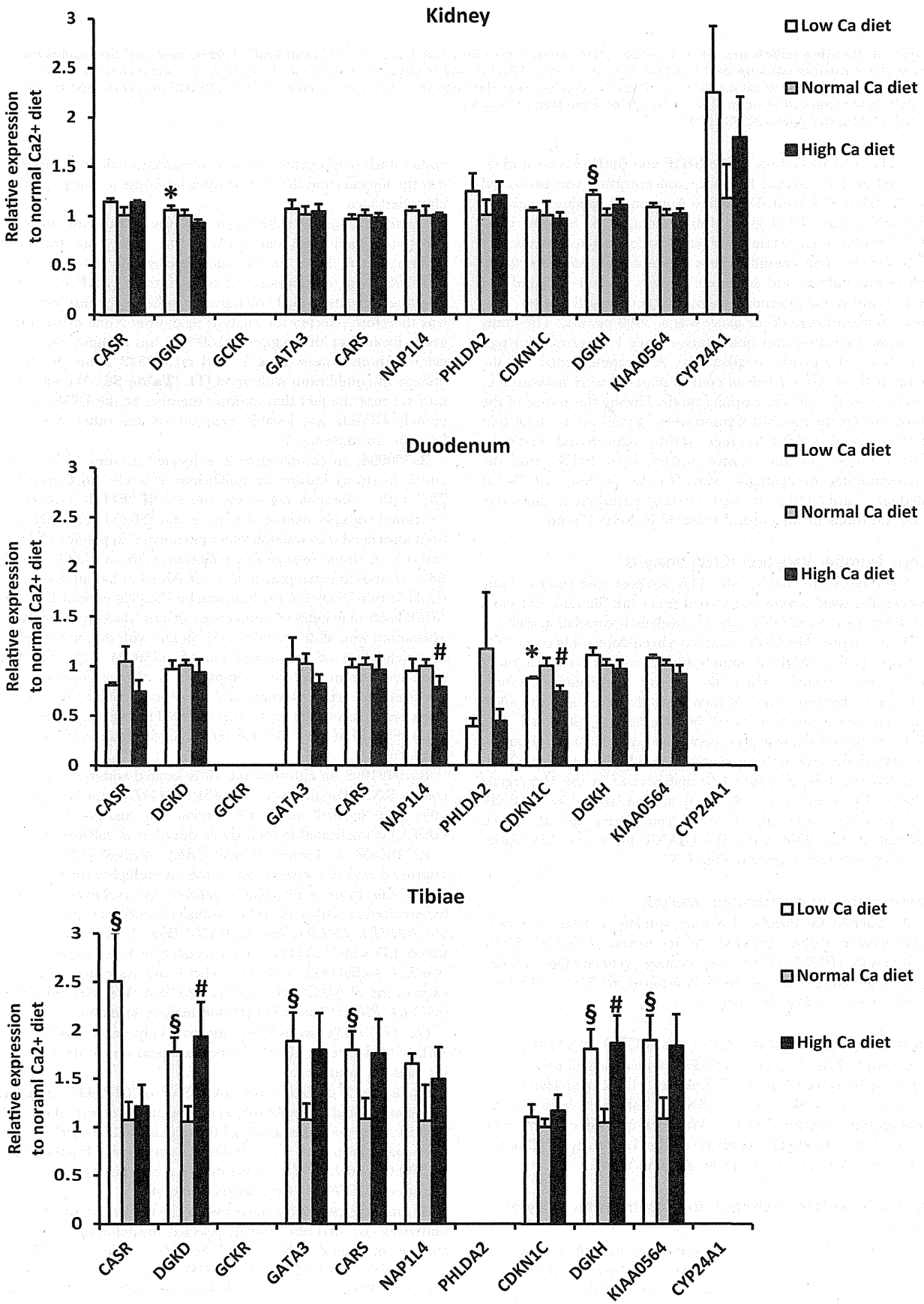
Kidney
Relative expression to normal Ca2+ diet
3
2.5
2
1.5
1
0.5
0
Low Ca diet
Normal Ca diet
High Ca diet
*
§
CASR
DGKD
GCKR
GATA3
CARS
NAP1L4
PHLDA2
CDKN1C
DGKH
KIAA0564
CYP24A1
Duodenum
Relative expression to normal Ca2+ diet
Low Ca diet
Normal Ca diet
High Ca diet
#
*
#
CASR
DGKD
GCKR
GATA3
CARS
NAP1L4
PHLDA2
CDKN1C
DGKH
KIAA0564
CYP24A1
Tibiae
Relative expression to noraml Ca2+ diet
Low Ca diet
Normal Ca diet
High Ca diet
§
§
#
§
§
§
#
§
CASR
DGKD
GCKR
GATA3
CARS
NAP1L4
PHLDA2
CDKN1C
DGKH
KIAA0564
CYP24A1

Figure 4. Relative mRNA expression of identified genes from mice fed a low (0.17%) and high (1.69%) calcium diet compared to mice fed a normal calcium diet (0.82%). Data are means± SEM of values obtained from 5 mice for each diet group. Expression levels were normalized to actin. Statistical significance of the difference between diets was calculated using unpaired t-test. *: $P \leq 0.05$ (low compared to high); §: $P \leq 0.05$ (low compared to normal); # $P \leq 0.05$ (high compared to normal).
doi:10.1371/journal.pgen.1003796.g004

[7]. PTH was looked-up in the SHIP and SHIP-Trend studies. The serum parathyroid hormone concentration was measured on the IDS-iSYS Multi-Discipline Automated Analyser with the IDS-iSYS Intact PTH assay (Immunodiagnostic Systems Limited, Frankfurt am Main, Germany) according to the instructions for use. This chemiluminescence immunoassay detects the full-length parathyroid hormone (amino acids 1–84) and the large parathyroid hormone fragment (amino acids 7–84). The measurement range of the assay was 5–5000 pg/mL. The limits of blank, detection and quantitation were 1.3 pg/mL, 1.4 pg/mL, and 3.6 pg/mL, respectively. As recommended by the manufacturer, three levels of control material were measured in order to verify a decent working mode. During the course of the study, the coefficients of variation were 14.02% at low, 6.64% at medium, and 6.84% at high serum parathyroid hormone concentrations in the control material in SHIP and the corresponding percentages were 16.8% at low, 10.7% at medium, and 9.0% at high serum parathyroid hormone concentrations in the control material in SHIP-Trend.

Copy Number Variation (CNV) analysis

The Hypergene dataset (a 4206 samples case-control study concerning hypertension genotyped using the Illumina 1M chip) has been used to call CNVs and to check their correlation with the SNPs of interest. The CNVs calls have been done using pennCNV software [19]. A SNP by sample matrix with the copy number status was created. Then the square correlation (Pearson correlation) between value of each SNP of interest and the SNPs copy number status in a +/−2 Mb region was calculated. The SNPs of interest for which no correspondence has been found in the Hypergene dataset have been replaced by the closest SNPs in high linkage disequilibrium (LD) and present in the Hypergene dataset. LD between the SNPs of interest and a list of SNPs tagging CNVs from the GIANT consortium has also been calculated. The SNPs from the GIANT list are in LD higher than 0.8 with their corresponding CNV.

Gene ontology classification analysis

We queried the AmiGo 1.8 gene ontology database for each gene located within ±250 kb of the seven replicated SNPs, including rs1801725 (*CASR*). (http://amigo.geneontology.org/cgi-bin/amigo/go.cgi, last accessed November 6, 2012). We used *Homo sapiens* as a filter for species.

Expression quantitative trait locus (eQTL) Analyses

For each of the 7 replicated SNPs, we identified all proxy SNPs with $r^2>0.8$ in HapMap CEU (releases 21, 22, and HapMap 3 vers. 2) using the online SNAP database (http://www.broadinstitute.org/mpg/snap/). We then queried each of these 40 SNPs in the eQTL database of the University of Chicago (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/).

Rationale for gene selection for experimental analyses in mouse

The rs1801725 SNP encodes a missense variant in exon 7 of the CASR gene leading to an alanine to serine substitution (A986S). Given the key physiological role of CASR in calcium homeostasis (monogenic disorders of calcium balance), this gene was the logical candidate for analysis in mouse at this previously identified locus.

For the 6 newly identified loci, the precise rationale for gene selection varied from one locus to the other, but the main criteria was to focus on the most biologically relevant gene. Rs1550532 on chromosome 2 is an intronic SNP of *DGKD*, which was the most likely biological candidate for this locus and was therefore selected for analysis in mouse. None of the other genes located in this region (±250 Kb) has a known link with calcium homeostasis (**Box 1**) and rs1550532 is not in strong linkage disequilibrium with an eQTL (**Table S6**). We also took into account the fact that another member of the DGK family, namely DGKH was located near one of the other replicated loci, on chromosome 13.

Rs780094, on chromosome 2, is located in intro 16 of *GCKR* and is in strong linkage disequilibrium ($r^2 = 0.93$) in Caucasians [20], with a common non-synonymous SNP (P446L, rs1260326) associated with glucokinase activity *in vitro* [20,21]. This SNP has been associated with multiple other phenotypes in previous GWAS and it is in strong linkage disequilibrium with an eQTL (**Table S6**). Previous fine mapping analysis of this locus has attributed the signal from rs780094 to the functional rs1260326 variant [20]. *The GCKR* locus may indirectly influence calcium concentrations via its association with albumin levels [22]. In line with this, we observed an attenuation of the association of rs780094 with albumin-corrected serum calcium compared to the association with uncorrected serum calcium and we found *GCKR* not to be expressed in any of the key organs involved in calcium homeostasis that we tested in mice. We selected *GCKR* for analysis in mouse at this locus.

Rs10491003 on chromosome 10 is located within a long non-coding RNA. For this locus, we selected *GATA3*, the nearest and only gene located within this region, for analysis in mouse. GATA3 is implicated in monogenic disorders of calcium balance.

Rs7481584 is located within *CARS* (intronic SNP) in an imprinted region known to play a role in multiple cancers, which makes this locus a plausible candidate for malignancy-related hypercalcemia. Other plausible biological candidates in this locus are *NAP1L4, PHLDA2 and CKDN1C* (**Box 1**). Rs7481584 is in strong LD with 2 eQTLs, one associated with the expression of *NAP1L4* (rs2583435) and the other one associated with the expressions of *SLC22A18* and *SLC22A18AS*. We selected *CARS, NAP1L4, PHLDA2 and CKDN1C* for analyses in mouse.

For rs7336933, we selected the two only genes (*DGKH* and *KIAA0564*) located under this association peak on chromosome 13 for analyses in mouse.

Finally, rs1570669 is an intronic SNP of *CYP24A1*, a strong biological candidate implicated in monogenic disorders of calcium balance. The two other genes of this region (*BCAS1* and *PFDN4*) have no known link with calcium homeostasis. Furthermore, rs1570669 and *PFDN4* are separated by a recombination hot spot. We selected *CYP24A1* for analysis in mouse.

As animal experiments started while the replication process was underway, we had also initially selected the following genes for analysis in mouse: RSG14 and *SLC34A1* at locus rs4074995 (discovery P value = 2.4E-07), *VKORC1L1* at locus rs17711722 (discovery P value = 2.8E-11), *PYGB* at locus rs2281558 (discovery

Box 1. Genes Located within Replicated Loci for Serum Calcium

We here summarize the information on genes located within ±250 kb from the top SNP at each locus. Because it is a gene dense region, details of genes located in the *GCKR* genomic region are presented in **Table S4**.

Chromosome 2, locus rs1550532

DGKD rs1550532 is an intronic SNP located near the 5′UTR region of *DGKD*. DGKD encodes diacylglycerol kinase delta, a member of the diacylglycerol kinase (DGK) enzyme family. Alternative splicing of the DGKD gene results in two isoforms, which differ in their expression profiles and regulatory mechanisms [24]. DGKs play an important role in signal transduction by modulating the balance between the diacylglycerol (DAG) and phosphatidic acid (PA), important second messengers in signaling cascades. Recent findings suggest that DAG is involved in calcium signaling in parathyroid cells [25]. CASR signaling influences intracellular DAG levels in cardiomyocytes [26].

SAG encodes S-antigen (also called arrestin), a soluble photoreceptor protein expressed in the retina and pineal gland. Mutations in this gene are associated with Oguchi disease (OMIM#258100), a rare autosomal recessive form of night blindness. Arrestin is a calcium-binding protein that plays an important role in phototransduction.

ATG16L1 encodes autophagy related 16-like 1 protein, part of a complex involved in autophagia. Mutations in this gene are responsible for inflammatory bowel disease 10 (OMIM # 611081). There is no known direct link with calcium signaling.

SCARNA5 and SCARNA6 encode small Cajal body-specific RNAs 5 and 6, which are small nuclear RNAs, belonging to non-coding RNAs involved in the RNA-processing machinery. There is no known direct link with calcium signaling.

USP40 encodes ubiquitin specific peptidase 40. USP40 functions as a deubiquinating enzyme involved in the degradation of unwanted intracellular proteins in eukaryocytic cells. There is no known direct link with calcium signaling.

INPP5D encode inositol polyphosphate-5-phosphatase, expressed in hematopoietic cells. This protein regulates myeloid cell proliferation. The presence of a recombination peak between this gene and rs1550532 makes it an unlikely candidate for this signal.

Chromosome 10, locus rs10491003

rs10491003, located within a long non-coding RNA with *GATA3* as its nearest gene may influence the expression of *GATA3* [27].

GATA3: *GATA3* encodes a GATA transcription factor involved in T cell lymphopoiesis [28], renal and vestibular morphogenesis [29,30], and parathyroid gland development [31]. *GATA3* haploinsufficiency causes hypoparathyroidism and hypocalcemia in the autosomal dominant HDR syndrome (hypoparathyroidism, sensorineural deafness and renal dysplasia) (OMIM#146255) [32,33]. Although *GATA3* is the closest gene to rs10491003, this variant lies 1.2 Mbp downstream from that gene. However, *GATA3* has a very large flanking regulatory region - greater than 450 kbp - [34] and mammalian enhancers may lie more than 1 Mbp away from the gene they regulate [35]. *GATA3* may play a role in preserving high degree of differentiation of parathyroid gland and of calcium transporting epithelia [36].

Chromosome 11, locus rs7481584

This region is located in the imprinted gene domain of 11p15.5, an important tumor suppressor gene region [37].

CARS: rs7481584 is an intronic SNP of *CARS*. *CARS* encodes a cysteinyl-tRNA synthetase and is located within the imprinted gene domain of 11p15.5. This region is linked to Beckwith-Wiedemann syndrome, which is associated with hypocalcemia and hypercalciuria.

NAP1L4 encodes nucleosome assembly protein 1-like 4, a member of the nucleosome assembly protein, potentially involved in histone chaperoning and ubiquitously expressed. NAP1L1 and NAP1L4 have been recently identified as being involved in the regulation of DGKH nucleocytoplasmic shuttling [38]. A link with calcium homeostasis could be possible via the DGKs pathway.

PHLDA2 encodes pleckstrin homology-like domain, family A, member 2. This gene has been recently highlighted as potentially relevant for osteoporosis on the basis of a bioinformatics pathway analysis approach [39]. Imprinting of this gene appears to play a role in fetal growth, including fetal bone growth, birth weight and bone mass in childhood.[40,41,42,43] In cancer, PHLDA2 is activated by parathyroid hormone-like hormone (PTHLH) [44]. PTHLH is associated with malignancy-related hypercalcemia [45], lactation [46], the expression of PHLDA2 is upregulated in osteosarcoma progression [47].

OSBPL5 encodes oxysterol binding protein-like 5, an intracellular lipid receptor involved in cholesterol balance. There is no known direct link with calcium homeostasis.

MRGPRE and MRGPRG encode MAS-related G-protein-coupled receptors, member E and G. This family of receptors is expressed in nociceptive sensory neurons. There is no known direct link with calcium homeostasis.

C11orf36 encodes MRGPRG antisense RNA 1. Little is known about this gene.

SNORA54 encodes small nucleolar RNA, H/ACA box. The gene product belongs to non-coding RNAs involved in the RNA-processing machinery. There is no known direct link with calcium homeostasis.

SLC22A18 and SLC22A18AS encode solute carrier family 22, member 1 and solute carrier family 22, member 1 antisense. SLC22A18 is an organic cation transporter. Mutations in *SLC22A18* have been found in several cancers. There is no known direct link with calcium homeostasis.

CDKN1C encodes cyclin-dependent kinase inhibitor 1C (p57, Kip2), a protein involved in cell-cycle progression. This imprinted gene is responsible for the IMAGe syndrome (OMIM#300290) characterized by intrauterine growth restriction, metaphyseal dysplasia, delayed bone aging, adrenal hypoplasia congenital, genital anomalies, and sometimes hypercalciuria [48].

KCNQ1 encode potassium voltage-gated channel, KQT-like subfamily, member 1. ***KCNQ1OT1*** represents KCNQ1 opposite strand transcript 1 and is an unspliced long non-coding RNA, which regulates the transcription of many target genes. Mutations in *KCNQ1* are associated with hereditary long and short QT syndromes (OMIM#192500 & 609621), Jervell and Lange-Nielsen syndrome (OMIM#220400), familial atrial fibrillation (OMIM#607554), type 2 diabetes. *KCNQ1* is also imprinted in a tissue-specific manner. There is no known direct link with calcium homeostasis.

Chromosome 13, locus rs7336933

DGKH encodes diacylglycerol kinase eta, a member of the diacylglycerol kinase (DGK) enzyme family. See *DGKD* (above) for discussion.

KIAA0564: this gene encodes a large uncharacterized protein containing a putative ATP-ase domain. The sequence of this gene is conserved across a large array of organisms,

from humans to mouse, zebrafish and to *C. elegans*, which suggests an important biological function. Yet, little is known on the nature of the function of this gene so far.

Chromosome 20, locus rs1570669

CYP24A1: rs1570669 is an intronic SNP of *CYP24A1*. *CYP24A1* encodes a cytochrome P450 enzyme that hydroxylates 1,25-$(OH)_2$D, into metabolites targeted for degradation and appears to be one of the central regulator of 1,25-$(OH)_2$-D metabolism. CYP24A1 is highly regulated by its own substrate 1,25$(OH)_2$-D, as well as by PTH [49,50], serum phosphate and fibroblast growth factor-23 (FGF-23) [51,52,53]. Sequence variants of *CYP24A1* impacting on 1,25$(OH)_2$-D metabolism have been described recently and explain the strong heritability of 1,25$(OH)_2$-D concentrations.

BCAS1 encodes breast carcinoma amplified sequence 1, considered as an oncogene. *BCAS1* is highly differentially expressed in some cancers. However, there is no direct link with calcium homeostasis.

PFDN4 encodes prefoldin subunit4. Prefoldin is a chaperone complex involved in polypeptide folding. There is no known link of this gene with calcium homeostasis.

P value = 6.4E-07), *CD109* at locus rs9447004 (discovery P value = 8.1E-06). No gene was selected for the rs2885836 and rs11967485 and rs12150338 loci in the absence of obvious candidate. Results for these unreplicated loci can be found in **Figures S6, S7 and S8**. We present these results for quality control purposes: *SLC34A1* (also known as NAPI-3 or NPT2), which encodes solute carrier family 34 (sodium phosphate), member 1, was expressed in the kidney, but neither in duodenum nor in bone, as expected based on current knowledge on this phosphate transporter. In the kidney *SLC34A1* was mainly expressed proximally and *SLC34A1* expression was upregulated under low calcium diet, which is in line with the known function of this gene.

Mouse experiments

Five C57bl/6 mice (Janvier) per group were fed, for one week, three different diets in which the percentage of calcium were 0.17% (low calcium diet), 0.82% (normal calcium diet) and 1.69% (high calcium diet) and had free access to water. 12:12 hours light/dark alternance was imposed. At the end of the week of the specific diet, spot urine were collected and mice were anesthetized. Blood was collected by retro-orbital puncture. Organs were immediately harvested and snap frozen. RNA was extracted using Trizol (Invitrogen) and reversed transcribed with PrimeScriptTM RT reagent Kit (Takara Bio Inc). Calcium, sodium, phosphate and creatinine in plasma and urine were analyzed at the central lab of the Lausanne University hospital using a Cobas-Mira analyzer (Roche).

Microdissection. A separate set of three mice was kept under normal calcium diet. Proximal Tubule (Prox), thick ascending limb of the loop of Henle (TAL), distal convoluted tubule and connecting tubule (DCT-CNT) and cortical collecting duct (CCD) were isolated by microdissection of the left kidney after the mice were perfused with Liberase TM (Roche Diagnostics) [23]. RNA was extracted from the above mentioned tubules following TRI Reagent Solution protocol (Applied Biosystems) and purified with RNeasy Micro Kit (Qiagen). Reversed transcription was performed with PrimeScriptTM RT reagent Kit (Takara Bio Inc). Quantitative PCRs were performed (7500 Software v 2.0.4.) using TaqMan gene expression assays for the different genes (Applied Biosystems) and comparative CT method was applied. Expression levels were normalised to beta actin as endogenous reference gene.

Statistics. Comparison of groups was performed using unpaired Student's t-test.

Supporting Information

Figure S1 QQ-plot of uncorrected serum calcium GWAS meta-analysis. Quantile-quantile plot showing observed p-values of the uncorrected serum calcium meta-analysis vs. expected p values by chance. The second genomic control step was applied to correct for the post meta-analysis of $\lambda = 1.03$.
(PDF)

Figure S2 Regional association plot for the CASR locus. Regional association plot showing −log10 p-values for the association of all SNPs ordered by their chromosomal position with uncorrected serum calcium at the *CASR* loci. The −log10 P value for each SNP is colored according to the correlation of the corresponding SNP with the SNP showing the lowest p-value (index SNP) within the locus using different colors for selected levels of linkage disequilibrium (r^2). Correlation structures correspond to HapMap 2 CEU.
(PDF)

Figure S3 Regional association plot for the newly identified loci. Regional association plot showing −log10 p-values for the association of all SNPs ordered by their chromosomal position with uncorrected serum calcium within the replicated loci. The −log10 P value for each SNP is colored according to the correlation of the corresponding SNP with the SNP showing the lowest p-value (index SNP) within the locus using different colors for selected levels of linkage disequilibrium (r^2). Correlation structures correspond to HapMap 2 CEU.
(PDF)

Figure S4 Manhattan plot of corrected serum calcium. Manhattan plot showing −log10 (P values) for all SNPs analyzed, ordered by their chromosomal position. The values correspond to the association of albumin-corrected serum calcium, including age and sex as covariates in the model as well as study-specific covariates if needed.
(PDF)

Figure S5 QQ-plot of corrected serum calcium. Quantile-quantile plot showing observed p-values of the corrected serum calcium meta-analysis vs. expected P values by chance in Europeans at discovery. The second genomic control step was applied to correct for the post meta-analysis of $\lambda = 1.03$.
(PDF)

Figure S6 Relative expression of genes in non-replicated loci in kidney, duodenum and tibia. The expression (based on delta CT normalized to actin) of the selected genes is compared to the expression of the *CASR* gene in the duodenum, thereby providing a relative expression. Cut-off was set at delta CT≤15. Data are means ± SEM of values obtained from 5 mice fed a normal diet.
(PDF)

Figure S7 Relative expression in segments of kidney tubules of genes located in non-replication loci. The renal tubular segments analyzed were the proximal tubule (PROX), the thick ascending

limb of the loop of Henle (TAL), the distal convoluted tubule and connecting tubule (DCT-CNT), and the cortical collecting duct (CCD). The expression (based on the delta CT) of the selected genes is compared to the expression of the CASR gene in the PROX. Data are means of values obtained from 3 mice fed a normal diet. *GCKR* was not expressed.
(PDF)

Figure S8 Relative expression of genes in non-replicated loci under various calcium diets. Data are means± SEM of values obtained from 5 mice fed a low (0.17%) and high (1.69%) calcium diet compared to mice fed a normal calcium diet (0.82%). Expression levels were normalized to actin. Statistical difference was calculated using unpaired t-test. *: P value≤0.05 (low compared to high); §: P value≤0.05 (low compared to normal); #: P value≤0.05 (high compared to normal).
(PDF)

Table S1 Characteristics of study participants in discovery and replication cohorts. Data are mean (SD) unless otherwise specified for each discovery and replication studies.
(DOCX)

Table S2 SNPs brought forward for replication that did not replicate. Chr, chromosome. A1, effect allele. A2, non-effect allele. Effect A1, regression coefficient for the A1 allele. SE, standard error. Freq A1,frequency of allele A1.
(DOCX)

Table S3 SNPs with P value<5*E-05 for uncorrected calcium in Europeans (discovery). Chr, chromosome. Position, position on build 36. A1, allele 1 (effect allele). A2, allele 2. Freq A1, frequency of allele 1. InRefGen, gene symbol if SNP is located within a specific gene.
(DOCX)

Table S4 Comparison of association with uncorrected versus corrected serum calcium. Chr, chromosome. Freq A1, frequency of allele A1. Beta, regression coefficient for the A1 allele. SE, standard error. A1, allele 1 (effect allele). Only replicated loci are included in this table.
(DOCX)

Table S5 Genome-wide significant loci for corrected calcium in Europeans (discovery). Chr, chromosome. Position, position on build 36. A1, allele 1 (effect allele). A2, allele 2. Freq A1, frequency of allele 1. InRefGen, gene symbol if SNP is located within a specific gene.
(DOCX)

Table S6 eQTL analysis for the seven genome-wide replicated loci for serum calcium. We used the online eQTL database of the University of Chicago (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/., last accessed, November 5, 2012). All eQTL were acting in *cis*.
(DOCX)

Table S7 Details on genes located in the GCKR genomic region.
(DOCX)

Table S8 Gene Ontology classification (AmiGo). Data are GO numbers, ontology and mechanism/location from the AmiGo 1.8 gene ontology database for each gene located within ±250 kb of the seven replicated SNPs, including rs1801725 (CASR).
(DOCX)

Table S9 OMIM disorders associated with the genes located within the replicated loci. This table includes all Mendelian disorders or other types of genetic disorders included in the OMIM database described for each gene located within ±250 kb of any of the six new loci and for *CASR*.
(DOCX)

Table S10 Association of replicated serum calcium loci in other ethnic groups. Chr, chromosome. Position, position on build 36. A1, allele 1 (effect allele). A2, allele 2. Freq A1, frequency of allele 1. Effect A1, regression coefficient for the A1 allele. SE, standard error. NA, not available.
(DOCX)

Table S11 Plasma and Urine electrolytes values by calcium diet in mice. Data are means ± SEM of values obtained from 3 to 5 mice. *: P value≤0.05 compared to normal or high calcium diet.
(DOCX)

Table S12 Study information.
(DOCX)

Table S13 Genotyping information for each cohort (discovery, replication and look-ups).
(DOCX)

Text S1 Study specific acknowledgements.
(DOCX)

Acknowledgments

The full list of acknowledgments for each study is provided in the Supporting Information files (**Text S1**).

Author Contributions

Conceived and designed the experiments: CMOS HWu QY KK WHLK HWa OB CSF MBoc. Performed the experiments: CMOS HWu QY KK IG AMZ AK CSt ZK TH GE LJL VG DEA LF DSS BMP OHF AHo AGU JCMW IJD JMS MP EMB PV SBe CH VV SBa IR OP JW HC JSK JCC AHa CSc MN RB UV PBM MJB JMT JMG TDS WH APM LL WM BOB BRW CG CM WHLK HWa OB CSF MBoc. Analyzed the data: CMOS HWu QY KK IG AMZ AK AT ZK YL TH JD KL AVS GE LJL VG TT GL AD FR LML LP FM JH AMac SBe CH VV JFW WZ AHo AGU FR KE SP MM FDE AMah WH APM MEK ACB CG WHLK HWa OB CSF MBoc. Contributed reagents/materials/analysis tools: IG YL VG EB DH AS DSS BMP AGU FR IJD JMS MP PV GD VV CH JSK JCC PBM MM SYS APM MEK TM WHLK HWa CSF MBoc. Wrote the paper: CMOS IG AK HWa OB CSF MBoc. Revising and reviewing the manuscript for important intellectual content: CMOS HWu QY KK IG AMZ AK CSto AT ZK MM AD WZ GE GL TT LP LML CH KL KM SP DF RS SU ACB MEK AMah FDE VG LJL AMac EB DEA CT YN MJB JMG JMT DSS BMP SBe PV VV AFW TZ MBob IK PN EMB KE JD TBH SBa DH ABS GG DR APA AR TM CM GD JMS JCC BOB BRW JH FM SHW HC APM OHF AHo AGU FR UV AHa RB WH SYS PL HH CSc PBM PG NP MC CG WM LL TDS AVS IR JFW OP IJD MP LF YL BK JSK JCMW MN WHLK HWa OB CSF MBoc.

References

1. Whitfield JB, Martin NG (1984) The effects of inheritance on constituents of plasma: a twin study on some biochemical variables. Ann Clin Biochem 21 (Pt 3): 176–183.
2. Williams PD, Puddey IB, Martin NG, Beilin LJ (1992) Platelet cytosolic free calcium concentration, total plasma calcium concentration and blood pressure in human twins: a genetic analysis. Clin Sci (Lond) 82: 493–504.

3. Hunter DJ, Lange M, Snieder H, MacGregor AJ, Swaminathan R, et al. (2002) Genetic contribution to renal function and electrolyte balance: a twin study. Clin Sci (Lond) 103: 259–265.
4. O'Seaghdha CM, Yang Q, Glazer NL, Leak TS, Dehghan A, et al. (2010) Common variants in the calcium-sensing receptor gene are associated with total serum calcium levels. Human Molecular Genetics 19: 4296–4303.
5. Kapur K, Johnson T, Beckmann ND, Sehmi J, Tanaka T, et al. (2010) Genome-wide meta-analysis for serum calcium identifies significantly associated SNPs near the calcium-sensing receptor (CASR) gene. PLoS Genet 6: e1001035.
6. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, et al. (2012) Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet 44: 491–501.
7. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, et al. (2010) Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet 376: 180–188.
8. Bibert S, Hess SK, Firsov D, Thorens B, Geering K, et al. (2009) Mouse GLUT9: evidences for a urate uniporter. Am J Physiol Renal Physiol 297: F612–619.
9. Zuber AM, Centeno G, Pradervand S, Nikolaeva S, Maquelin L, et al. (2009) Molecular clock is involved in predictive circadian adjustment of renal function. Proc Natl Acad Sci U S A 106: 16523–16528.
10. Peacock M (2010) Calcium metabolism in health and disease. Clinical journal of the American Society of Nephrology : CJASN 5 Suppl 1: S23–30.
11. Shu L, Ji J, Zhu Q, Cao G, Karaplis A, et al. (2011) The calcium-sensing receptor mediates bone turnover induced by dietary calcium and parathyroid hormone in neonates. J Bone Miner Res 26: 1057–1071.
12. Cheval L, Pierrat F, Dossat C, Genete M, Imbert-Teboul M, et al. (2011) Atlas of gene expression in the mouse kidney: new features of glomerular parietal cells. Physiol Genomics 43: 161–173.
13. Chabardes-Garonne D, Mejean A, Aude JC, Cheval L, Di Stefano A, et al. (2003) A panoramic view of gene expression in the human kidney. Proc Natl Acad Sci U S A 100: 13710–13715.
14. Murakami T, Sakane F, Imai S, Houkin K, Kanoh H (2003) Identification and characterization of two splice variants of human diacylglycerol kinase eta. J Biol Chem 278: 34364–34372.
15. Shulga YV, Topham MK, Epand RM (2011) Regulation and functions of diacylglycerol kinases. Chem Rev 111: 6186–6208.
16. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190–2191.
17. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, et al. (2012) Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nature genetics 44: 491–501.
18. Kestenbaum B, Glazer NL, Kottgen A, Felix JF, Hwang SJ, et al. (2010) Common genetic variants associate with serum phosphorus concentration. J Am Soc Nephrol 21: 1223–1232.
19. Wang K, Li M, Hadley D, Liu R, Glessner J, et al. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome research 17: 1665–1674.
20. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, et al. (2008) Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes 57: 3112–3121.
21. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, et al. (2009) The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Human molecular genetics 18: 4081–4088.
22. Franceschini N, van Rooij FJ, Prins BP, Feitosa MF, Karakas M, et al. (2012) Discovery and Fine Mapping of Serum Protein Loci through Transethnic Meta-analysis. American journal of human genetics 91: 744–753.
23. Zuber AM, Centeno G, Pradervand S, Nikolaeva S, Maquelin L, et al. (2009) Molecular clock is involved in predictive circadian adjustment of renal function. Proceedings of the National Academy of Sciences of the United States of America 106: 16523–16528.
24. Sakane F, Imai S, Yamada K, Murakami T, Tsushima S, et al. (2002) Alternative splicing of the human diacylglycerol kinase delta gene generates two isoforms differing in their expression patterns and in regulatory functions. J Biol Chem 277: 43519–43526.
25. Okada Y, Imendra KG, Miyazaki T, Hotokezaka H, Fujiyama R, et al. (2011) High extracellular Ca2+ stimulates Ca2+-activated Cl- currents in frog parathyroid cells through the mediation of arachidonic acid cascade. PLoS One 6: e19158.
26. Zheng H, Liu J, Liu C, Lu F, Zhao Y, et al. (2011) Calcium-sensing receptor activating phosphorylation of PKCdelta translocation on mitochondria to induce cardiomyocyte apoptosis during ischemia/reperfusion. Mol Cell Biochem 358: 335–343.
27. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, et al. (2012) Landscape of transcription in human cells. Nature 489: 101–108.
28. Hosoya T, Maillard I, Engel JD (2010) From the cradle to the grave: activities of GATA-3 throughout T-cell development and differentiation. Immunol Rev 238: 110–125.
29. Grote D, Souabni A, Busslinger M, Bouchard M (2006) Pax 2/8-regulated Gata 3 expression is necessary for morphogenesis and guidance of the nephric duct in the developing kidney. Development 133: 53–61.
30. Haugas M, Lillevali K, Salminen M (2012) Defects in sensory organ morphogenesis and generation of cochlear hair cells in Gata3-deficient mouse embryos. Hear Res 283: 151–161.
31. Grigorieva IV, Thakker RV (2011) Transcription factors in parathyroid development: lessons from hypoparathyroid disorders. Ann N Y Acad Sci 1237: 24–38.
32. Van Esch H, Groenen P, Nesbit MA, Schuffenhauer S, Lichtner P, et al. (2000) GATA3 haplo-insufficiency causes human HDR syndrome. Nature 406: 419–422.
33. Bilous RW, Murty G, Parkinson DB, Thakker RV, Coulthard MG, et al. (1992) Brief report: autosomal dominant familial hypoparathyroidism, sensorineural deafness, and renal dysplasia. N Engl J Med 327: 1069–1074.
34. Lakshmanan G, Lieuw KH, Lim KC, Gu Y, Grosveld F, et al. (1999) Localization of distant urogenital system-, central nervous system-, and endocardium-specific transcriptional regulatory elements in the GATA-3 locus. Mol Cell Biol 19: 1558–1568.
35. Khandekar M, Suzuki N, Lewton J, Yamamoto M, Engel JD (2004) Multiple, distant Gata2 enhancers specify temporally and tissue-specific patterning in the developing urogenital system. Mol Cell Biol 24: 10263–10276.
36. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z (2006) GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. Cell 127: 1041–1055.
37. Li S, Li J, Tian J, Dong R, Wei J, et al. (2012) Characterization, tissue expression, and imprinting analysis of the porcine CDKN1C and NAP1L4 genes. J Biomed Biotechnol 2012: 946527.
38. Okada M, Hozumi Y, Ichimura T, Tanaka T, Hasegawa H, et al. (2011) Interaction of nucleosome assembly proteins abolishes nuclear localization of DGKzeta by attenuating its association with importins. Exp Cell Res 317: 2853–2863.
39. Xiao H, Shan L, Zhu H, Xue F (2012) Detection of significant pathways in osteoporosis based on graph clustering. Mol Med Report 6: 1325–1332.
40. Lim AL, Ng S, Leow SC, Choo R, Ito M, et al. (2012) Epigenetic state and expression of imprinted genes in umbilical cord correlates with growth parameters in human pregnancy. J Med Genet.
41. Apostolidou S, Abu-Amero S, O'Donoghue K, Frost J, Olafsdottir O, et al. (2007) Elevated placental expression of the imprinted PHLDA2 gene is associated with low birth weight. J Mol Med (Berl) 85: 379–387.
42. Ishida M, Monk D, Duncan AJ, Abu-Amero S, Chong J, et al. (2012) Maternal inheritance of a promoter variant in the imprinted PHLDA2 gene significantly increases birth weight. Am J Hum Genet 90: 715–719.
43. Lewis RM, Cleal JK, Ntani G, Crozier SR, Mahon PA, et al. (2012) Relationship between placental expression of the imprinted PHLDA2 gene, intrauterine skeletal growth and childhood bone mass. Bone 50: 337–342.
44. Huang J, Wang L, Jiang M, Lin H, Qi L, et al. (2012) PTHLH coupling upstream negative regulation of fatty acid biosynthesis and Wnt receptor signal to downstream peptidase activity-induced apoptosis network in human hepatocellular carcinoma by systems-theoretical analysis. J Recept Signal Transduct Res 32: 250–256.
45. Miraliakbari BA, Asa SL, Boudreau SF (1992) Parathyroid hormone-like peptide in pancreatic endocrine carcinoma and adenocarcinoma associated with hypercalcemia. Hum Pathol 23: 884–887.
46. VanHouten J, Dann P, McGeoch G, Brown EM, Krapcho K, et al. (2004) The calcium-sensing receptor regulates mammary gland parathyroid hormone-related protein production and calcium transport. J Clin Invest 113: 598–608.
47. Li Y, Meng G, Guo QN (2008) Changes in genomic imprinting and gene expression associated with transformation in a model of human osteosarcoma. Exp Mol Pathol 84: 234–239.
48. Arboleda VA, Lee H, Parnaik R, Fleming A, Banerjee A, et al. (2012) Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. Nat Genet 44: 788–792.
49. Zierold C, Reinholz GG, Mings JA, Prahl JM, DeLuca HF (2000) Regulation of the procine 1,25-dihydroxyvitamin D3-24-hydroxylase (CYP24) by 1,25-dihydroxyvitamin D3 and parathyroid hormone in AOK-B50 cells. Arch Biochem Biophys 381: 323–327.
50. Shinki T, Jin CH, Nishimura A, Nagai Y, Ohyama Y, et al. (1992) Parathyroid hormone inhibits 25-hydroxyvitamin D3-24-hydroxylase mRNA expression stimulated by 1 alpha,25-dihydroxyvitamin D3 in rat kidney but not in intestine. J Biol Chem 267: 13757–13762.
51. Shimada T, Hasegawa H, Yamazaki Y, Muto T, Hino R, et al. (2004) FGF-23 is a potent regulator of vitamin D metabolism and phosphate homeostasis. J Bone Miner Res 19: 429–435.
52. Wohrle S, Bonny O, Beluch N, Gaulis S, Stamm C, et al. (2011) FGF receptors control vitamin D and phosphate homeostasis by mediating renal FGF-23 signaling and regulating FGF-23 expression in bone. J Bone Miner Res 26: 2486–2497.
53. Perwad F, Zhang MY, Tenenhouse HS, Portale AA (2007) Fibroblast growth factor 23 impairs phosphorus and vitamin D metabolism in vivo and suppresses 25-hydroxyvitamin D-1alpha-hydroxylase expression in vitro. Am J Physiol Renal Physiol 293: F1577–1583.

OPEN ACCESS Freely available online

PLOS | ONE

Impact of *PSCA* Variation on Gastric Ulcer Susceptibility

Chizu Tanikawa[1], Keitaro Matsuo[2], Michiaki Kubo[5], Atsushi Takahashi[5], Hidemi Ito[2], Hideo Tanaka[2], Yasushi Yatabe[3], Kenji Yamao[4], Naoyuki Kamatani[5], Kazuo Tajima[2], Yusuke Nakamura[1,6], Koichi Matsuda[1]*

1 Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **2** Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan, **3** Department of Pathology and Molecular Diagnostics, Aichi Cancer Center Research Institute, Aichi, Japan, **4** Department of Gastroenterology, Aichi Cancer Center Research Institute, Aichi, Japan, **5** Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan, **6** Departments of Medicine and Surgery, and Center for Personalized Therapeutics, The University of Chicago, Chicago, Illinois, United States of America

Abstract

Peptic ulcer is one of the most common gastrointestinal disorders with complex etiology. Recently we conducted the genome wide association study for duodenal ulcer and identified disease susceptibility variations at two genetic loci corresponding to the *Prostate stem cell antigen* (*PSCA*) gene and the *ABO blood group* (*ABO*) gene. Here we investigated the association of these variations with gastric ulcer in two Japanese case-control sample sets, a total of 4,291 gastric ulcer cases and 22,665 controls. As a result, a C-allele of rs2294008 at *PSCA* increased the risk of gastric ulcer with odds ratio (OR) of 1.13 (*P* value of 5.85×10^{-7}) in an additive model. On the other hand, SNP rs505922 on *ABO* exhibited inconsistent result between two cohorts. Our finding implies presence of the common genetic variant in the pathogenesis of gastric and duodenal ulcers.

Citation: Tanikawa C, Matsuo K, Kubo M, Takahashi A, Ito H, et al. (2013) Impact of *PSCA* Variation on Gastric Ulcer Susceptibility. PLoS ONE 8(5): e63698. doi:10.1371/journal.pone.0063698

Editor: Giuseppe Novelli, Tor Vergata University of Rome, Italy

Received November 12, 2012; **Accepted** April 7, 2013; **Published** May 21, 2013

Funding: The Aichi part of the study was supported by a grant-in-aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan [Priority Areas of Cancer (No. 17015018) and Innovative Areas (No. 221S0001)] and from Ministry of Health, Welfare and Labours of Japan [H24-3-G002 and H23-Application(cancer)-G002]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: koichima@ims.u-tokyo.ac.jp

Introduction

Peptic ulcer is the most common disease in the gastrointestinal tract with symptoms of nausea, vomiting, and abdominal pain, and sometimes causes bleeding and perforation with acute peritonitis. Lifetime prevalence of peptic ulcer is 10–15% in the Japanese and 4–10% in Caucasians [1–3]. Approximately 70% of gastric ulcer patients and 90% of duodenal ulcer patients are associated with *H. pylori* infection [4]. Since eradication of *H. pylori* by antibiotics in combination with proton pomp inhibitor can effectively cure peptic ulcer [5], *H. pylori* is shown to be the major cause of peptic ulcer. Although nearly 50% of individuals on the earth are infected with *H. pylori*, most of them remain asymptomatic indicating that the clinical outcome after the *H. pylori* infection varies substantially between individuals. These inter-individual diversities are affected by various factors including bacteria subtypes, host response, and their interaction. *Duodenal ulcer promoting gene A* (*dupA*) in *H. pylori* was indicated to induce interleukin (IL)-8 that increases the risk of duodenal ulcer and decreases the risk of gastric cancer [6,7]. Nonsteroidal anti-inflammatory drugs (NSAIDs) and smoking are known risk factors for peptic ulcer [8,9]. In addition to these bacterial and environmental factors, host genetic factors had been implicated to have some roles in the risk of peptic ulcer. Proband-wise concordance rate of peptic ulcer in monozygotic twins was as high as 23.6% while that in dizygotic twins was 14.8%. Several candidate gene approaches revealed the possible association of genetic variations in *IL-6*, *IL-8*, *IL-10* [10], *TNF*, *LTA* [11], and *COX1* [12] with peptic ulcer risk.

In our previous genome wide association study (GWAS) of duodenal ulcer using a total of 7,035 cases and 25,323 controls, we identified the significant association of genetic variations at *PSCA* (*prostate stem cell antigen*) and the *ABO* blood group with duodenal ulcer [13]. The C allele of rs2294008 at *PSCA* increased the risk of duodenal ulcer (odds ratio (OR) of 1.84 with *P* value of 3.92×10^{-33}) in a recessive model, while it decreased the risk of gastric cancer (OR of 0.79 with *P* value of 6.79×10^{-12}) as reported previously [14]. Our functional analyses revealed that the T allele of SNP rs2294008 creates an upstream translational initiation codon and add the signal peptide sequences at the N-terminal portion, resulting in alteration of the protein subcellular localization from cytoplasm to cell surface. SNP rs505922 on *ABO* was also associated with duodenal ulcer in a recessive model (OR of 1.32 with *P* value of 1.15×10^{-10}). Since *H. pylori* infection and non-steroidal anti-inflammatory drugs induce gastroduodenal mucosal injury which would cause duodenal and gastric ulcer, we examined the role of variants in the *PSCA* and *ABO* genes on gastric ulcer risk among Japanese population.

Results

A total of 4,291 gastric ulcer cases and 22,665 controls without having the past history of duodenal ulcer or continuous NSAID intake were recruited from the BioBank Japan and the Aichi

Cancer Center (**Table 1**). We then genotyped SNP rs2294008 and rs505922 in two case-control sample sets and examined the association with gastric ulcer in three genetic models (additive, recessive, and dominant model) (**Table 2**). To increase the statistical power of this study, we used subjects with either of 22 diseases as control samples. Therefore we evaluated the confounding effect of disease mix control samples used in this analysis. SNPs rs2294008 and rs505922 did not show significant association between case-mix controls (n = 19,884) and healthy volunteers (n = 2,781) (**Table S1**). In addition, both SNPs did not show the significant deviation from HWE (Hardy-Weinberg equilibrium) in each disease group. Therefore disease mix controls seem not to largely affect the association result in our analysis.

The results of association analyses revealed that gastric ulcer patients had a higher frequency of C allele at rs2294008 than the control group in both sets (39.7% vs 36.9% and 40.1% vs 37.0%, respectively). A meta-analysis of the two studies showed the significant association of rs2294008 in an additive model with no evidence of heterogeneity ($P = 5.85 \times 10^{-7}$ with OR of 1.13), although the association was not statistically significant among Aichi Cancer Center cohort probably due to smaller sample size. Risk alleles (C allele at rs2294008) in the two sample sets were consistent between duodenal ulcer and gastric ulcer, indicating the role of *PSCA* variation as common genetic factors for peptic ulcer. However impact of this variation on gastric ulcer risk was not as strong as those on duodenal ulcer reported previously [13].

On the other hands, SNP rs505922 showed inconsistent results between two cohorts. A T allele of rs505922 increased gastric ulcer risk in all three genetic models in BioBank Japan cohort. However, gastric ulcer patients exhibited lower frequency (53.5%) of a T allele than the healthy controls (55.1%) in the Aichi Cancer Center cohort. Therefore, further association analysis is essential to determine the role of *ABO* variations on gastric ulcer susceptibility.

Since we have genotyping results of 1,862 gastric ulcer cases and 17,482 controls analyzed by Illumina Human Hap610-Quad genechip, we conducted whole genome screening using these sample set. Although 62 SNPs exhibited suggestive associations with *P* values of less than 1×10^{-4}, no SNPs cleared genome wide significant threshold (**Table S2** and **Figure S1**). Thus, our sample set did not have sufficient statistical power to detect gastric ulcer susceptibility loci by GWAS.

We also investigated the association of previously reported genes with gastric ulcer (**Table 3**). We selected 32 SNPs at five gene loci that had been genotyped by Illumina Human Hap610-Quad genechip. As a result, two loci at *LTA* and *PTGS1* indicated suggestive association ($P = 1.64 \times 10^{-3}$ and 0.0376), although these associations were not statistically significant after Bonferroni's correction ($P < 0.00156 = 0.05/32$). Thus further analyses are necessary to elucidate the role of these variations on gastric ulcer.

Discussion

The development of gastric ulcer is determined by the interplay between gastric acid secretion and mucosal resistance, however their underling pathogenesis has not been fully elucidated. Gastric mucus, a gelatinous material secreted by gastric mucous cells, serves as an unstirred layer through which the diffusion of acid and pepsin is reduced. We here found that variation in the *PSCA* gene was significantly associated with gastric ulcer. PSCA was initially identified as a tumor antigen that was highly expressed in prostate, bladder, and pancreatic cancer tissues [15,16]. Since tumor cells treated with anti-PSCA antibody exhibited a growth suppressive effect [17,18], cell surface-PSCA is considered to play an important role in cell proliferation. In contrast, down-regulation of PSCA in gastric and esophageal cancer tissues was also reported [19,20]. Thus the role of PSCA in carcinogenesis is still controversial [21]. These diverse effects of PSCA among various cancer types might be partially explained by the effect of genetic variation. Individuals carrying the T allele at rs2294008 express PSCA proteins with an additional fragment of nine amino acids at the N-terminal portion [13]. On the other hand, individuals carrying the C allele at rs2294008 express a shorter PSCA protein which lacks the signal peptide and is predicted to be localized in the cytoplasm without glycosylation [22]. We also found that the cytosolic shorter PSCA protein was more susceptible to proteasomal degradation than the long PSCA protein at the cell-surface. Since PSCA-derived peptides were reported to be a target of T-cell-based immunotherapy for advanced prostate cancer [23], the shorter PSCA protein would cause the activation of CD4-positive and/or CD8-positive T cells and subsequently promote epithelial mucosal injury [24]. In contrast, the long PSCA protein at the cell surface might facilitate mucosal repair by enhancing epithelial cell proliferation. In addition, T allele of SNP rs2294008 was shown to be associated with higher mRNA and protein expression [25]. Thus the impact of PSCA on gastric ulcer and carcinogenesis could be regulated by the *PSCA* variation.

H. pylori plays an important role in the development of gastritis, peptic ulcers, and gastric cancer, and the eradication of *H. pylori* was shown to reduce the recurrence of gastric ulcer [26] and prevent the onset of gastric cancer [27]. Since vertical transmission during childhood is the major source of infection, family history of *H. pylori* infection or *H. pylori*-related diseases is a risk factor for *H. pylori* infection [28–30]. In addition, recent accumulated evidences

Table 1. Characteristics of study population.

Samples	Source	Platform	Number of samples	Female (%)	Age (mean +/− SD)
Gastric ulcer[a]	BioBank Japan	Illumina HumanHap 610	1,862	32.0 (%)	66.0+/−10.7
		Invader assay	2,004	35.3 (%)	66.5+/−11.7
	Aichi Cancer Center	TaqMan	425	48.7 (%)	55.6+/−12.3
Control[a]	BioBank Japan[b]	Illumina HumanHap 610	17,482	54.1 (%)	62.2+/−13.0
	BioBank Japan[b]	Illumina HumanHap 550	3,309	66.0 (%)	43.8+/−16.2
	Aichi Cancer Center	TaqMan	1,874	38.7 (%)	53.7+/−14.6

[a]Subjects with a history of gastric cancer or duodenal ulcer were excluded from cases and controls.
[b]Control samples consist of patients with colon cancer, breast cancer, diabetes, arteriosclerosis obliterans, atrial fibrillation, brain infarction, drug response, amyotrophic lateral sclerosis, liver cancer, liver cirrhosis, osteoporosis, fibroid, cervical cancer, chronic hepatitis B, ovarian cancer, pulmonary tuberculosis, keloid, drug eruption, hematological cancer, uterus cancer, heat cramp, endometriosis, and 907 healthy volunteers.
doi:10.1371/journal.pone.0063698.t001

Table 2. Association of PSCA and ABO SNPs with gastric ulcer.

SNP		Case			Control			Additive			Recessive			Dominant		
Chr		CC	CT	TT	RAF[a]	CC	CT	TT	RAF[a]	*P* [b]	OR [b] (95% C.I.)	P_{Het} [c]	*P* [b]	OR [b] (95% C.I.)	P_{Het} [c]	*P* [b]	OR [b] (95% C.I.)	P_{Het} [c]
rs2294008 8q24/*PSCA*	BBJ	##	###	###	#####	###	9577	###	#####	2.55×10^{-6}	### (1.07–1.18)		6.12×10^{-4}	1.18 (1.07–1.30)		3.52×10^{-5}	### (1.08–1.25)	
	ACC	70	201	154	#####	235	917	722	#####	9.09×10^{-2}	### (0.98–1.33)	#####	3.10×10^{-2}	1.38 (1.03–1.84)	#####	0.380	### (0.89–1.37)	0.657
	meta[d]									5.85×10^{-7}	### (1.08–1.18)		8.94×10^{-5}	1.20 (1.09–1.31)		2.60×10^{-5}	### (1.08–1.24)	
rs505922 9q34/*ABO*	BBJ	##	###	###	#####	###	#####	###	#####	9.39×10^{-4}	### (1.03–1.14)	6.91×10^{-2}	7.72×10^{-4}	1.13 (1.05–1.22)		5.26×10^{-2}	### (1.00–1.19)	7.90×10^{-2}
	ACC	97	201	127	#####	379	925	570	#####	0.407	### (0.81–1.09)		0.829	0.97 (0.77–1.23)	#####	0.232	### (0.67–1.10)	
	meta[d]									3.88×10^{-3}	### (1.02–1.12)		1.75×10^{-3}	1.12 (1.04–1.20)		0.141	### (0.98–1.15)	

We analyzed 4,291 gastric ulcer cases and 22,665 controls.

[a]rs2294008 (C allele) and rs505922 (T allele).

[b]P values were obtained using chi-square test. To calculate odds ratios (OR), non risk alleles were considered as references.

[c]Heterogeneity across two stages was assessed by Cochran Q test.

[d]OR and P values were obtained using the Mantel-Haenszel fixed-effects model in the meta analysis.

doi:10.1371/journal.pone.0063698.t002

revealed a number of risk factors of gastric cancer (T allele at rs2294008, blood type A, decreased gastric acid, intake of proton pump inhibitor/H_2 blocker, and *CagA* in *H. pylori* [31]) or peptic ulcer (C allele at rs2294008, blood type O, NSAID intake, *dupA* in *H. pylori*) [32]. In addition, *CYP2C19* genotype was associated with the response to triple anti-*H. pylori* therapy including proton pump inhibitor [33]. However, our previous analysis revealed that SNP rs2294008 and rs505922 did not associated with *H. pylori* prevalence [13]. Taking the above information into account, the estimation of disease risk and drug efficacy would enable us to determine the appropriate treatment protocol for *H. pylori* carriers.

Here we found that *PSCA* variant was significantly associated with gastric ulcer. In our previous analysis, PSCA variation did not associate with *H. pylori* prevalence [13]. Since *H. pylori* infection was associated with many diseases such as MALT lymphoma [34], idiopathic thrombocytopenic purpura [35], atrophic gastritis [36], and NSAID-induced gastric ulcer, it is very interesting to evaluate the effect of *PSCA* variation on these diseases. We hope our findings would contribute to the elucidation of disease pathogenesis as well as to the establishment of personalized medical treatments in the future.

Methods

Ethics Statement

This research project was approved by the ethical committees at the University of Tokyo, RIKEN, and Aichi Cancer Center. All participants provided written informed consent as approved by the ethical committees of the University of Tokyo and Aichi Cancer Center.

Study participants

The demographic details of study participants are summarized in Table 1. A total of 3,866 gastric ulcer patients, and 20,791 gastric ulcer negative controls were obtained from BioBank Japan that was initiated in 2003 with the funding from the Ministry of Education, Culture, Sports, Science and Technology, Japan [37]. In the BioBank Japan Project, DNA and serum of patients with 47 diseases were collected through collaborating network of 66 hospitals throughout Japan. The list of participating hospitals is shown in the following website (http://biobankjp.org/plan/member_hospital.html). A total of 425 gastric ulcer cases and 1,874 healthy controls were obtained from the Aichi Cancer Center. The diagnosis of gastric ulcer was based on clinical, endoscopic, and histological features. List of disease-mix control samples used in this study was shown in **Table S1**. We excluded patients with duodenal ulcer or gastric cancer from both cases and controls. Deregulation of PSCA was reported in many types of malignancy such as prostate, pancreatic, lung, bladder, gastric, cholangiocarcinoma, and esophageal cancer [14–16,20,38,39]. In addition, *ABO* locus was previously shown to be associated with various diseases such as myocardial infarction and pancreatic cancer [40,41]. Therefore, we excluded subjects with these diseases from case mix controls. We also excluded the subjects with continuous NSAID intake.

SNP Genotyping

Genotyping platforms used in this study are shown in Table 1. A total of 1,862 gastric ulcer cases and 20,791 gastric ulcer negative control samples were genotyped with Illumina Human Hap610-Quad or with Human Hap550v3. The other samples were genotyped by the Invader assay system (Third Wave Technologies, Madison, WI) or Taqman assay.

Table 3. Association of variations on candidate genes with Gastric ulcer.

SNP	Gene	relative loc	Chr	Position	Gastric ulcer	
					[a] *P*	[b] OR (95% C.I.)
rs3024505	*IL10*	1044	1	2.05E+08	0.817	### (0.71–1.31)
rs3024498	*IL10*	0	1	2.05E+08	0.569	### (0.55–3.01)
rs1554286	*IL10*	0	1	2.05E+08	0.911	### (0.93–1.08)
rs3021094	*IL10*	0	1	2.05E+08	0.178	### (0.98–1.12)
rs3024490	*IL10*	0	1	2.05E+08	0.902	### (0.93–1.07)
rs2222202	*IL10*	0	1	2.05E+08	0.975	### (0.74–1.33)
rs1800896	*IL10*	−1058	1	2.05E+08	0.766	### (0.87–1.20)
rs2844484	*LTA*	−3869	6	31644203	4.41×10^{-2}	### (1.00–1.15)
rs2009658	*LTA*	−1849	6	31646223	0.453	### (0.94–1.14)
rs2844482	*LTA*	−326	6	31647746	0.484	### (0.94–1.13)
rs1800683	*LTA*	−22	6	31648050	1.64×10^{-3}	### (0.83–0.96)
rs2229094	*LTA*	0	6	31648535	0.163	### (0.98–1.16)
rs2229092	*LTA*	0	6	31648736	0.295	### (0.90–1.42)
rs1041981	*LTA*	0	6	31648763	1.75×10^{-3}	### (0.83–0.96)
rs3093662	*TNF*	0	6	31652168	0.220	### (0.91–1.51)
rs3093668	*TNF*	383	6	31654474	0.335	### (0.87–1.51)
rs833068	*VEGFA*	0	6	43850505	0.427	### (0.96–1.10)
rs833069	*VEGFA*	0	6	43850557	0.401	### (0.96–1.10)
rs3025010	*VEGFA*	0	6	43855555	0.976	### (0.93–1.07)
rs3025033	*VEGFA*	0	6	43859053	0.282	### (0.96–1.14)
rs3025035	*VEGFA*	0	6	43859337	0.841	### (0.93–1.09)
rs6900017	*VEGFA*	4261	6	43866463	9.77×10^{-2}	### (0.99–1.15)
rs2069837	*IL6*	0	7	22734552	0.728	### (0.93–1.12)
rs2066992	*IL6*	0	7	22734774	0.953	### (0.93–1.09)
rs1554606	*IL6*	0	7	22735232	0.813	### (0.61–1.86)
rs10242595	*IL6*	2611	7	22740756	0.799	### (0.87–1.11)
rs1236913	*PTGS1*	0	9	1.24E+08	0.389	### (0.74–1.12)
rs1213266	*PTGS1*	0	9	1.24E+08	0.263	### (0.81–1.06)
rs4836885	*PTGS1*	0	9	1.24E+08	0.964	### (0.82–1.23)
rs6478565	*PTGS1*	0	9	1.24E+08	0.318	### (0.78–1.08)
rs10306163	*PTGS1*	0	9	1.24E+08	0.119	### (0.80–1.03)
rs10306202	*PTGS1*	1540	9	1.24E+08	3.76×10^{-2}	### (0.75–0.99)

We analyzed 1,862 gastric ulcer cases and 17,482 controls in this analysis. Chr.,chromosome; Postion in the NCBI Build 36.3.
[a]P values were calculated by Cochran Armitage trend test.
[b]OR, odds ratio was calculated by considering the major allele as the reference.
doi:10.1371/journal.pone.0063698.t003

Statistical Analysis

The association of SNPs rs2294008 and rs505922 with gastric ulcer was tested by chi-square test. The Odds ratios were calculated by considering the protective allele as the reference allele. The association of SNPs genotyped by Illumina Human Hap610-Quad with gastric ulcer was tested by multivariate logistic regression analysis upon adjusting for age at recruitment and gender using PLINK [42]. Heterogeneity across two stages was examined by Cochran Q test [43].

Supporting Information

Figure S1 Manhattan plot showing the genome-wide P values of association. The P values were obtained by logistic regression analysis upon adjustment for age and gender. The y-axis represents the –log10 P values of 480,566 SNPs, and their chromosomal positions are shown on x-axis.
(TIF)

Table S1 Genotype frequency of two SNPs in disease mix controls.
(DOCX)

Table S2 The result of association analysis of Gastric ulcer in GWAS.
(DOCX)

Acknowledgments

We would like to thank all the patients and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan, who

donated their DNA for this work. We also thank Ayako Matsui and Hiroe Tagaya (the University of Tokyo), and the technical staff of the Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN for their technical support.

Author Contributions

Conceived and designed the experiments: CT K. Matsuda YN. Performed the experiments: CT K. Matsuo MK. Analyzed the data: CT AT NK HI. Contributed reagents/materials/analysis tools: HT YY KT KY. Wrote the paper: CT YN K. Matsuda.

References

1. Schlemper RJ, van der Werf SD, Vandenbroucke JP, Biemond I, Lamers CB (1993) Peptic ulcer, non-ulcer dyspepsia and irritable bowel syndrome in The Netherlands and Japan. Scand J Gastroenterol Suppl 200: 33–41.
2. Araki S, Goto Y (1985) Peptic ulcer in male factory workers: a survey of prevalence, incidence, and aetiological factors. J Epidemiol Community Health 39: 82–85.
3. Schlesinger PK, Robinson B, Layden TJ (1992) Epidemiology considerations in peptic ulcer disease. J Assoc Acad Minor Phys 3: 70–77.
4. Marshall BJ (1994) Helicobacter pylori. Am J Gastroenterol 89: S116–128.
5. Hopkins RJ, Girardi LS, Turney EA (1996) Relationship between Helicobacter pylori eradication and reduced duodenal and gastric ulcer recurrence: a review. Gastroenterology 110: 1244–1252.
6. Lu H, Hsu PI, Graham DY, Yamaoka Y (2005) Duodenal ulcer promoting gene of Helicobacter pylori. Gastroenterology 128: 833–848.
7. Argent RH, Burette A, Miendje Deyi VY, Atherton JC (2007) The presence of dupA in Helicobacter pylori is not significantly associated with duodenal ulceration in Belgium, South Africa, China, or North America. Clin Infect Dis 45: 1204–1206.
8. Wolfe MM, Lichtenstein DR, Singh G (1999) Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. N Engl J Med 340: 1888–1899.
9. Chen TS, Lee YC, Li FY, Chang FY (2005) Smoking and hyperpepsinogenemia are associated with increased risk for duodenal ulcer in Helicobacter pylori-infected patients. J Clin Gastroenterol 39: 699–703.
10. Kang JM, Kim N, Lee DH, Park JH, Lee MK, et al. (2009) The effects of genetic polymorphisms of IL-6, IL-8, and IL-10 on Helicobacter pylori-induced gastroduodenal diseases in Korea. J Clin Gastroenterol 43: 420–428.
11. Lanas A, García-González MA, Santolaria S, Crusius JB, Serrano MT, et al. (2001) TNF and LTA gene polymorphisms reveal different risk in gastric and duodenal ulcer patients. Genes Immun 2: 415–421.
12. Arisawa T, Tahara T, Shibata T, Nagasaka M, Nakamura M, et al. (2007) Association between genetic polymorphisms in the cyclooxygenase-1 gene promoter and peptic ulcers in Japan. Int J Mol Med 20: 373–378.
13. Tanikawa C, Urabe Y, Matsuo K, Kubo M, Takahashi A, et al. (2012) A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. Nat Genet 44: 430–434, S431–432.
14. Sakamoto H, Yoshimura K, Saeki N, Katai H, Shimoda T, et al. (2008) Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. Nat Genet 40: 730–740.
15. de Nooij-van Dalen AG, van Dongen GA, Smeets SJ, Nieuwenhuis EJ, Stigter-van Walsum M, et al. (2003) Characterization of the human Ly-6 antigens, the newly annotated member Ly-6K included, as molecular markers for head-and-neck squamous cell carcinoma. Int J Cancer 103: 768–774.
16. Gu Z, Thomas G, Yamashiro J, Shintaku IP, Dorey F, et al. (2000) Prostate stem cell antigen (PSCA) expression increases with high gleason score, advanced stage and bone metastasis in prostate cancer. Oncogene 19: 1288–1296.
17. Marra E, Uva P, Viti V, Simonelli V, Dogliotti E, et al. (2010) Growth delay of human bladder cancer cells by Prostate Stem Cell Antigen downregulation is associated with activation of immune signaling pathways. BMC Cancer 10: 129.
18. Gu Z, Yamashiro J, Kono E, Reiter RE (2005) Anti-prostate stem cell antigen monoclonal antibody 1G8 induces cell death in vitro and inhibits tumor growth in vivo via a Fc-independent mechanism. Cancer Res 65: 9495–9500.
19. Study Group of Millennium Genome Project for C, Sakamoto H, Yoshimura K, Saeki N, Katai H, et al. (2008) Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. Nat Genet 40: 730–740.
20. Bahrenberg G, Brauers A, Joost HG, Jakse G (2000) Reduced expression of PSCA, a member of the LY-6 family of cell surface antigens, in bladder, esophagus, and stomach tumors. Biochem Biophys Res Commun 275: 783–788.
21. Saeki N, Gu J, Yoshida T, Wu X (2010) Prostate stem cell antigen: a Jekyll and Hyde molecule? Clin Cancer Res 16: 3533–3538.
22. Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14: 897–911.
23. Raff AB, Gray A, Kast WM (2009) Prostate stem cell antigen: a prospective therapeutic and diagnostic target. Cancer Lett 277: 126–132.
24. Ohara T, Morishita T, Suzuki H, Masaoka T, Ishii H (2003) Perforin and granzyme B of cytotoxic T lymphocyte mediate apoptosis irrespective of Helicobacter pylori infection: possible act as a trigger of peptic ulcer formation. Hepatogastroenterology 50: 1774–1779.
25. Kohaar I, Porter-Gill P, Lenz P, Fu YP, Mumy A, et al. (2013) Genetic variant as a selection marker for anti-prostate stem cell antigen immunotherapy of bladder cancer. J Natl Cancer Inst 105: 69–73.
26. Marshall BJ, Goodwin CS, Warren JR, Murray R, Blincow ED, et al. (1988) Prospective double-blind trial of duodenal ulcer relapse after eradication of Campylobacter pylori. Lancet 2: 1437–1442.
27. Fukase K, Kato M, Kikuchi S, Inoue K, Uemura N, et al. (2008) Effect of eradication of Helicobacter pylori on incidence of metachronous gastric carcinoma after endoscopic resection of early gastric cancer: an open-label, randomised controlled trial. Lancet 372: 392–397.
28. Leung WK, Ng EK, Lam CC, Chan KF, Chan WY, et al. (2006) Helicobacter pylori infection in 1st degree relatives of Chinese gastric cancer patients. Scand J Gastroenterol 41: 274–279.
29. Brenner H, Bode G, Boeing H (2000) Helicobacter pylori infection among offspring of patients with stomach cancer. Gastroenterology 118: 31–35.
30. Brenner H, Rothenbacher D, Bode G, Adler G (1998) Parental history of gastric or duodenal ulcer and prevalence of Helicobacter pylori infection in preschool children: population based study. BMJ 316: 665.
31. Hatakeyama M (2009) Helicobacter pylori and gastric carcinogenesis. J Gastroenterol 44: 239–248.
32. Yamaoka Y (2010) Mechanisms of disease: Helicobacter pylori virulence factors. Nat Rev Gastroenterol Hepatol. England. pp. 629–641.
33. Furuta T, Shirai N, Takashima M, Xiao F, Hanai H, et al. (2001) Effect of genotypic differences in CYP2C19 on cure rates for Helicobacter pylori infection by triple therapy with a proton pump inhibitor, amoxicillin, and clarithromycin. Clin Pharmacol Ther 69: 158–168.
34. Montalban C, Santon A, Boixeda D, Bellas C (2001) Regression of gastric high grade mucosa associated lymphoid tissue (MALT) lymphoma after Helicobacter pylori eradication. Gut 49: 584–587.
35. Michel M, Cooper N, Jean C, Frissora C, Bussel JB (2004) Does Helicobater pylori initiate or perpetuate immune thrombocytopenic purpura? Blood 103: 890–896.
36. Rokkas T, Pistiolas D, Sechopoulos P, Robotis I, Margantinis G (2007) The long-term impact of Helicobacter pylori eradication on gastric histology: a systematic review and meta-analysis. Helicobacter 12 Suppl 2: 32–38.
37. Nakamura Y (2007) The BioBank Japan Project. Clin Adv Hematol Oncol 5: 696–697.
38. Kawaguchi T, Sho M, Tojo T, Yamato I, Nomi T, et al. (2010) Clinical significance of prostate stem cell antigen expression in non-small cell lung cancer. Jpn J Clin Oncol 40: 319–326.
39. Ono H, Hiraoka N, Lee YS, Woo SM, Lee WJ, et al. (2012) Prostate stem cell antigen, a presumable organ-dependent tumor suppressor gene, is down-regulated in gallbladder carcinogenesis. Genes Chromosomes Cancer 51: 30–41.
40. Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, et al. (2011) Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. Lancet 377: 383–392.
41. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, et al. (2009) Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet 41: 986–990.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
43. Breslow N, Day N (1987) Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. IARC Sci Publ: 1–406.

OPEN ACCESS Freely available online

PLOS | ONE

Genome Wide Association Study of Age at Menarche in the Japanese Population

Chizu Tanikawa[1☯], Yukinori Okada[2,3,4☯], Atsushi Takahashi[2], Katsutoshi Oda[5], Naoyuki Kamatani[2], Michiaki Kubo[2], Yusuke Nakamura[1,6], Koichi Matsuda[1]*

1 Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **2** Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan, **3** Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **4** Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, **5** Department of Obstetrics and Gynecology, Faculty of Medicine, The University of Tokyo, Tokyo, Japan, **6** Departments of Medicine and Surgery, and Center for Personalized Therapeutics, The University of Chicago, Chicago, Illinois, United States for America

Abstract

Age at menarche (AAM) is a complex trait involving both genetic and environmental factors. To identify the genetic factors associated with AAM, we conducted a large-scale meta-analysis of genome-wide association studies using more than 15,000 Japanese female samples. Here, we identified an association between SNP (single nucleotide polymorphism) rs364663 at the *LIN28B* locus and AAM, with a P-value of 5.49×10^{-7} and an effect size of 0.089 (year). We also evaluated 33 SNPs that were previously reported to be associated with AAM in women of European ancestry. Among them, two SNPs rs4452860 and rs7028916 in *TMEM38B* indicated significant association with AAM in the same directions as reported in previous studies ($P=0.0013$ with an effect size of 0.051) even after Bonferroni correction for the 33 SNPs. In addition, six loci in or near *CCDC85A*, *LOC100421670*, *CA10*, *ZNF483*, *ARNTL*, and *RXRG* exhibited suggestive association with AAM ($P<0.05$). Our findings elucidated the impact of genetic variations on AAM in the Japanese population.

Citation: Tanikawa C, Okada Y, Takahashi A, Oda K, Kamatani N, et al. (2013) Genome Wide Association Study of Age at Menarche in the Japanese Population. PLoS ONE 8(5): e63821. doi:10.1371/journal.pone.0063821

Editor: Chunyan He, The University of Tokyo, Japan

Received October 16, 2012; **Accepted** March 26, 2013; **Published** May 7, 2013

Funding: This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government. This work was also supported by grant from the Ministry of Education, Culture, Sports, Science and Technology (C. Tanikawa and K. Matsuda) and Shiseido female researcher science grant (C. Tanikawa). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: koichima@ims.u-tokyo.ac.jp

☯ These authors contributed equally to this work.

Introduction

Age at menarche (AAM), the onset of the first menstrual period in girls, is considered as a landmark of female pubertal development. Menarche generally occurs after a series of complex neuroendocrine events leading to full activation of the hypothalamic-pituitary-gonadal axis [1]. Menarche is associated with physical, emotional, and social development [2]. In addition, AAM was shown to be associated with the risk of various diseases. Early AAM is reported to be one of the significant risk factors for depression [3], eating disorders [4], obesity [5], diabetes [6], breast cancer [7], and coronary heart disease [8]. On the other hand, late AAM has been associated with osteoporosis [9] and taller adult stature [10]. Therefore, the identification of loci contributing to variation in AAM could lead to a better understanding of a wide range of phenotypes.

AAM is known to be a complex trait determined by an array of genetic and environmental variables [11–13]. Twin and family studies suggest a significant genetic contribution to AAM with a heritability of more than 50% [12,14]. Several genetic variations within candidate genes such as the estrogen receptor genes (*ESR1* and *ESR2*) [15,16], *CYP19A1*[17], and the *SHBG* gene [18] were shown to be associated with AAM. To date, a number of genome-wide linkage analyses [14,19,20] and genome-wide association studies (GWAS)[21–25] for genes underlying variation in AAM have been performed. In 2009, the association of genetic variations in *LIN28B* with AAM was identified by four independent groups. Currently, more than 30 loci have been shown to be significantly associated with AAM. However, most of these studies were conducted using women of European ancestry. Here we performed a large scale meta-analysis of GWAS using more than 15,000 Japanese female samples.

Results

A total of 15,495 Japanese female subjects from four GWAS using different SNP genotyping systems were enrolled in this analysis. Characteristics of samples and genotyping methods are summarized in **Table 1**. All the subjects were of Japanese origin and obtained from the Biobank Japan Project [26]. Samples consist of patients that were classified into 33 disease groups. The average and S.D. of AAM in each disease cohort is shown in **Table S1**. Some diseases such as breast cancer and osteoporosis are likely to be associated with early or late AAM, as reported previously [7,9]. We also found that AAM was negatively associated with birth year ($p<0.0001$). Thus we used disease status and birth year as covariates in this study. Genotyping was performed with over 500,000 SNP markers using Illumina

HumanHap 550 Genotyping BeadChip, Illumina610-Quad Genotyping BeadChip, or Illumina Omni Express (Illumina, CA, USA). We applied stringent quality control criteria as mentioned in the methods section. We also conducted principal component analysis [27] to evaluate potential population stratification. To extend the genomic coverage and conduct meta-analysis, we subsequently performed a whole-genome imputation of the SNPs, using HapMap Phase II genotype data [28]. After the imputation, we performed SNP quality control (minor allele frequency≥0.01 and an imputation score (*Rsq* value by MACH software [29])≥0.7) and found that more than two million autosomal SNPs satisfied these criteria.

The associations of these imputed SNPs with AAM were evaluated using a linear regression model [30] and meta-analysis. Quantile-Quantile plots of P-values indicated the Inflation factors to be as low as 1.039 (**Figure 1**), suggesting no substantial population stratification existed in our study population. Although we could not identify significant association that satisfied the genome-wide significant threshold ($P<5\times10-8$), SNP rs364663 which is located within intron 2 of the *LIN28B* gene at 6q21 indicated the strongest association with a P-value of 5.49×10^{-7} (**Table 2**, **Figure 2**). SNP rs364663 exhibited the association with AAM in all four cohorts without significant heterogeneity in both effect sizes and directions (P-value for heterogeneity = 0.29; **Table 3**), suggesting that the observed associations at *LIN28B* are not the result of false-positives due to study-specific bias. Regional p-value plots indicated that all of the AMM-associated SNPs were clustered around the *LIN28B* locus (**Figure 2**). Previously reported SNPs rs314263 and rs7759938 near the *LIN28B* locus [21,24] also associated with AAM ($P=1.03\times10-6$ and $1.12\times10-6$, respectively) in the same direction. In addition to the 6q21 locus, 17 SNPs in seven genomic regions exhibited suggestive association with a p-value of $<1\times10-5$ (**Table 2**). To further investigate the physiological role of these loci, the associations of these variations with body mass index (BMI) and height were evaluated using previously published results in 26,620 Japanese subjects [31]. SNP rs9404590 near the *LIN28B* locus associated with height with p-value of 0.0003 (Table S2), but we did not find significant association ($P<0.01$) between these loci and BMI (Table S2).

Since AAM is associated with various disease risks, we conducted separate analyses for each disease and then performed meta-analysis for the top 42 loci. As a result, some SNPs showed slightly stronger association, but none cleared the genome wide significant threshold (Table S3). Similar to the current study, our group had previously conducted QTL analyses using disease status as a covariate and successfully identified many QTL loci [31–35]. Therefore, different background due to disease status was unlikely to significantly affect the result of association analysis.

Additionally, we examined the loci already known to show significant association with AAM in women of European ancestry [22–25,36]. We selected 37 SNPs for this candidate analysis and successfully obtained the genotyping results of 33 SNPs (**Table 4**), and the risk allele was consistent with previous reports for 31 SNPs. In addition, eight SNPs in or near *RXRG*, *CCDC85A*, *LOC100421670*, *TMEM38B*, *ZNF483*, *ARNTL*, and *CA10* indicated possible associations with AAM ($P<0.05$). Among them, rs4452860 and rs7028916 at the *TMEM38B* locus exhibited significant association even after Bonferroni's correction ($P<0.0015 = 0.05/33$). Taken together, these variations as well as *LIN28B* are likely to be common AAM loci, although their effect sizes are different between women of European ancestry and those of Japanese.

Table 1. Characteristics of study population.

Cohorts	Number of Samples	Source	Platform	Inflation factor	SNP number	Age (S.D.)	Diseases
Cohort1	11,454	BioBank Japa n	Illumina HumanHap 610	1.052	2,263,308	59.81+−13.25	Cancer(Colorectal, breast, lung, gastric, pancreas, liver, cholangiocarcinoma), diabetes mellitus, myocardial infarction, brain infarction, arteriosclerosis obliterans, Arrythimia,drug eruption, liver cirrhosis, amyotrophic lateral sclerosis, osteoporosis, fibroid, Rheumatoid arthritis, and drug response
Cohort2	941	BioBank Japan	Illumina HumanHap 550	1.001	2,220,799	47.01+−15.04	Cancer (cervival, uterus, esophageal, hematopoietic, cholangiocarcinoma, ovarian, pancreas, liver), chronic hepatitis B, pulmonary tuberculosis, keloid, drug eruption, heat cramp
Cohort3	1957	BioBank Japan	Illumina OmniExpress	1.045	2,283,889	60.90+−9.47	Cancer (esophageal, uterus) brain aneurysm, chronic obstractive lung disease, glaucoma
Cohort4	1,143	BioBank Japan	Illumina HumanHap 550	1.008	2,220,799	37.86+−8.13	Endometriosis
Metaanalysis	15,495			1.039	2,310,762	57.56+−14.15	

doi:10.1371/journal.pone.0063821.t001

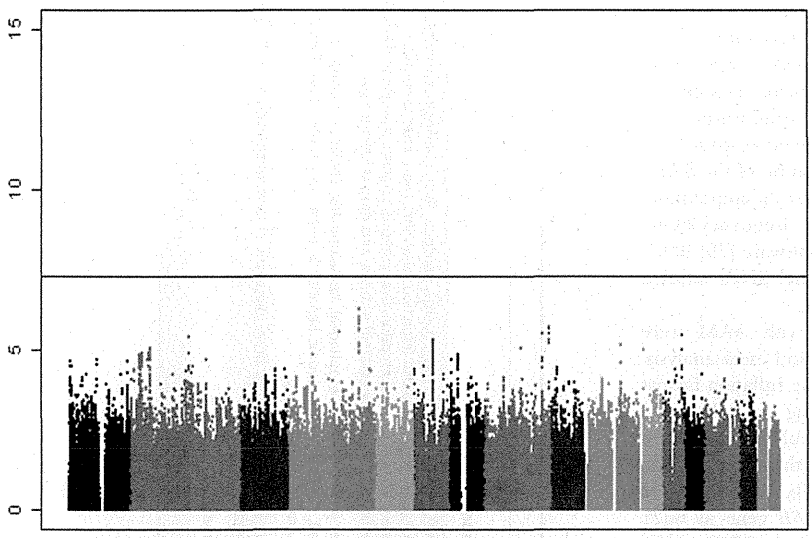

Figure 1. Results from meta-analysis of four genome-wide association studies. A total of 15,495 female samples were analyzed in this study.
doi:10.1371/journal.pone.0063821.g001

Discussion

AAM is a complex trait that is influenced by both genetic and environmental factors. In recent years, genome wide association analyses have become a standard method to identify genetic factors related with various diseases and phenotypes. In 2002 our group performed the first GWAS for myocardial infarction and successfully identified *LTA* as a disease susceptibility gene. Using this method, we have identified a number of loci associated with various phenotypes and common diseases [37–42].

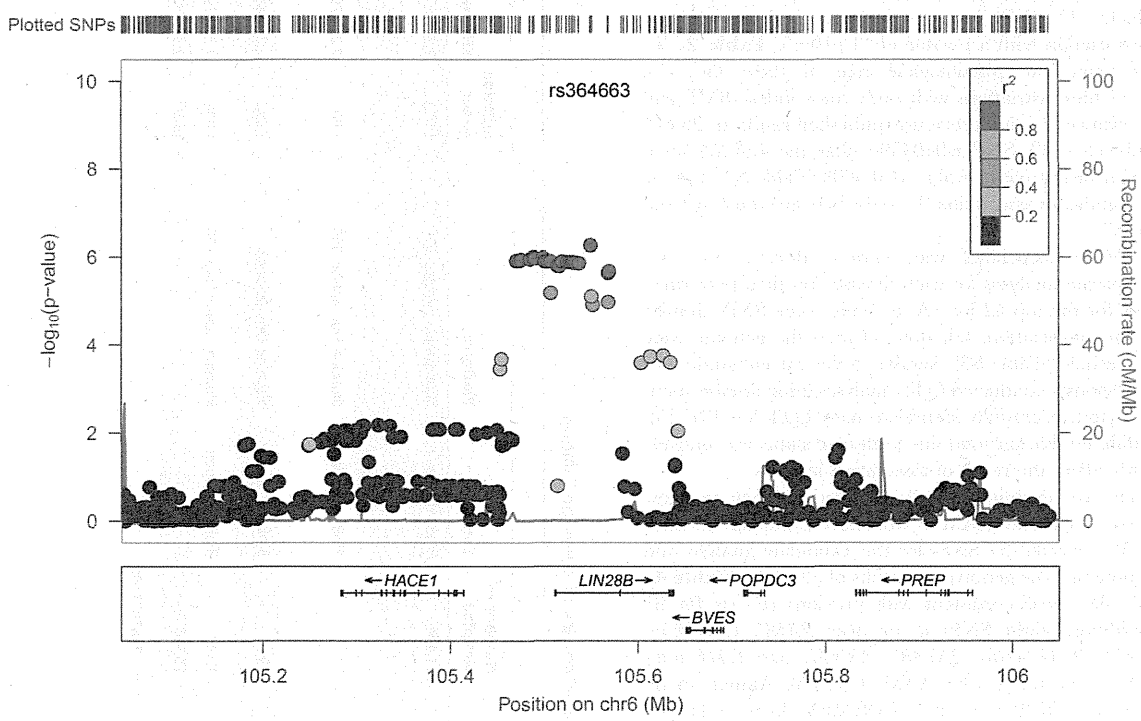

Figure 2. Regional association plot at rs364663. Upper panel; *P* values of genotyped SNPs are plotted (as $-\log_{10}$ values) against their physical location on chromosome 6 (NCBI Build 36). Estimated recombination rates from HapMap JPT shows the local LD structure. Inset; Colors of other SNPs indicate LD with rs2596542 according to a scale from $r^2=0$ to $r^2=1$ based on pair-wise r^2 values from HapMap JPT. Lower panel; Gene annotations from the University of California Santa Cruz genome browser.
doi:10.1371/journal.pone.0063821.g002