

# Genome-wide association study of circulating retinol levels

Alison M. Mondul<sup>1</sup>, Kai Yu<sup>2</sup>, William Wheeler<sup>5</sup>, Hong Zhang<sup>6</sup>, Stephanie J. Weinstein<sup>1</sup>, Jacqueline M. Major<sup>1</sup>, Marilyn C. Cornelis<sup>7</sup>, Satu Männistö<sup>10</sup>, Aditi Hazra<sup>8</sup>, Ann W. Hsing<sup>3</sup>, Kevin B. Jacobs<sup>4,11</sup>, Heather Eliassen<sup>12</sup>, Toshiko Tanaka<sup>13</sup>, Douglas J. Reding<sup>14</sup>, Sara Hendrickson<sup>7,9</sup>, Luigi Ferrucci<sup>13</sup>, Jarmo Virtamo<sup>10</sup>, David J. Hunter<sup>8</sup>, Stephen J. Chanock<sup>4</sup>, Peter Kraft<sup>8</sup> and Demetrius Albanes<sup>1,\*</sup>

<sup>1</sup>Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, <sup>2</sup>Biostatistics Branch, Division of Cancer Epidemiology and Genetics, <sup>3</sup>Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics and <sup>4</sup>Core Genotyping Facility, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA, <sup>5</sup>Information Management Services, Inc., Silver Spring, MD, USA, <sup>6</sup>Institute of Biostatistics, School of Life Science, Fudan University, Shanghai, China, <sup>7</sup>Department of Nutrition, <sup>8</sup>Program in Molecular and Genetic Epidemiology and <sup>9</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA, <sup>10</sup>Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, <sup>11</sup>BioInformed LLC, Gaithersburg, MD, USA, <sup>12</sup>Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, <sup>13</sup>Clinical Research Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA and <sup>14</sup>Marshfield Clinic Research Foundation, Marshfield, WI, USA

Received April 27, 2011; Revised August 18, 2011; Accepted August 25, 2011

Retinol is one of the most biologically active forms of vitamin A and is hypothesized to influence a wide range of human diseases including asthma, cardiovascular disease, infectious diseases and cancer. We conducted a genome-wide association study of 5006 Caucasian individuals drawn from two cohorts of men: the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study and the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. We identified two independent single-nucleotide polymorphisms associated with circulating retinol levels, which are located near the transthyretin (*TTR*) and retinol binding protein 4 (*RBP4*) genes which encode major carrier proteins of retinol: rs1667255 ( $P = 2.30 \times 10^{-17}$ ) and rs10882272 ( $P = 6.04 \times 10^{-12}$ ). We replicated the association with rs10882272 in *RBP4* in independent samples from the Nurses' Health Study and the Invecchiare in Chianti Study (InCHIANTI) that included 3792 women and 504 men ( $P = 9.49 \times 10^{-5}$ ), but found no association for retinol with rs1667255 in *TTR* among women, thus suggesting evidence for gender dimorphism ( $P$ -interaction =  $1.31 \times 10^{-5}$ ). Discovery of common genetic variants associated with serum retinol levels may provide further insight into the contribution of retinol and other vitamin A compounds to the development of cancer and other complex diseases.

## INTRODUCTION

Retinol, one of the most biologically active forms of vitamin A, has been hypothesized to influence a wide range of

human diseases, including cardiovascular disease, infectious diseases and cancer at multiple sites (1). For example, in addition to the classic etiologic role of vitamin A deficiency in xerophthalmia, a pathologic dryness of the eyes which can

\*To whom correspondence should be addressed at: Division of Cancer Epidemiology and Genetics, National Cancer Institute, EPS-320, Bethesda, MD 20892, USA. Tel: +1 3014965626; Fax: +1 3014966829; Email: daa@nih.gov

lead to blindness (2), higher vitamin A status has been related to an increased risk of cardiovascular disease (3) and, most recently, to increased prostate cancer risk (4). Circulating concentrations of retinol are influenced by many factors in addition to diet and supplements, including gut absorption, cleavage of pre-retinol carotenoid compounds, transport and hepatic storage/release. Identification of common variants that influence retinol levels may help shed light on the etiology of several diseases, including cancer.

There is evidence that genetic variants influence circulating retinol. Family studies have estimated that 30% of the variation in serum retinol is heritable (5). One case study demonstrated that a mutation in the gene encoding retinol-binding protein (RBP), one of the two major transport proteins for retinol in circulation, resulted in abnormally low retinol concentrations (6), while inactivation of the transthyretin (*TTR*) gene, the other major retinol transport protein, also resulted in hypovitaminosis A in mice (7). One genome-wide association study (GWAS) study recently examined the genetics of serum retinol concentration in a population without conditions associated with retinol deficiency or *TTR* abnormalities, but failed to identify variants related to serum retinol at the genome-wide significance level (8). Thus, it remains unclear whether common single-nucleotide polymorphisms (SNPs) can explain variations in retinol concentration within the normal range. We therefore conducted a GWAS, similar to recent studies of other micronutrient traits (9,10), that pooled data from two cohort studies, and replicated our findings in two other independent cohorts.

## RESULTS

The individual SNP-serum retinol association *P*-values from the initial GWAS are plotted by chromosome in Supplementary Material, Figure S1. There were 10 SNPs on chromosomes 10 and 18 that reached genome-wide significance ( $P < 5 \times 10^{-8}$ ) for association with circulating retinol concentration. We examined the association between all SNPs with initial-scan *P*-values below  $10^{-5}$  ( $n = 121$ ) in the replication set; no additional SNPs reached genome-wide significance after meta-analysis with these data. The two highly significant SNPs from chromosome 10 were in the gene neighborhood of the *RBP4* gene, which encodes retinol-binding protein 4 (RBP4), one of the two major carriers of retinol in serum. The strongest signal in this region was for rs10882272 ( $P = 6.04 \times 10^{-12}$ ; Table 1 and Fig. 1). In the pooled analysis, the significance of an additional SNP that reached genome-wide significance in this region (rs11187545) was greatly attenuated when it was included in the conditional regression model with rs10882272, showing no evidence for independence and suggesting that the signals from the two SNPs originate from a common locus (even though the two SNPs are not highly correlated;  $r^2 = 0.15$ ) (Fig. 1). Also shown in the figure are the recombination hotspots which support the signal most likely being from one allele. The underlying linkage disequilibrium, as demonstrated with both  $D' = 0.83$  and the haplotype structure, indicate that the variants are well correlated and differ slightly in minor allele frequency.

rs10882272 was independently significant in two of the replication sets (NHS-CGEMS,  $P = 0.003$ ; and InCH-males,  $P = 0.025$ ), as well as in the replication sets combined ( $P = 9.49 \times 10^{-5}$ ), and it remained highly significant in the meta-analysis of the GWAS and replication data ( $P = 6.51 \times 10^{-15}$ ) (Table 1). There was no heterogeneity across studies for this SNP ( $P = 0.67$ , Table 1). The estimated relative difference in mean retinol levels per copy of the rs10882272-C allele from the overall meta-analysis was a decrease of 3.0% (Table 1), and we estimated that this SNP accounted for 0.5% of the variation in serum retinol levels.

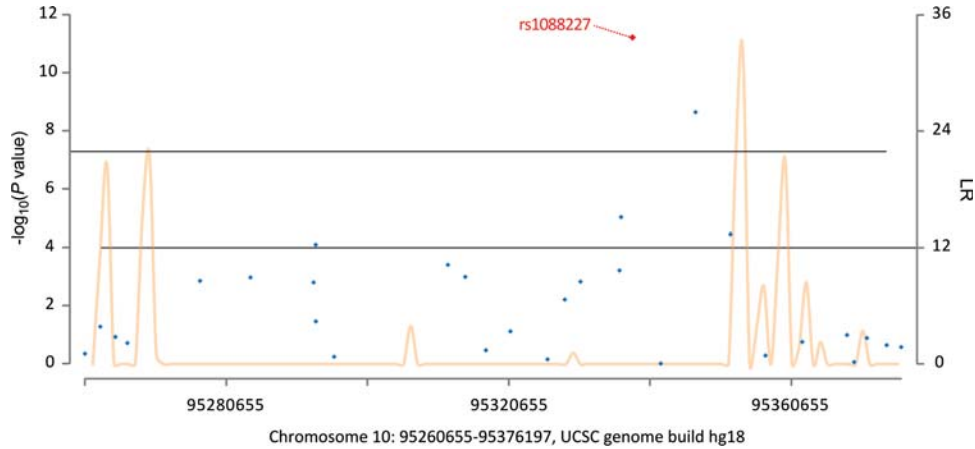
We observed a cluster of eight SNPs on chromosome 18 that were significantly associated with serum retinol in the initial GWAS ( $P < 5 \times 10^{-8}$ ). These SNPs are near *TTR*, the gene encoding TTR, which dimerizes with RBP4 and is therefore also involved in retinol (as well as thyroid hormone) transport in circulation. The strongest signal in this region was for rs1667255 (Table 1 and Fig. 2). When the other seven SNPs that reached genome-wide significance in the pooled analysis (i.e. rs1667254, rs1616887, rs1667234, rs723744, rs4799585, rs9304102, rs1621308) were included in the regression model with rs1667255, their significance was greatly reduced, indicating signal from a common locus. rs1667255 did not reach statistical significance in the replication data set ( $P = 0.08$ , Table 1), but did exhibit highly significant heterogeneity in the strength of association across studies ( $P = 0.0005$ , Table 1), with similar magnitude of association in the ATBC and PLCO male cohorts, somewhat attenuated beta for two of the replication studies (NHS-CHD and InCH-males), and lack of association in the other replication studies of women. The combined meta-analysis yielded significance at the  $P = 6.35 \times 10^{-14}$  level (Table 1), and the random effects meta-analysis *P*-value was 0.085. Combining ATBC and PLCO cohorts, the estimated beta and standard deviation are 0.039 and 0.0046, while in the NHS and InCH-female set the estimated beta and standard deviation are 0.0065 and 0.0059. These yield a formal *z*-test *P*-value of  $1.31 \times 10^{-5}$  for the difference in the strength of SNP association between the male and female samples. Comparing those with two copies of the minor allele to those with zero copy, the difference in mean retinol ranged from  $-2.3$  to 8.7% across the GWAS and replication cohorts (Table 1), and we estimated in the pooled PLCO and ATBC studies that this SNP accounted for 1.7% of the variation in circulating retinol levels. These results were unchanged when we restricted the analysis to participants who did not use supplemental vitamin A of any kind (data not shown).

We conducted an analysis in the pooled discovery GWAS data combining the two identified SNPs (rs10882272 on chromosome 18, and rs1667255 on chromosome 10) and found that individuals with two copies of both variant alleles (i.e. rs1667255: C/C, and rs10882272: C/C) had 12.7–15.1% higher serum retinol than those who were homozygous for the common allele for both SNPs (ATBC: 15.1%, 95% CI: 14.3–18.4%; PLCO: 12.7%, 95% CI: 11.6–13.7%). We estimated that the two SNPs together accounted for 2.3% of the variance in serum retinol levels. A sensitivity analysis conducted among the 2184 controls from the ATBC and PLCO nested sets revealed identical regression parameter estimates

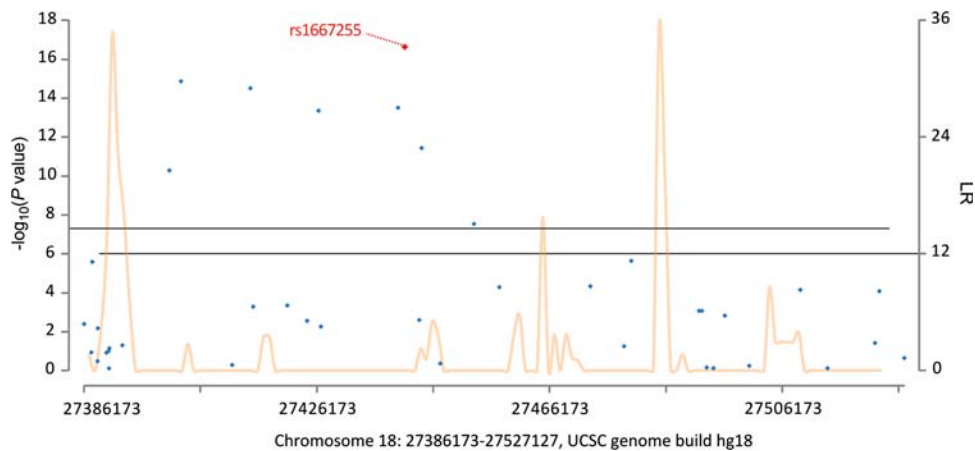
**Table 1.** SNPs associated with serum retinol at  $P < 5 \times 10^{-7}$ -GWAS and replication studies

SNP <sup>a</sup>	Minor allele (freq)	Chr	Location <sup>b</sup>	Gene neighborhood	Population	<i>P</i> -value <sup>c</sup>	Mean serum retinol levels (µg/l) by genotype (copies of minor allele)			Beta <sup>d</sup>	SE <sup>d</sup>	<i>P</i> for heterogeneity <sup>e</sup>
							0	1	2			
rs10882272	C (0.35)	10	95338172	<i>RBP4</i>	GWAS studies							
					Pooled	$6.04 \times 10^{-12}$	–	–	–	–0.03	0.005	
					ATBC	$3.16 \times 10^{-10}$	579.7	567.4	540.7	–0.03	0.005	
					PLCO	0.004	696.9	668.3	653.2	–0.03	0.01	
					Replication studies							
					Total	$9.49 \times 10^{-5}$	–	–	–	–0.02	0.006	
					NHS-T2D	0.23	591.0	565.7	563.9	–0.03	0.02	
					NHS-CHD	0.82	589.3	582.7	604.9	–0.005	0.02	
					NHS-CGEMS	0.003	597.4	581.2	574.1	–0.02	0.008	
					InCH-males	0.025	724.5	710.5	686.4	–0.033	0.014	
					InCH-females	0.24	657.8	649.6	640.7	–0.016	0.014	
					Combined meta-analysis	$6.51 \times 10^{-15}$	–	–	–	–0.03	0.004	0.67
					rs1667255	C (0.31)	18	27441277	<i>TTR</i>	GWAS studies		
Pooled	$2.30 \times 10^{-17}$	–	–	–						0.04	0.005	
ATBC	$6.66 \times 10^{-16}$	555.5	577.8	602.3						0.04	0.005	
PLCO	0.0002	654.2	688.2	711.1						0.04	0.01	
Replication studies												
Total	0.08	–	–	–						0.009	0.005	
NHS-T2D	0.76	566.8	581.1	578.1						0.005	0.02	
NHS-CHD	0.17	577.4	594.4	591.9						0.02	0.02	
NHS-CGEMS	0.37	582.9	587.0	595.5						0.007	0.008	
InCH-males	0.06	690.2	710.3	743.8						0.03	0.015	
InCH-females	0.72	649.9	655.6	635.3						–0.005	0.014	
Combined meta-analysis	$6.35 \times 10^{-14}$	–	–	–						0.03	0.004	0.0005

<sup>a</sup>NCBI dbSNP identifier.<sup>b</sup>NCBI Human Genome Build 36 location.<sup>c</sup>Result of the 1 df test based on linear regression using an additive model and adjusting for case status, age at blood collection (continuous), BMI (continuous), serum cholesterol (continuous), study and eigenvectors chosen to adjust for population stratification. Meta-analysis *P*-values were based on the combined Wald statistics weighted by the square-root of the associated sample size.<sup>d</sup>SE, standard error. The regression  $\beta$  and its standard error are calculated using natural-log transformed retinol. The meta-analysis  $\beta$  and standard error are calculated using the fixed-effect model.<sup>e</sup>Assessed using the Cochrane's Q statistic.



**Figure 1.** LD structure of chromosome 10.  $P$ -values generated from ATBC and PLCO data. LR, the recombination rate on a logarithmic scale with 12 being 'notable' for a hotspot.



**Figure 2.** LD structure of chromosome 18.  $P$ -values generated from ATBC and PLCO data. LR, the recombination rate on a logarithmic scale with 12 being 'notable' for a hotspot.

to those obtained using all participants (per minor allele change in log retinol levels =  $-0.03$  and  $0.04$  for rs10882272 and rs1667255, respectively).

The results above were unchanged when participants using supplemental vitamin A from either individual supplements or multivitamins (i.e. 10% in ATBC, 37% in PLCO and 41% in NHS) were excluded from the analysis (data not shown). We also tested the gene-serum retinol associations within strata of vitamin A supplement use: the betas, SEs and  $P$ -values for *RBP4* rs10882272 among vitamin A supplement users and non-users in our GWAS sample were  $-0.030/0.012/P = 0.008$  and  $-0.033/0.005/P = 1.41 \times 10^{-10}$ , respectively. Similarly for *TTR* rs1667255, they were  $0.040/0.012/P = 0.001$  and  $0.041/0.005/P = 3.38 \times 10^{-15}$ , respectively. The gene-serum retinol associations were also identical for *RBP4* for higher and lower baseline serum retinol strata ( $-0.010/0.005/P = 0.025$  and  $-0.013/0.004/P = 0.002$ , respectively), but showed signal only in the above-the-median group for *TTR* ( $0.019/0.004/P = 1.31 \times 10^{-5}$ ) and not those with lower serum retinol levels ( $0.006/0.005/P = 0.24$ ).

## DISCUSSION

In this GWAS, we identified two distinct regions that influence circulating retinol levels. The SNPs showing the strongest signal localize to regions that include the biologically plausible candidate genes, *RBP4* and *TTR*, which encode the two major carrier proteins involved in retinol transport in circulation. Previous studies found mutations in the *RBP4* and *TTR* genes that resulted in abnormally low retinol levels (6,7); however, no variants have been associated with altered retinol within the normal range, including from a previous GWAS that found no variants associated at the genome-wide significance level (8). Although our analyses suggest that the multiple SNPs we found in association with serum retinol in each gene neighborhood were from common loci, there may be additional functional loci in each of these genes that remain to be found through further study, including fine mapping.

Our results differed from those of the one previous GWAS of serum retinol concentration, which identified no variants associated with serum retinol at the genome-wide significance level

**Table 2.** Descriptive characteristics of the participating GWAS cohorts

	Cohort ATBC	PLCO	NHS	InCH
No. of cases/no. of controls	2336/1678	486/506	1413/1359	0/1124
Location	Finland	Multi-center, USA	USA	Italy
Median age (years)	58 (54–62)	65 (61–68)	60 (55–64)	71 (66–77)
Sex (% male)	100	100	0	45
Smoking (% current)	100	9	15	19
Dietary vitamin A intake (IU/day)	4203 (2711–6602)	11 823 (8321–16 041)	10 682 (7653–15 467)	997 (593–2157)
Supplemental vitamin A use (% yes)	10	37	41	Not queried
Years of blood collection	1985–1988	1993–2001	1989–1990	1998–2000
Blood sample	Serum	Serum	Plasma	Plasma
Median (interquartile range) of serum retinol ( $\mu\text{g/l}$ )	572 (496–654)	672 (562–794)	563 (485–657)	538 (461–634)

(8). One possible explanation for this difference is that the previous study was likely underpowered to detect associations of the magnitude that we observe here; i.e. with an initial GWAS sample size of  $n = 1242$ , they had  $\sim 13\%$  power to detect a SNP explaining 1.5% of the variation in serum retinol levels.

It should also be noted that for the SNP in the *TTR* gene neighborhood associated with serum retinol (rs1667255), there was highly significant heterogeneity between the discovery and replication cohorts. It seems unlikely that systematic differences across the studies, for example in genotyping or laboratory analyses, could account for this difference. It is notable that for the other SNP associated with circulating retinol, rs10882272, there was no evidence of heterogeneity and the findings were consistent across the studies. Another possible explanation has to do with the discovery cohorts consisting only of men, whereas women made up 88% of the replication samples, with the one male replication set showing a borderline non-significant association ( $P = 0.06$ ). The gender difference between the discovery and replication sets may also explain why rs1667255 failed to reach statistical significance in the NHS replication samples, despite having adequate power to detect an association (i.e. 99% power to detect a SNP explaining 1.7% of the variation in serum retinol at  $P = 0.05$ ). This heterogeneity raises the possibility of a biologically based sex difference in the transport of retinol (and possibly thyroid hormones) similar to gender dimorphisms observed for some other traits, including childhood obesity (11) and adult waist-to-hip ratio (12). The suggested gender dimorphism in the association between SNPs in the *TTR* gene neighborhood and serum retinol concentration should be examined in additional studies.

Our investigation has many strengths, including large sample size, an independent population for replication, and measurement of serum retinol using the accepted, gold standard assay method across studies. The present analysis included only Caucasians, however, and re-examination in populations of Asian and African descent is warranted. Because retinol deficiency is not common in developed populations, too few individuals had evidence of retinol deficiency (i.e.  $n = 25$  at  $<300 \mu\text{g/l}$ ); therefore, this specific outcome could not be evaluated.

## Conclusions

In this GWAS, variants near the genes encoding two major proteins involved in the transport of retinol and other

vitamin A compounds in circulation were associated with retinol concentration, and for one (*TTR*), gender dimorphism was suggested. Given the known associations between circulating retinol and human disease, understanding the underlying mechanisms that determine retinol concentrations and status may provide insight into how retinol transport and metabolism are related to disease risk, including cancer. Future studies should examine these findings in other populations.

## MATERIALS AND METHODS

### Sample description: discovery samples

We conducted a GWAS analysis based on two cohorts with prospectively collected serum retinol levels: (i) the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study, a randomized trial of  $\alpha$ -tocopherol and  $\beta$ -carotene for cancer prevention that was conducted among male smokers in southwestern Finland, (13) and (ii) the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial (14), a multi-center cancer screening effectiveness trial conducted in the USA which included both smokers and non-smokers. Details of the participants from each cohort are presented in Table 2. In the ATBC Study, participants were previously selected for nested case–control sets to conduct GWAS analyses to identify genetic determinants of lung, pancreatic, bladder and advanced prostate cancers. The present analyses are conducted in these participants ( $n = 4014$ ) in whom genome-wide scans and information on serum retinol exists. In PLCO, participants were men who were previously selected for a nested case–control set to conduct GWAS analyses to identify genetic determinants of prostate cancer; 992 participants with existing genetic and serum retinol data were available for the present study.

### Sample description: replication samples

We attempted to replicate in two independent cohorts the most significant findings (all SNPs  $P < 10^{-5}$ ,  $n = 121$ ) from the initial GWAS: 2772 women from the Nurses' Health Study (NHS) who were previously genotyped as part of three nested case–control GWAS studies of coronary heart disease (CHD), type 2 diabetes (T2D) and breast cancer [Cancer Genetic Markers of Susceptibility (CGEMS)] (15,16); and, 1124 men and women from the InCHIANTI Study (InCH) (8).



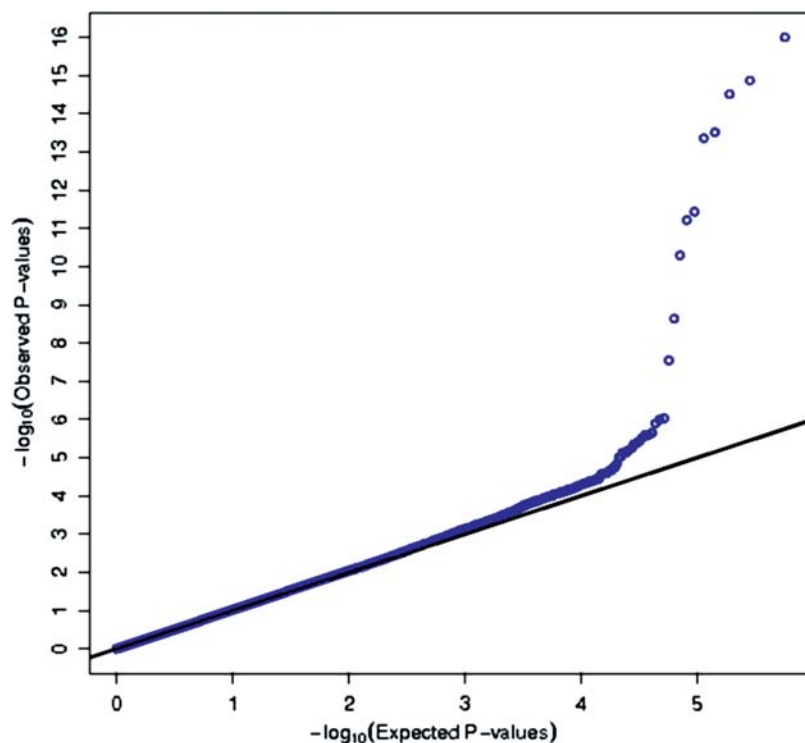


Figure 3. Quantile-quantile plot of 562 105 SNPs from the pooled analysis of the two GWAS scans ( $n = 5\,006$ ;  $\lambda_{GC} = 1.018$ ).

### Genotyping and serum retinol measurement

All genotyping for the initial GWAS was conducted at the National Cancer Institute Core Genotyping Facility (CGF). The discovery samples (ATBC Study and PLCO) were genotyped using the Illumina HumanHap550/610 arrays. The replication samples from NHS were genotyped on either the Illumina HumanHap550 (breast cancer) or Affymetrix 6.0 (CHD and T2D) platforms; this genotyping was performed at the CGF (breast cancer), Merck laboratories (CHD) and the Broad Institute (T2D). Replication study samples from InCH were genotyped on the Illumina HumanHap550 at the Laboratory of Neurogenetics of the National Institute on Aging. Imputation was performed for SNPs not available from genotyping using the hidden-Markov model algorithm implemented in MACH using the HapMap CEU reference panel build 36, R22. The majority of the imputed SNPs had a high imputation quality score ( $\text{MACH-R}^2 > 0.6$ ). Details of the quality-control assessment of genotypes including the sample completion rates, SNP call rates, concordance rates, fitness for deviation from Hardy-Weinberg proportions and final sample selection for analyses are described elsewhere (15-19).

In the ATBC Study, retinol concentration was determined in baseline fasting serum samples using reversed-phase liquid chromatography with diode-array UV detection (20). After ethanol/ether extraction and injection into a Hypersil ODS column with isocratic methanol mobile phase and flow rate of 0.9 ml/min for 9 min run, retinol was monitored at 305 nm wavelength. All samples were protected from light and stored at  $-70^\circ\text{C}$  until they were assayed. The coefficient of variation (CV) for the retinol measurement was 2.2%.

PLCO serum samples collected at study entry were processed and stored within 2 h of collection and were also stored at  $-70^\circ\text{C}$  and measured using reversed-phase liquid chromatography with diode-array UV detection. After ethanol/hexane extraction, injection into an Adsorbosphere HS C column with a step gradient to acetonitrile-methanol-dichloromethane mobile phase, and flow rate of 0.9 ml/min for 23 min, retinol was monitored at 325 nm wavelength (21). Although the PLCO participants were not required to fast prior to blood collection, fasting and non-fasting serum retinol values are not significantly different (22). The CV for serum retinol in PLCO was 5.1%. NHS non-fasting plasma samples were stored at  $-130^\circ\text{C}$  and retinol was determined by reversed-phase high-performance liquid chromatography with UV detection as described by El-Soheily *et al.* (23) with modifications. Samples were processed by organic phase extraction in hexane and retinol was isolated from the extract using a C18 column (150 mm  $\times$  4.6 mm, 3  $\mu\text{m}$  particle size) with a mobile phase mixture of acetonitrile, tetrahydrofuran, methanol and 1% ammonium acetate solution (68:22:7:3) flowing at 1 ml/min. Concentration was determined at 300 nm wavelength (23). Retinol was run in 14 batches; CVs were  $<11\%$  for all except 3 batches which had CVs ranging from 16 to 24% ( $N = 663$  samples). Plasma and serum retinol concentrations have been shown to be quite comparable (24). Blood samples in InCH were collected in the morning after a 12 h overnight fast and aliquots of plasma were stored at  $-80^\circ\text{C}$ . Retinol was measured using an isocratic high-performance liquid chromatography method as described by Sowell *et al.* (25). Retinol was isolated using a 150  $\times$  4.6 mm octadecylsilane column packed with 5  $\mu\text{m}$  particles with a mobile phase of an equivolume solution of ethanol and acetonitrile

containing 0.1 ml of diethylamine per liter of solvent. The flow rate was 0.9 ml/min and the concentration was determined at 325 nm wavelength. The within-run and between-run coefficients of variation were 3.3 and 2.8%.

### Statistical analysis

For the initial GWAS in the ATBC and PLCO, we performed linear regression assuming an additive genetic model of the genotype data and adjusting for cancer case status, age at blood collection (continuous), body mass index (BMI) (continuous), serum cholesterol (continuous), study and eigenvectors chosen to adjust for population stratification. When combining SNPs from two genes for analysis, we created a score indicating the total number of alleles across SNPs conferring higher retinol and performed linear regression assuming an additive genetic model over the number of alleles adjusting for the same factors listed above. Retinol concentrations were transformed by taking the natural log, which was identified as the optimal transformation using the Box-Cox procedure. The Wald test was used to test the association between each SNP ( $n = 562\ 105$ ) and serum retinol. All analyses were conducted using the R software (version 2.10.1). Results from the quantile–quantile plot of  $P$ -values from the pooled analysis of the GWAS cohorts indicated no systematic type-I error inflation ( $\lambda_{GC} = 1.018$ , Fig. 3).

The five replication studies were analyzed separately using linear regression adjusted for case–control status, age at blood collection, BMI and serum cholesterol. Serum retinol was transformed by taking the natural log and then adjusted for batch. Results across the replication studies were synthesized using a fixed-effects model; heterogeneity in SNP-retinol associations was assessed using Cochran's  $Q$  statistic.

For the meta-analysis combining the discovery and replication studies, we combined the study-specific beta estimates for the discovery (pooled ATBC and PLCO samples) and the five replication studies (T2D, CHD, CGEMS, InCH-male and InCH-female) using a fixed-effects model; heterogeneity in SNP-retinol associations was assessed using Cochran's  $Q$  statistic.

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

*Conflict of Interest statement.* None declared.

### FUNDING

The ATBC Study was supported by US Public Health Service contracts N01-CN-45165, N01-RC-45035, N01-RC-37004 and HHSN261201000006C from the National Cancer Institute, Department of Health and Human Services, and by funding from the Intramural Research Program of the National Cancer Institute. S.H. is supported in part by training grant NIH 5 T32 CA09001-35. Funding to pay the Open Access publication charges for this article was provided by the Intramural Program of the US National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

### REFERENCES

- Krinsky, N.I. and Johnson, E.J. (2005) Carotenoid actions and their relation to health and disease. *Mol. Aspects Med.*, **26**, 459–516.
- Sommer, A. (1998) Xerophthalmia and vitamin A status. *Prog. Retin. Eye Res.*, **17**, 9–31.
- Sesso, H.D., Buring, J.E., Norkus, E.P. and Gaziano, J.M. (2005) Plasma lycopene, other carotenoids, and retinol and the risk of cardiovascular disease in men. *Am. J. Clin. Nutr.*, **81**, 990–997.
- Mondul, A.M., Watters, J.L., Mannisto, S., Weinstein, S.J., Snyder, K., Virtamo, J. and Albanes, D. (2011) Serum retinol and risk of prostate cancer. *Am. J. Epidemiol.*, **173**, 813–821.
- Gueguen, S., Leroy, P., Gueguen, R., Siest, G., Visvikis, S. and Herbeth, B. (2005) Genetic and environmental contributions to serum retinol and alpha-tocopherol concentrations: the Stanislas Family Study. *Am. J. Clin. Nutr.*, **81**, 1034–1044.
- Biesalski, H.K., Frank, J., Beck, S.C., Heinrich, F., Illek, B., Reifen, R., Gollnick, H., Seeliger, M.W., Wissinger, B. and Zrenner, E. (1999) Biochemical but not clinical vitamin A deficiency results from mutations in the gene for retinol binding protein. *Am. J. Clin. Nutr.*, **69**, 931–936.
- Episkopou, V., Maeda, S., Nishiguchi, S., Shimada, K., Gaitanaris, G.A., Gottesman, M.E. and Robertson, E.J. (1993) Disruption of the transthyretin gene results in mice with depressed levels of plasma retinol and thyroid hormone. *Proc. Natl Acad. Sci. USA*, **90**, 2375–2379.
- Ferrucci, L., Perry, J.R., Matteini, A., Perola, M., Tanaka, T., Silander, K., Rice, N., Melzer, D., Murray, A., Cluett, C. *et al.* (2009) Common variation in the beta-carotene 15,15'-monooxygenase 1 gene affects circulating levels of carotenoids: a genome-wide association study. *Am. J. Hum. Genet.*, **84**, 123–133.
- Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K.C., McCullough, M.L., Gallicchio, L., Jacobs, E.J., Ascherio, A., Helzlsouer, K., Jacobs, K.B. *et al.* (2010) Genome-wide association study of circulating vitamin D levels. *Hum. Mol. Genet.*, **19**, 2739–2745.
- Major, J.M., Yu, K., Wheeler, W., Zhang, H., Cornelis, M.C., Wright, M.E., Yeager, M., Snyder, K., Weinstein, S.J., Mondul, A. *et al.* (2011) Genome-wide association study identifies common variants associated with circulating vitamin E levels. *Hum. Mol. Genet.*, **20**, 3876–3883.
- Jacobsson, J.A., Danielsson, P., Svensson, V., Klovin, J., Gyllensten, U., Marcus, C., Schiöth, H.B. and Fredriksson, R. (2008) Major gender difference in association of FTO gene variant among severely obese children with obesity and obesity related phenotypes. *Biochem. Biophys. Res. Commun.*, **368**, 476–482.
- Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Magi, R. *et al.* (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.*, **42**, 949–960.
- (1994) The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann. Epidemiol.*, **4**, 1–10.
- Hayes, R.B., Sigurdson, A., Moore, L., Peters, U., Huang, W.Y., Pinsky, P., Reding, D., Gelmann, E.P., Rothman, N., Pfeiffer, R.M. *et al.* (2005) Methods for etiologic and early marker investigations in the PLCO trial. *Mutat. Res.*, **592**, 147–154.
- Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870–874.
- Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Pare, G. *et al.* (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum. Mol. Genet.*, **19**, 2706–2715.
- Amundadottir, L., Kraft, P., Stolzenberg-Solomon, R.Z., Fuchs, C.S., Petersen, G.M., Arslan, A.A., Bueno-de-Mesquita, H.B., Gross, M., Helzlsouer, K., Jacobs, E.J. *et al.* (2009) Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat. Genet.*, **41**, 986–990.
- Landi, M.T., Chatterjee, N., Yu, K., Goldin, L.R., Goldstein, A.M., Rotunno, M., Mirabello, L., Jacobs, K., Wheeler, W., Yeager, M. *et al.* (2009) A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.*, **85**, 679–691.

19. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A. *et al.* (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.*, **40**, 310–315.
20. Milne, D.B. and Botnen, J. (1986) Retinol, alpha-tocopherol, lycopene, and alpha- and beta-carotene simultaneously determined in plasma by isocratic liquid chromatography. *Clin. Chem.*, **32**, 874–876.
21. Steghens, J.P., van Kappel, A.L., Riboli, E. and Collombel, C. (1997) Simultaneous measurement of seven carotenoids, retinol and alpha-tocopherol in serum by high-performance liquid chromatography. *J. Chromatogr. B Biomed. Sci. Appl.*, **694**, 71–81.
22. Sporn, M.B., Roberts, A.B. and Goodman, D.S. (1994) *The Retinoids: Biology, Chemistry, and Medicine*. New York: Raven Press.
23. El-Soheby, A., Baylin, A., Kabagambe, E., Ascherio, A., Spiegelman, D. and Campos, H. (2002) Individual carotenoid concentrations in adipose tissue and plasma as biomarkers of dietary intake. *Am. J. Clin. Nutr.*, **76**, 172–179.
24. Olmedilla-Alonso, B., Granado-Lorencio, F. and Blanco-Navarro, I. (2005) Carotenoids, retinol and tocopherols in blood: comparability between serum and plasma (Li-heparin) values. *Clin. Biochem.*, **38**, 444–449.
25. Sowell, A.L., Huff, D.L., Yeager, P.R., Caudill, S.P. and Gunter, E.W. (1994) Retinol, alpha-tocopherol, lutein/zeaxanthin, beta-cryptoxanthin, lycopene, alpha-carotene, trans-beta-carotene, and four retinyl esters in serum determined simultaneously by reversed-phase HPLC with multiwavelength detection. *Clin. Chem.*, **40**, 411–416.