③液体クロマトグラフ質量分析計を用いた糖リン酸・有機酸などアニオン系分子を中心とした分析系

④液体クロマトグラフ質量分析計を用いたアミノ酸・有機酸などカチオン系分子を中心とした分析系

得られた測定結果をそれぞれデータ解析に供し、半定量データの取得を行った。

（倫理面への配慮）

前向き大規模コホート研究において既に収集されている血漿検体の分析を実施するにあたり、はじめに、神戸大学大学院医学研究科等医学倫理委員会の承認を得るとともに、その承認内容に基づき、研究を実施した。

C. 研究結果

はじめに、4つの分析プラットホームによる血漿検体分析の妥当性について検討し、それぞれの分析プラットホームで分析可能な血漿中代謝物リストを作成した。続けて、多目的コホート研究参加者のうち、アンケート情報がない対象者から構築したコホート内症例対照研究の血漿検体（Training set）の分析を開始し、現在、分析進行中である。

D. 考察

ガスクロマトグラフ質量分析計、あるいは、液体クロマトグラフ質量分析計を用いた 4 つの分析プラットホームを採用することで、数多くの血漿中代謝物の分析が可能である。分析できる代謝物数が多ければ、大腸がんに対する新規リスク要因を見つける可能性も高くなり、低分子代謝物を網羅的に解析する技術メタボローム解析を、前向き大規模コホート研究において既に収集されている血漿検体を用いた新規リスク要因探索に供することは意義があるとともに、その実現性も高いと考える。

E. 結論

本年度、ガスクロマトグラフ質量分析計、あるいは、液体クロマトグラフ質量分析計を用いた 4 つの分析プラットホームによる血漿代謝物分析システムを構築でき、さらに、前向き大規模コホート研究において既に収集されている血漿検体のうち、Training set の分析を開始できた。当初の計画通り、来年度も引き続き、Training set の分析を進めていき、大腸がんに対する新規リスク要因候補を見出していく。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

1) 山中　広大、小林　隆、西海　信、東　健、吉田　優（2015）血清メタボローム解析にもとづく早期大腸癌診断のバイオマーカー. G.I.Research 2 月号, 第 23 巻, 第 1 号, 46-51.

2) Fujiwara Y, Kobayashi T, Chayahara N, Imamura Y, Toyoda M, Kiyota N, Mukohara T, Nishiumi S, Azuma T, Yoshida M, Minami H. (2014) Metabolomics evaluation of serum markers for cachexia and their intra-day variation in patients with advanced pancreatic cancer. PLoS One. 9, e113259.

3) Matsubara A, Izumi Y, Nishiumi S, Suzuki M, Azuma T, Fukusaki E, Bamba T, Yoshida M. (2014) Supercritical fluid extraction as a preparation method for mass spectrometry of dried blood spots. Journal of Chromatography B, 969,

199-204.

2. 学会発表
該当なし

H. 知的財産権の出願・登録状況（予定を含む）
1. 特許取得
該当なし

2. 実用新案登録
該当なし

3. その他
該当なし

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（業務項目）

リスク予測モデルの構築と検証に関する研究

担当責任者　　口羽　文　　(独)国立がん研究センター研究支援センター生物統計部
生物統計室研究員

研究要旨

当該担当責任者は、ケース・コホート研究デザインおよびコホート内症例対照研究デザインにより構築した予測モデルの精度評価方法に関する統計学的方法論の検討を行うことを業務項目として分担している。平成 26 年度は、ケース・コホート研究デザインおよびコホート内症例対照研究デザインから得られるデータを用いて C-index を推定する方法について検討することにした。Inverse Probability Weighting（IPW）法では、まず各対象者への重み w を求める必要がある。研究デザインに基づくサンプリング確率のほか、あるデータが観測されるかどうかを適切な回帰モデルを用いて推定する方法も考えられる。C-index 推定に対する重みの推定方法の影響は、今後の検討課題である。また、分散の推定方法も整理する必要がある。予測モデルの良さを評価するための指標の一つである C-index に対して、ケース・コホート研究あるいはコホート内症例対照研究から推定するために重み付き C-index の定義をまとめた。平成 27 年度は、考察で述べた課題に加え、C-index 以外の予測性能評価指標についても整理を行い、本委託事業で構築する予測モデルの評価方法を確立する予定である。

A. 研究目的

ゲノムワイド関連研究（GWAS）に代表されるように、血液サンプルから得られる網羅的なオミックスデータの解析から、SNP などの遺伝要因やバイオマーカーが、新たなリスク要因として次々と発見されている。これらの要因が，疾患の発症予測にどの程度寄与するかを評価することは重要な課題の一つである。

予測モデルの良さは、主に calibration（モデルから予測される疾患発症確率と実際のリスクとの一致の程度）と discrimination（予測モデルが、将来疾患を発症する人としない人をどの程度区別できるか）の二つの面から評価される。ROC 曲線の曲線下面積（AUC）は、discrimination を評価するための代表的な統計量であり、最も良く用いられる指標の一つである。

本委託事業で行われる多くの研究では、ケース・コホート研究デザイン、あるいはコホート内症例対照研究デザインを用いて SNP や他のバイオマーカーと疾患との関連を検討することが計画されている。本研究では「多目的コホート研究」から得られたサンプリングデータを用いて予測モデルの良さを評価するための方法を整理する。本

年度は、C-index の推定について検討することにした。

B. 研究方法

ケース・コホート研究デザインあるいはコホート内症例対照研究デザインから得られたデータを用いて予測モデルの評価を行うために、重み付き C-index の定義をまとめた。

（倫理面への配慮）

方法論の検討であり、倫理面への配慮が必要となるデータなどは使用していない。

C. 研究結果

本研究で、予測確率とは予測モデルから推定される疾患発症確率のことと定義する。C-index は、対象者のペアが実際のアウトカムと予測確率に関して concordance か discordance かによって定義される。疾患発症有無などの 2 値のアウトカムの予測を考える場合には、C-index は AUC と等しくなり、疾患を発症する人の方がしない人よりも大きい予測確率を持つ確率と解釈できる。

Harrell ら（1996）は、C-index を生存時間データへ拡張している。この場合、C-index は疾患を早く発症する人の方が高い予測確率を持つ確率と定義される。ここで、有効ペア(i,j)を、対象者のあらゆるペアのうち、少なくともどちらかが疾患を発症しているペアとする。ここでは、対象者 j は必ず疾患を発症しているものとする。対象者 i の予測確率を p_i、観測された疾患発症までの時間を t_i とする。$C_{i,j}$ は対象者ペア(i, j)が concordance かどうかを示す指示変数とし、$p_i>p_j$ かつ $t_i<t_j$ あるいは $p_i<p_j$ かつ $t_i>t_j$ であれば $C_{i,j}$ =1、それ以外では $C_{i,j}$ =0 とする。$D_{i,j}$ は discordant ペアを示す指示変数とし、$p_i>p_j$ かつ $t_i>t_j$ あるいは $p_i<p_j$ かつ $t_i<t_j$ であれば $D_{i,j}$ =1、それ以外では $D_{i,j}$ =0 とする。$C_i=\sum_j C_{i,j}$、$D_i=\sum_j D_{i,j}$ とすると、Harrell ら（1996）の C-index は以下のように推定できる。

$$\hat{C}index = \frac{\sum_i C_i}{\sum_i (C_i + D_i)}$$

ケース・コホート研究、あるいはコホート内症例対照研究における曝露効果の推定方法として、サンプリング確率を用いた Inverse Probability Weighting（IPW）法に基づく方法がさまざま提案されている。ここで、C-index についても重み付き推定を考える。対象者 i のサンプリング確率の逆数 w_i を重みとして用いれば、ケース・コホート研究、あるいはコホート内症例対照研究データを用いて C-index は以下のように推定できる（Ganna ら, 2012）。ただし、疾患を発症した対象者は確率 1 でサンプリングされるものとする。

$$\hat{wC}index = \frac{\sum_i w_i C_i}{\sum_i w_i (C_i + D_i)}$$

D. 考察

IPW 法では、まず各対象者への重み w を求める必要がある。デザインに基づくサンプリング確率のほか、あるデータが観測されるかどうかを適切な回帰モデルを用いて推定する方法も考えられる。C-index 推定に対する重みの推定方法の影響は今後の検討課題である。また、分散の推定方法も整理する必要がある。

E. 結論

予測モデルの良さを評価するための指標の一つである C-index に対して、ケース・コホート研究あるいはコホート内症例対照研究から推定するために重み付き C-index の定義をまとめた。平成 27 年度は、考察で述べた課題に加え、C-index 以外の予測性能評価

指標についても整理を行い、本委託事業で構築する予測モデルの評価方法を確立する予定である。

F. 健康危険情報
なし

G. 研究発表
1. 論文発表
1) Inamura K, Yamauchi M, Nishihara R, Lochhead P, Qian ZR, **Kuchiba A**, Kim SA, Mima K, Sukawa Y, Jung S, Zhang X, Wu K, Cho E, Chan AT, Meyerhardt JA, Harris CC, Fuchs CS, Ogino S. Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis. **Journal of the National Cancer Institute**. 2014; 106(9).

2) Li T, Liao X, Lochhead P, Morikawa T, Yamauchi M, Nishihara R, Inamura K, Kim SA, Mima K, Sukawa Y, **Kuchiba A**, Imamura Y, Baba Y, Shima K, Meyerhardt JA, Chan AT, Fuchs CS, Ogino S, Qian ZR. SMO expression in colorectal cancer: associations with clinical, pathological and molecular features. **Annals of Surgical Oncology**. 2014; 21(13):4164-73.

3) Imamura Y, Lochhead P, Yamauchi M, **Kuchiba A**, Qian ZR, Liao X, Nishihara R, Jung S, Wu K, Nosho K, Wang YE, Peng S, Bass AJ, Haigis KM, Meyerhardt JA, Chan AT, Fuchs CS, Ogino S. Analyses of Clinicopathological, Molecular, and Prognostic Associations of KRAS Codon 61 and Codon 146 Mutations in Colorectal Cancer: Cohort Study and Literature Review. **Molecular Cancer**. 2014; 13(1):135.

2. 学会発表
1) **Kuchiba A**, Wang M, Spiegelman D. Two-stage approach for identifying tumor subtypes associated with an exposure. The 2014 Joint Statistical Meetings, Boston, MA. August 2-7, 2014.

H. 知的財産権の出願・登録状況（予定を含む）
1. 特許取得
なし

2. 実用新案登録
なし

3. その他
なし

学 会 等 発 表 実 績

委託業務題目「前向き大規模コホート研究において既に収集されているがん罹患前試料・情報を用いた発がんリスク要因の探索と層別化に関する研究」

機関名　国立がん研究センターおよび国立大学法人神戸大学

１．学会等における口頭・ポスター発表

発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
Two-stage approach for identifying tumor subtypes associated with an exposure.（口頭）	Kuchiba A, Wang M, Spiegelman D.	The 2014 Joint Statistical Meetings, Boston, MA	August 2-7, 2014	国外

２．学会誌・雑誌等における論文掲載

掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌等名）	発表した時期	国内・外の別
Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A.	Wang H, Burnett T, Kono S, Haiman CA, Iwasaki M, Wilkens LR, Loo LW, Van Den Berg D, Kolonel LN, Henderson BE, Keku TO, Sandler RS, Signorello LB, Blot WJ, Newcomb PA, Pande M, Amos CI, West DW, Bézieau S, Berndt SI, Zanke BW, Hsu L; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Linder NM	Nat Commun	2014 Aug	国外

Plasma isoflavones and risk of primary liver cancer in Japanese women and men with hepatitis virus infection: a nested case-control study.	Michikawa T, Inoue M, Sawada N, Tanaka Y, Yamaji T, Iwasaki M, Shimazu T, Sasazuki S, Mizokami M, Tsugane S; for the Japan Public Health Center-based Prospective Study Group.	Cancer Epidemiol Biomarkers Prev	2015 Mar	国外
Plasma insulin, C-peptide and blood glucose and the risk of gastric cancer: the Japan Public Health Center-based prospective study.	Hidaka A, Sasazuki S, Goto A, Sawada N, Shimazu T, Yamaji T, Iwasaki M, Inoue M, Noda M, Tajiri H, Tsugane S; JPHC Study Group.	Int J Cancer	2015 Mar	国外
Genetic polymorphisms of ADH1B, ADH1C and ALDH2, alcohol consumption, and the risk of gastric cancer: the Japan Public Health Center-based prospective study.	Hidaka A, Sasazuki S, Matsuo K, Ito H, Sawada N, Shimazu T, Yamaji T, Iwasaki M, Inoue M, Tsugane S; JPHC Study Group.	Carcinogenesis	2015 Feb	国外
血清メタボローム解析にもとづく早期大腸癌診断のバイオマーカー	山中　広大、小林　隆、西海　信、東　健、吉田　優	G.I.Research	2015年 2月	国内
Metabolomics evaluation of serum markers for cachexia and their intra-day variation in patients with advanced pancreatic cancer.	Fujiwara Y, Kobayashi T, Chayahara N, Imamura Y, Toyoda M, Kiyota N, Mukohara T, Nishiumi S, Azuma T, Yoshida M, Minami H.	PLoS ONE	2014 Nov	海外

Supercritical fluid extraction as a preparation method for mass spectrometry of dried blood spots	Matsubara A, Izumi Y, Nishiumi S, Suzuki M, Azuma T., Fukusaki E., Bamba T., Yoshida M.	Journal of Chromatography B	2014 Aug	海外
Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis.	Inamura K, Yamauchi M, Nishihara R, Lochhead P, Qian ZR, Kuchiba A, Kim SA, Mima K, Sukawa Y, Jung S, Zhang X, Wu K, Cho E, Chan AT, Meyerhardt JA, Harris CC, Fuchs CS, Ogino S.	The Journal of the National Cancer Institute	2014 Sep	国外
SMO expression in colorectal cancer: associations with clinical, pathological and molecular features.	Li T, Liao X, Lochhead P, Morikawa T, Yamauchi M, Nishihara R, Inamura K, Kim SA, Mima K, Sukawa Y, Kuchiba A, Imamura Y, Baba Y, Shima K, Meyerhardt JA, Chan AT, Fuchs CS, Ogino S, Qian ZR.	Annals of Surgical Oncology	2014 Dec	国外
Analyses of Clinicopathological, Molecular, and Prognostic Associations of KRAS Codon 61 and Codon 146 Mutations in Colorectal Cancer: Cohort Study and Literature Review.	Imamura Y, Lochhead P, Yamauchi M, Kuchiba A, Qian ZR, Liao X, Nishihara R, Jung S, Wu K, Nosho K, Wang YE, Peng S, Bass AJ, Haigis KM, Meyerhardt JA, Chan AT, Fuchs CS, Ogino S.	Molecular Cancer	2014 May	国外

nature COMMUNICATIONS

ARTICLE

Received 21 Jan 2014 | Accepted 7 Jul 2014 | Published 8 Aug 2014

DOI: 10.1038/ncomms5613

Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in *VTI1A*

Hansong Wang[1], Terrilea Burnett[1], Suminori Kono[2], Christopher A. Haiman[3], Motoki Iwasaki[4], Lynne R. Wilkens[1], Lenora W.M. Loo[1], David Van Den Berg[3], Laurence N. Kolonel[1], Brian E. Henderson[3], Temitope O. Keku[5], Robert S. Sandler[5], Lisa B. Signorello[6,7], William J. Blot[8,9], Polly A. Newcomb[10], Mala Pande[11], Christopher I. Amos[12], Dee W. West[13], Stéphane Bézieau[14], Sonja I. Berndt[15], Brent W. Zanke[16], Li Hsu[10], Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)*, Noralane M. Lindor[17], Robert W. Haile[18], John L. Hopper[19], Mark A. Jenkins[19], Steven Gallinger[20], Graham Casey[3], Colon Cancer Family Registry (CCFR)*, Stephanie L. Stenzel[3], Fredrick R. Schumacher[3], Ulrike Peters[10], Stephen B. Gruber[3], Colorectal Transdisciplinary Study (CORECT)*, Shoichiro Tsugane[4], Daniel O. Stram[3] & Loïc Le Marchand[1]

The genetic basis of sporadic colorectal cancer (CRC) is not well explained by known risk polymorphisms. Here we perform a meta-analysis of two genome-wide association studies in 2,627 cases and 3,797 controls of Japanese ancestry and 1,894 cases and 4,703 controls of African ancestry, to identify genetic variants that contribute to CRC susceptibility. We replicate genome-wide statistically significant associations ($P<5\times10^{-8}$) in 16,823 cases and 18,211 controls of European ancestry. This study reveals a new pan-ethnic CRC risk locus at 10q25 (rs12241008, intronic to *VTI1A*; $P=1.4\times10^{-9}$), providing additional insight into the aetiology of CRC and highlighting the value of association mapping in diverse populations.

[1] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96822, USA. [2] Department of Preventive Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8581, Japan. [3] Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California 90033, USA. [4] Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo 104-0045, Japan. [5] Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [6] Department of Epidemiology, Harvard School of Public Health, and the Channing Division of Network Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [7] Dana-Farber/Harvard Cancer Center, Boston, Massachusetts 02115, USA. [8] Division of Epidemiology, Vanderbilt University Medical Center/Vanderbilt-Ingram Cancer Center, Nashville, Tennessee 37235, USA. [9] International Epidemiology Institute, Rockville, Maryland 20850, USA. [10] Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington 19024, USA. [11] Department of Gastroenterolgy—Research, the University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [12] Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire 03755, USA. [13] Cancer Prevention Institute of California, Fremont, California 94538, USA. [14] Service de Génétique Médicale, CHU Nantes, 44093 Nantes, France. [15] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [16] Division of Hematology, Faculty of Medicine, University of Ottawa and Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada K1J8M5. [17] Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona 85054, USA. [18] Stanford Cancer Institute, Stanford, California 94305, USA. [19] Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria 3010, Australia. [20] Cancer Care Ontario, Toronto, Ontario, Canada M5G 2L3. * List of members and affiliations appears at the end of the paper. Correspondence and requests for materials should be addressed to L.L.M. (email: loic@cc.hawaii.edu).

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer deaths in the United States. Genetics is known to play an important role in CRC susceptibility[1]. However, genome-wide association studies (GWASs), mostly conducted in European-descent populations, have only identified 30 common risk variants (22 independent loci) for CRC, markedly fewer than for prostate or breast cancer.

To discover additional risk loci for this cancer, we combine, via a meta-analysis, two GWASs of CRC in populations of Japanese and African American ancestry. The top associations for single-nucleotide polymorphisms (SNPs) in *VTI1A* are replicated in European-descent populations.

Results

In the first GWAS, Japanese samples ($n=6,424$) were identified from the Multiethnic Cohort study (MEC), the Colorectal Cancer Family Registry (CCFR), the Japan Public Health Center cohort study (JPHC) and three case–control studies in Hawaii (CR2&3) and in Fukuoka and Nagano, Japan (Supplementary Table 1). Blood leukocyte DNA samples were genotyped on the Illumina 1M-Duo or the Illumina 660W-Quad arrays, yielding, after quality control (QC) procedures, data for 323,852 SNPs available for all Japanese samples (see Methods and Supplementary Methods). Un-typed markers or markers with partly missing values were imputed with BEAGLE[2] using East Asians from the 1000 Genomes Project (phase 1, release 3) as the reference panel. The second GWAS of African American samples ($n=6,597$) (Supplementary Table 2) were identified from the MEC, CCFR, the Southern Community Cohort Study (SCCS), the MD Anderson Cancer Center, the University of North Carolina CanCORS study (UNC-CanCORS) and Rectal Cancer Study (UNC-Rectal), and from the Prostate, Lung, Colorectal and Ovarian Cancer Screening (PLCO) Trial. African American samples were genotyped using the Illumina 1M-Duo bead arrays (except 170 PLCO subjects on Illumina Omni 2.5M). Imputation was performed with BEAGLE using Europeans and Africans from the 1000 Genomes Project (phase 1, release 3) as reference panels. Over 4.2 million genotyped or imputed autosomal markers were available for both studies.

In both GWASs, cases and controls were well matched with regard to genetic ancestry based on principal component (PC) analyses (Methods and Supplementary Figs 1 and 2). We used logistic regression within each ethnic group to test for the association of SNP dosage with CRC risk, adjusting for age at blood draw, sex and the first four PCs. The genomic control[3] inflation factor (λ) was 1.04 for each individual study, indicating little effect of population stratification after controlling for global ancestries (Supplementary Fig. 3).

After combining the two GWASs, we observed three SNPs in the *VTI1A* gene on chromosome 10q25 to be statistically significant at the genome-wide significance level ($P<5\times10^{-8}$; Fig. 1; Table 1). The strongest association was for rs12241008 (114,280,702 bp) (odds ratio (OR) = 1.19, 95% confidence interval (CI) 1.12–1.26, $P=2.9\times10^{-8}$, allele frequency 0.19 and 0.25 in African Americans and Japanese, respectively), with highly consistent associations in both populations ($I^2=0$). The other two SNPs, rs7894915 (114,277,039 bp, $P=4.8\times10^{-8}$) and rs10082356 (114,278,181 bp, $P=4.9\times10^{-8}$), were in high linkage disequilibrium (LD) with rs12241008 (r^2 from 0.80 to 1.0 in East Asians, Africans and Europeans), with risk estimates almost identical to those for rs12241008 (Supplementary Tables 3 and 4). This locus has not previously been reported to be associated with CRC.

We subsequently replicated these associations in two large CRC consortia of European-descent populations (allele frequency 0.09): Colorectal Transdisciplinary Study (CORECT) with 7,561 cases and 6,328 controls from eight participating studies (combined OR = 1.09, $P=0.036$, Table 1), and the Genetics

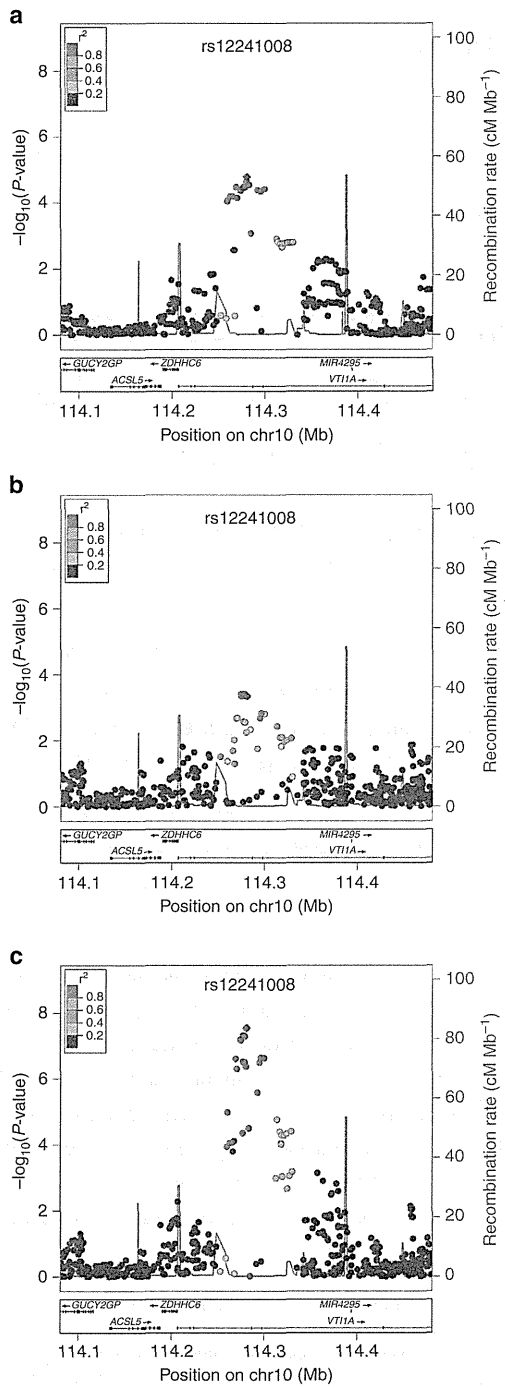

Figure 1 | Regional *P*-value plots for the new colorectal cancer susceptibility locus at 10q25. Results in the Japanese (**a**) ($n=6,424$), African Americans (**b**) ($n=6,595$) and in the combined data (Japanese and African Americans) (**c**) are displayed. The SNP with the smallest *P*-value from meta-analysis in the combined data ($n=13,019$), rs12241008, is shown as a purple diamond. r^2 is in relation to this SNP from the 1000 Genomes Project in East Asians (**a**,**c**) or in Africans (**b**). The plots were generated using LocusZoom[26].

Table 1 | Most strongly associated SNP in the new colorectal cancer susceptibility locus 10q25.

SNP	BP	Alleles*	Study	RAF		Sample size		OR (95% CI)	P	I^2 (%)
				Control	Case	Control	Case			
rs12241008	114280702	C/T	AA†	0.19	0.22	4,702	1,893	1.19 (1.08-1.31)	4.6×10^{-4}	
			JPN‡	0.25	0.28	3,797	2,627	1.19 (1.10-1.29)	1.6×10^{-5}	
			AA + JPN			8,499	4,520	1.19 (1.12-1.26)	2.9×10^{-8}	0
			Replication							
			CORECT†	0.094	0.10	6,328	7,561	1.09 (1.01-1.19)	0.036	40
			GECCO‡	0.090	0.097	11,883	9,262	1.09 (1.02-1.17)	0.018	0
			CORECT + GECCO			18,211	16,823	1.09 (1.03-1.15)	0.0015	0
			Combined			26,710	21,343	1.13 (1.09-1.18)	1.4×10^{-9}	53

AA, African Americans; CI, confidence interval; JPN, Japanese; OR, odds ratio; RAF, risk allele frequency.
*Risk allele/other allele.
†Genotyped in AA and CORECT substudies.
‡Imputed with $R^2 \geq 0.90$ in JPN and in GECCO substudies.

and Epidemiology of Colorectal Cancer Consortium (GECCO) with 9,262 cases and 11,883 controls from 18 participating studies (combined OR = 1.09, $P = 0.018$, Table 1). The combined P-value for rs12241008 in the Japanese, African Americans and Europeans was 1.4×10^{-9} (OR = 1.13, 95% CI 1.09–1.18). A meta-analysis using individual study-level statistics yielded similar results ($P = 1.5 \times 10^{-9}$). Although risk estimates were consistent across individual studies ($I^2 = 8\%$, $P_{het} = 0.35$, d.f. = 27) (see the forest plot in Supplementary Fig. 4), there was some evidence for heterogeneity in effects across ethnic groups ($I^2 = 53\%$, $P_{het} = 0.12$, d.f. = 2). This possible heterogeneity in effects, along with the low allele frequencies observed in European-descent populations (and therefore low power), could partially explain why previous GWASs in Europeans failed to identify this locus, and thus emphasizes the importance of conducting GWASs in ethnically diverse populations.

Similar results were observed for rs7894915 and rs10082356 when the data were combined with the European-descent GWASs ($P = 1.6 \times 10^{-9}$ and 1.5×10^{-9}, respectively). Nine other SNPs in this region (located within 12 kb in the same LD block as rs12241008 in East Asians, Africans and Europeans, Supplementary Fig. 5) also had P-values $< 5 \times 10^{-8}$ when all data were combined (Supplementary Tables 3 and 4). However, the strongest association signal was still with rs12241008 (Supplementary Fig. 5) and none of the other nearby SNPs within 200 kb represented an independent signal after conditioning on rs12241008 in the African American and Japanese GWASs. Results for these 12 SNPs were similar (change in OR $< 1.2\%$) with or without adjustment for local ancestry estimates among African Americans (Methods).

There was no important heterogeneity in ORs in the Japanese or the African American data by anatomical site (colon versus rectal cancer) (P-values > 0.74), by stage (regional/distant versus local/*in situ*) (P-values > 0.6), by age of diagnosis (P-values > 0.7) or by sex ($P = 0.04$ in African Americans and 0.80 in Japanese) for the most significant marker rs12241008 (stratified analysis results are in Supplementary Table 5).

No association reached the genome-wide significance threshold (5×10^{-8}) in the Japanese GWAS when analysed separately. One SNP on chromosome 7, rs79453636, passed this threshold ($P = 2.9 \times 10^{-8}$) in the African American study (Supplementary Fig. 3). However, this association was not replicated (Supplementary Table 6) in the Japanese or in the combined European-descent data (P-values > 0.15).

Out of 30 known CRC susceptibility SNPs, 27 were available for analysis in the Japanese study and 23 (85%) effects were in the same direction as in the original GWAS reports (Supplementary Table 7). Replication and fine-mapping of the known risk loci in the African American study was summarized previously[4]. Twelve out of the 27 associations were replicated in the meta-analysis of the Japanese and African American data ($P < 0.05$) and 23 risk estimates were directionally consistent with those originally reported (Supplementary Table 7). Considerable heterogeneity in disease risk between the two ethnic groups ($I^2 > 50\%$) was observed for six SNPs (Supplementary Table 7).

The three genetic variants with the strongest association with CRC, rs12241008, rs7894915 and rs10082356, are located in intron 3 of the *VTI1A* gene, which encodes vesicle transport through interaction with t-SNAREs 1A. VTI1A is involved in regulating insulin-stimulated trafficking of secretory vesicles enriched with both GLUT4 (glucose transporter) and Acrp30 in adipocytes[5]; it also plays key roles in neuronal development[6] and in selectively maintaining spontaneous neurotransmitter release[7]. A recent GWAS in never-smoking Asian women has identified rs7086803 in intron 7 of *VTI1A* as a lung cancer susceptibility variant (218 kb from and not in LD with rs12241008)[8]. Interestingly, a gene fusion product, *VTI1A-TCF7L2*, was identified in colorectal tumours and shown to promote anchorage-independent growth of cultured tumour cells[9]. The fusion occurs between *VTI1A* exon 3 (chr10: 114,220,869) and *TCF7L2* exon 4 (chr10: 114,760,545) and results in the deletion of both the intron 3 CRC and intron 7 lung cancer risk variants. No coding variant is in LD with the top three SNPs. Among the three *VTI1A* SNPs associated with CRC in this study, rs7894915 and rs10082356 lie in predicted transcriptional regulatory regions, suggesting enhancer and promoter regulatory activities across multiple cell lines (Supplementary Table 8)[10]. We explored regulatory effects of the SNPs correlated with rs12241008 in a *cis*-expression quantitative trait loci analysis in 40 paired colon adjacent-normal and tumour tissue samples from European descent patients[11]. Among the SNPs in high LD ($r^2 > 0.8$) with rs12241008 in East Asians, the intronic SNP rs7081965 (alleles: A/T) affected *VTI1A* expression ($P = 0.003$) in colon tumour tissue. Rs7081965 is also in considerable LD with rs12241008 in Africans ($r^2 = 0.21$, $|D'| = 0.88$) and in Europeans ($r^2 = 0.24$, $|D'| = 1$) in the 1000 Genomes data. Although the association of rs7081965 with CRC was not statistically significant in this study (OR = 1.09 for allele T, $P = 8.2 \times 10^{-6}$ from the three ethnic groups combined), these results provide an interesting lead for future functional investigations.

In summary, this trans-ethnic GWAS identified a new CRC susceptibility locus at 10q25 with directionally consistent associations across three ethnic/racial populations, providing additional insight into the genetic architecture of CRC. Further

work is needed to dissect this genetic signal and to conduct functional studies to uncover the mechanisms underlying this association.

Methods

Japanese subjects and QC on genotypes. Details on study design and basic characteristics for each study are provided in Supplementary Methods. Briefly, 1,703 MEC Japanese American subjects were genotyped by the Broad Genotyping Center on the Illumina 1M-Duo Array and 1,602 (803 cases, 799 controls) passed their initial QC filters. To maximize sample size, initially 'failed' samples on five plates were re-clustered with a customized genotype calling algorithm—this step recovered 42 additional MEC subjects (23 cases, 19 controls), although not all SNPs on the array were preserved. To increase statistical power and to provide a larger control pool, 1,033 prostate cancer-free men and 808 breast cancer-free women genotyped on the Illumina 660W-Quad platform were drawn from the MEC prostate cancer[12] and breast cancer[13] studies, respectively.

Japanese from the following studies were all genotyped on the Illumina 1M-Duo array by the University of Southern California (USC) Epigenome Center: 697 from CCFR (384 cases, 313 controls), 155 cases from CR2&3, 1,463 from Fukuoka, Japan (685 cases, 778 controls), 212 from Nagano, Japan (106 cases, 106 controls) and 1,332 from JPHC (670 cases, 662 controls). In general, all genotyped samples were examined and excluded according to the following: (1) call rates $<90\%$, 95% or 97% depending on the batches, (2) missing on basic covariates (age, sex or disease status), (3) gender mismatch, that is, the reported sex was different from that estimated based on X chromosome inbreeding coefficient F, calculated by PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/), (4) ethnicity outliers, that is, subjects fell out of the Japanese cluster (by visual inspection) on PC plots, where PCs were derived for study subjects as well as unrelated HapMap CEU, YRI and JPT samples with our own *R* program (The Comprehensive R Archive Network http://www.r-project.org/), based on about 20k SNPs with inter-marker distance >100 kb, and (5) close ($\geq$2nd degree) relatives, where relationships were derived from estimated probabilities of sharing 0, 1 or 2 alleles based on genomic data (calculated by PLINK), and relatives were removed in the following order: subjects with most relatives, controls and subjects with lower call rates. All cases were verified by histological records to have invasive carcinoma of the colon or rectum. More details on genotype QC can be found in Supplementary Methods. After QC, the following subjects were retained in analysis: 3,094 from the MEC (797 cases, 2,297 controls), 285 from CCFR (276 cases and 9 controls), 134 cases from CR2&3, 1,411 from Fukuoka, Japan (662 cases, 749 controls), 207 from Nagano, Japan (105 cases, 102 controls) and 1,293 from the JPHC (653 cases, 640 controls).

African American subjects and QC on genotypes. Sample collection and genotyping QC have been described in detail elsewhere[4] and in Supplementary Methods. We genotyped 7,168 African American samples from six studies/centres: the MEC (442 cases, 4,620 controls), CCFR (999 cases, 290 controls), SCCS (164 cases, 160 controls), the MD Anderson Cancer Center (189 cases), UNC-CanCORS (84 AA cases) and UNC-Rectal (112 cases, 108 controls) on the Illumina 1M-duo platform. QC procedures for all subjects were similar to the criteria described for the Japanese study subjects. Included in analysis were 6,427 subjects (4,609 controls, 1,818 cases) on 1,049,327 markers. We also included 170 PLCO samples (76 cases, 94 controls) that were previously genotyped on the Illumina Omni 2.5M array and pre-filtered by the NCI genotyping centre for analysis (527,383 markers that overlapped with other studies). Overall, 6,597 subjects (1,894 cases, 4,703 controls) were used in association testing. Supplementary Table 2 shows the distribution of subjects by participating study.

Imputation. Prediction of un-typed or partly genotyped SNPs was performed with BEAGLE 3.3 (ref. 2) using the 1000 Genomes Project (phase 1, release 3) East Asians as reference panels for the Japanese data and Europeans and Africans for the African American data. Imputation was performed separately for the two ethnic groups with all cases and controls combined. Markers with minor allele frequencies <0.005 in reference panels were excluded from imputation. For the African American data, 10,050,748 markers with imputation accuracy $R^2>0.8$ were kept for association analysis; for Japanese data, 4,266,108 markers with imputation $R^2>0.95$ were retained. Altogether, 4,276,079 autosomal genotyped or imputed markers were available in both populations for meta-analysis.

Analysis of the Japanese and African American GWASs. PCs were calculated as in EIGENSTRAT[14] with our own *R* program, including unrelated HapMap CEU, YRI and JPT samples as population controls. Ethnicity outliers were identified on PC plots by visual inspection and subsequently removed. Pair-wise PC plots suggested that the first two PCs were most informative for global ancestry and the distribution of PCs was similar among all cases and controls in both Japanese and African Americans (Supplementary Figs 1 and 2). Logistic regression of CRC on allelic dosage with adjustment for age at blood draw, sex and the first four PCs was performed to obtain OR estimates and 95% CI of per increase in allele count with PLINK, where age was grouped as <55 years, 5-year intervals from 55 to 80 and ≥ 80 years. The genomic control factor (λ) was estimated from the median of the χ^2 statistics divided by 0.456.

Heterogeneity of genetic effects by site (colon versus rectal cancer, mutually exclusive), stage (regional/distant versus local/*in situ*) and age at diagnosis (≤ 55 versus >55 years) was tested in a case-only analysis. Effect modification by sex was assessed comparing the model with and without the cross-product term. These and additional stratified analyses by site, stage, age at diagnosis and sex were adjusted for age at blood draw, sex (where appropriate), the first four PCs and BMI.

Conditional analyses were performed to examine the independence of association signals in the chromosome 10 region, conditioning on the SNP with the smallest *P*-value. Significance of the additional contribution by other SNPs was calculated based on a likelihood ratio test. These analyses were carried out using *SAS* 9.3.

Local ancestry estimation for African Americans. The percentage of African ancestry (0, 50 or 100%, that is, half of the estimated number of African chromosomes) was inferred for each participant at the putative CRC risk locus on chromosome 10 (± 250 kb) with the LAMP program v2.4 (ref. 15). To summarize local ancestry, for each individual we averaged across all local ancestry estimates that are within the region. The effect of local ancestry was evaluated by examining the relative change in ORs with and without adjustment for local ancestry in logistic regression.

CORECT study for replication. The CORECT study meta-analysis was conducted using germline DNA in the Molecular Epidemiology of Colorectal Cancer study (MECC) (set 1: 484 cases and 498 controls; set 2: 1,120 cases and 820 controls), CCFR (set 1: 1,977 cases and 999 controls; set 2: 1,660 cases and 1,393 controls), Kentucky case–control study (1,038 cases and 1,134 controls), Newfoundland case–control study (548 cases and 538 controls), American Cancer Society CPS II nested case–control study (ACS/CPSII, 539 cases and 469 controls) and the Melbourne nested case–control study (195 cases and 477 controls). All subjects were self-reported whites. The majority of the studies were genotyped using the Affymetrix Axiom CORECT Set containing ~1.3 million SNPs and indels on two physical genotyping chips (Supplementary Table 3). Genotype data were screened based on filters such as call rates, concordance rates, sample relatedness and ethnic outliers. IMPUTE2 (ref. 16) was used to impute missing genotypes based on the cosmopolitan panel of reference haplotypes from Phase I of the 1000 Genomes Project. Imputed genotypes were screened based on stringent imputation quality and accuracy filters (info ≥ 0.7, certainty ≥ 0.9, concordance ≥ 0.9 between directly measured and imputed genotypes after masking input genotypes for genotyped markers only). Associations between genetic variants and CRC risk were tested using a log-additive genetic model within each study, allowing for study-specific adjustment for age, sex, study centre, genotyping batch and 2–4 PCs. More details of each participating study can be found in Supplementary Methods.

GECCO study for replication. The GECCO GWAS consortium has been described before[17–19]. The consortium consisted of European-descent participants within the French Association Study Evaluating RISK for sporadic CRC (ASTERISK, 948 cases and 947 controls); CR2&3 (87 cases and 125 controls); Darmkrebs: Chancen der Verhütung durch Screening (DACHS set 1: 1,710 cases and 1,707 controls; DACHS set 2: 675 cases and 498 controls); Diet, Activity, and Lifestyle Study (DALS set 1: 706 cases and 710 controls; DALS set 2: 410 cases and 464 controls); Health Professionals Follow-up Study (HPFS set 1: 227 cases and 230 controls; HPFS set 2: 176 cases and 172 controls); MEC (328 cases and 346 controls); Nurses' Health Study (NHS set 1: 394 cases and 774 controls; NHS set 2: 159 cases and 181 controls); Ontario Familial Colorectal Cancer Registry (OFCCR, 650 cases and 522 controls); Physician's Health Study (PHS, 382 cases and 389 controls); Postmenopausal Hormone study (PMH, 280 cases and 122 controls); Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO set 1: 533 cases and 1,976 controls; PLCO set 2: 486 cases and 415 controls); VITamins And Lifestyle (VITAL, 285 cases and 288 controls); and the Women's Health Initiative (WHI set 1: 470 cases and 1,529 controls; WHI set 2: 1,006 cases and 1,010 controls). All individual studies were genotyped on Illumina arrays on 240k–730k markers and went through rigorous QC. The genotype data were imputed to increase the density of genetic variants. The haplotypes from the 1000 Genomes Project Phase I were used as the reference panel. Logistic regression of CRC on SNP dosage effect on CRC risk was performed with adjustment for age, sex (when appropriate), centre (when appropriate), smoking status (PHS only), batch effects (ASTERISK only) and the first three PCs from EIGENSTRAT[13] to account for population substructure within each individual study. Additional details on sample collection, genotyping, QC and statistical methods are provided in Supplementary Methods.

All samples were collected with informed consent and all procedures were approved by the Human Research Institutional Review Boards (IRBs) at relevant institutions. Specifically, the study protocols of the Japanese and African Americans' GWASs were approved by the University of Hawaii Human Studies Program and University of Southern California IRB, the IRB in the National Cancer Center, Japan, the Ethics Committee of Kyushu University Faculty of Medical Sciences, the University of North Carolina IRB, Vanderbilt University IRB,

the Fred Hutchinson Cancer Research Center IRB and the MD Anderson Cancer Center IRB. The GECCO portion of this work was approved by the Fred Hutchinson Cancer Research Center IRB. The University of Southern California Health Sciences IRB approved all elements of the CORECT study.

Meta-analysis. A fixed-effect model with inverse variance weighting implemented in METAL[20] was used to combine the results from the Japanese and the African American studies and for further combining with replication studies. Heterogeneity measure I^2 was calculated and Cochran's Q statistic to test for heterogeneity was calculated[21]. For the 12 top hits in the *VTI1A* region at 10q25 (see text), OFCCR in GECCO was excluded because these SNPs did not pass the quality filters in this substudy (Table 1, Supplementary Table 4 and Supplementary Fig. 4). In Supplementary Fig. 5, SNPs that passed the filters in OFCCR were included whenever applicable.

References

1. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *New Engl. J. Med.* **343,** 78–85 (2000).
2. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am. J. Hum. Genet.* **81,** 1084–1097 (2007).
3. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
4. Wang, H. *et al.* Fine-mapping of genome-wide association study-identified risk loci for colorectal cancer in African Americans. *Hum. Mol. Genet.* **22,** 5048–5055 (2013).
5. Bose, A. *et al.* The v-SNARE Vti1a regulates insulin-stimulated glucose transport and Acrp30 secretion in 3T3-L1 adipocytes. *J. Biol. Chem.* **280,** 36946–36951 (2005).
6. Kunwar, A. J. *et al.* Lack of the endosomal SNAREs vti1a and vti1b led to significant impairments in neuronal development. *Proc. Natl Acad. Sci. USA* **108,** 2575–2880 (2011).
7. Ramirez, D. M., Khvotchev, M., Trauterman, B. & Kavalali, E. T. Vti1a identifies a vesicle pool that preferentially recycles at rest and maintains spontaneous neurotransmission. *Neuron* **73,** 121–134 (2012).
8. Lan, Q. *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat. Genet.* **44,** 1330–1335 (2012).
9. Bass, A. J. *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat. Genet.* **43,** 964–968 (2011).
10. The ENCODE Project Consortium *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
11. Loo, L. W. *et al.* cis-Expression QTL analysis of established colorectal cancer risk variants in colon tumors and adjacent normal tissue. *PLoS ONE* **7,** e30477 (2012).
12. Cheng, I. *et al.* Evaluating genetic risk for prostate cancer among Japanese and Latinos. *Cancer Epidemiol. Biomarkers Prev.* **21,** 2048–2058 (2012).
13. Siddiq, A. *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum. Mol. Genet.* **21,** 5373–5384 (2012).
14. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
15. Sankararaman, S., Sridhar, S., Kimmel, G. & Halperin, E. Estimating local ancestry in admixed populations. *Am. J. Hum. Genet.* **9,** 290–303 (2008).
16. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
17. Peters, U. *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum. Genet.* **131,** 217–234 (2012).
18. Peters, U. *et al.* Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology* **144,** 799–807 (2013).
19. Zanke, B. W. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **39,** 989–994 (2007).
20. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
21. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21,** 1539–1558 (2002).
22. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39,** 645–649 (2007).
23. Amundadottir, L. *et al.* Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat. Genet.* **41,** 986–990 (2009).
24. Petersen, G. M. *et al.* A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.* **42,** 224–228 (2010).
25. Landi, M. T. *et al.* 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 (2009).
26. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–2337 (2010).

Acknowledgements

We thank Dr Daniel Mirel, who supervised the genotyping of samples while working at the Broad Institute, Boston, MA, and Xin Sheng, Loreall Pooler, Dr Gary K. Chen and Alex H. Stram at the University of Southern California (USC), and Lucy Shen and Mike Loomis at the University of Hawai'i Cancer Center for their technical assistance. The colorectal cancer GWAS among Japanese and African Americans was funded through US National Institutes of Health (NIH) grants 1R01-CA126895, 1R01-CA126895-S1, 1R01-CA104132 and 2U24-CA074806. Genotyping of the additional MEC controls was funded through NIH grants R01-CA132839, RC2-CA148085, R01-CA1326792 and U01-HG004726, as well as a Department of Defense Breast Cancer Research Program, Era of Hope Scholar Award to CAH (W81XWH-08-1-0383). MEC was funded through NIH grants R37 CA54281, P01 CA033619, and R01 CA63464. The SCCS was funded by NIH grant R01CA092447. Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry, which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry. The UNC studies were supported by grants U01 CA 093326, P50 CA 106991 and R01 CA 66635. The MD Anderson data collection was supported in part by the MD Anderson University Cancer Fund, the MD Anderson Cancer Center Duncan Family Institute for Cancer Prevention and Risk Assessment, the Center for Clinical and Translational Sciences of the University of Texas Health Science Center at Houston, NCI Cancer Center Support Grant (CA16672) and NCI grant (K07CA160753). JPHC was supported by the National Cancer Center Research and Development Fund (since 2011) and a Grant-in-Aid for Cancer Research (from 1989 to 2010) from the Ministry of Health, Labor and Welfare of Japan. The Fukuoka Colorectal Cancer Study was funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan. The CORECT Study is supported by the National Cancer Institute as part of the GAME-ON consortium (U19 CA148107) with additional support from NCI grants (R01 CA81488, P30 CA014089), the National Human Genome Research Institute at the NIH (T32 HG000040) and the National Institute of Environmental Health Sciences at the NIH (T32 ES013678). CCFR (http://www.coloncfr.org/) is supported by the National Cancer Institute, NIH under RFA #CA-95-011 and through cooperative agreements with members of the Colon Cancer Family Registry and PIs of the Australasian Colorectal Cancer Family Registry (U01 CA097735), Familial Colorectal Neoplasia Collaborative Group (U01 CA074799) [USC], Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01 CA074800), Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783), Seattle Colorectal Cancer Family Registry (U01 CA074794) and the University of Hawaii Colorectal Cancer Family Registry (U01 CA074806). The Colon CFR GWAS work was supported by a National Cancer Institute grant (U01CA122839 and P30 CA014089), Australasian Colorectal Cancer Family Registry (U01 CA097735), Seattle Colorectal Cancer Family Registry (U01 CA074794) and Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute, NIH or any of the collaborating centres in the CCFRs, nor does it mention trade names, commercial products or organizations imply endorsement by the US Government or the CCFR. GECCO was funded by NIH grants U01 CA137088 and R01 CA059045. ASTERISK was supported by a Hospital Clinical Research Program (PHRC) and by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). CR2&3 was funded by NIH grant R01 CA60987. DACHS was funded by the German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1) and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814). DALS was funded by the NIH (R01 CA48998 to M.L.S.). HPFS is supported by the NIH (P01 CA 055075, UM1 CA167552, R01 137178, and P50 CA 127003), NHS by the NIH (R01 CA137178, P01 CA 087969 and P50 CA 127003) and PHS by the NIH (R01 CA042182). OFCCR was supported by the NIH through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see the CCFR section above. As a subset of ARCTIC, OFCCR is supported by a GL2 grant from the Ontario Research Fund, the Canadian Institutes of Health Research, and the Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society Research Institute. Thomas J. Hudson and Brent W. Zanke are recipients of Senior Investigator Awards from the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation. PLCO (http://dcp.cancer.gov/plco) was supported by the Intramural Research Program of the Division

of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS[22], CGEMS pancreatic cancer scan (PanScan)[23,24] and the Lung Cancer and Smoking study[25]. The prostate and PanScan study data sets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung data sets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study was provided by NIH, Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446 and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping. PMH was funded by the NIH grant R01 CA076366 to P.A. Newcomb. NHS was supported by NIH grants CA 087969, R01 137178, and P50 CA 127003. VITAL was funded by NIH grant K05 CA154337. The WHI program is funded by the National Heart, Lung, and Blood Institute, and NIH through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C. The ASTERISK study are very grateful to Dr Bruno Buecher, without whom this project would not have existed and also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students. The DACHS study thank all participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benscheid, Muhabbet Celik and Ursula Eilber at DACHS for excellent technical assistance. GECCO would like to thank all those at the Coordinating Center for helping to bring together the data and people that made this project possible. HPFS, NHS and PHS would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Dr Immaculata Devivo and Dr David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA and WY. In addition, this study was approved by the Connecticut Department of Public Health (DPH) Human Investigations Committee. Certain data used in this publication were obtained from the DPH. The authors assume full responsibility for analyses and interpretation of these data. The PLCO thank Drs Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff of the PLCO Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O'Brien and staff, Westat, Inc., and Drs Bill Kopp, Wen Shao, and staff, SAIC-Frederick, and most importantly the study participants for their contributions to making this study possible. The PMH study would like to thank the study participants and staff of the Hormones and Colon Cancer study. The WHI study thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

Author contributions

L.L.M., C.A.H. and D.O.S. contributed to the study concept and design. L.L.M. and T.B. organized the Japanese and African American colorectal cancer consortia. D.V.D.B. supervised genotyping of samples at USC. H.W., L.L.M., D.O.S. and S.L.S. contributed to the statistical analysis. H.W. and L.L.M. drafted the manuscript. L.L.M., L.N.K., B.E.H., S.K., M.I., T.O.K., R.S.S., L.B.S., W.J.B., P.A.N., M.P., C.I.A., D.W.W., S.B., S.I.B., B.W.Z., N.M.L., R.W.H., J.L.H., M.A.J., S.G., G.C., U.P., S.B.G. and S.T. conducted the epidemiological studies that contributed samples to the scan. All authors contributed to the writing of the manuscript, interpretation and discussion of the findings and approved the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wang, H. *et al.* Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in *VTI1A*. *Nat. Commun.* 5:4613 doi: 10.1038/ncomms5613 (2014).

GECCO consortium members:

John A. Baron[21], Sonja I. Berndt[15], Stéphane Bezieau[14], Hermann Brenner[22], Bette J. Caan[23], Christopher S. Carlson[10], Graham Casey[3], Andrew T. Chan[24,25], Jenny Chang-Claude[22], Stephen J. Chanock[15], David V. Conti[3], Keith Curtis[10], David Duggan[26], Charles S. Fuchs[24,25], Steven Gallinger[20], Edward L. Giovannucci[24,25], Stephen B. Gruber[3], Robert W. Haile[18], Tabitha A. Harrison[10], Richard B. Hayes[27], Michael Hoffmeister[22], John L. Hopper[19], Li Hsu[10], Thomas J. Hudson[28,29], David J. Hunter[24,25], Carolyn M. Hutter[15], Rebecca D. Jackson[30], Mark A. Jenkins[19], Shuo Jiao[10], Sébastien Küry[14], Loic Le Marchand[1], Mathieu Lemire[31], Noralane M. Lindor[17], Jing Ma[24,25], Polly A. Newcomb[10], Ulrike Peters[10], John D. Potter[10,32], Conghui Qu[10], Thomas Rohan[33], Robert E. Schoen[34], Fredrick R. Schumacher[3], Daniela Seminara[15], Martha L. Slattery[35], Stephen N. Thibodeau[36,37], Emily White[10] and Brent W. Zanke[16].

CORECT consortium members:

Kendra Blalock[10], Peter T. Campbell[38], Graham Casey[3], David V. Conti[3], Christopher K. Edlund[3], Jane Figueiredo[3], W. James Gauderman[3], Jian Gong[10], Roger C. Green[39], Stephen B. Gruber[3], John F. Harju[40], Tabitha A. Harrison[10], Eric J. Jacobs[38], Mark A. Jenkins[19], Shuo Jiao[10], Li Li[41], Yi Lin[10], Frank J. Manion[40], Victor Moreno[42], Bhramar Mukherjee[40], Ulrike Peters[10], Leon Raskin[3], Fredrick R. Schumacher[3], Daniela Seminara[15], Gianluca Severi[19], Stephanie L. Stenzel[3] and Duncan C. Thomas[3].

CCFR consortium members:

John A. Baron[21], Graham Casey[3], Duncan C. Thomas[3], Fredrick R. Schumacher[3], David V. Conti[3], Michelle Cotterchio[20], Dave Duggan[26], Steven Gallinger[20], Robert Haile[18], John L. Hopper[19], Mark A. Jenkins[19], Jeremy Jass[43], Loic Le Marchand[1], John Grove[1], Noralane M. Lindor[17], Steve Thibodeau[17], Polly A. Newcomb[10], John D. Potter[10,32], Sheri Schully[15], Daniela Seminara[15].

[21]Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599, USA. [22]German Cancer Research Center, 69120 Heidelberg, Germany. [23]Division of Research, Kaiser Permanente Medical Care Program, Oakland, California 94611, USA. [24]Harvard Medical School, Boston, Massachusetts 02115, USA. [25]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [26]Translational Genomics Research Institute, Phoenix, Arizona 85004, USA. [27]Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, New York, New York 10016, USA. [28]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 0A3, Canada. [29]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5G 0A3. [30]Division of Endocrinology, Diabetes and Metabolism, Ohio State University, Columbus, Ohio 43210, USA. [31]Ontario Institute for Cancer Research, Toronto, Ontario, Canada M5G 0A3. [32]Centre for Public Health Research, Massey University, 4474 Palmerston North, New Zealand. [33]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Yeshiva University, Bronx, New York 10461, USA. [34]Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213, USA. [35]Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, Utah 84132, USA. [36]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota 55905, USA. [37]Department of Laboratory Genetics, Mayo Clinic, Rochester, Minnesota 55905, USA. [38]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA. [39]Faculty of Medicine, Memorial University of Newfoundland, St John's, Newfoundland, Canada A1B 3X9. [40]University of Michigan Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan 48109, USA. [41]Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio 44106, USA. [42]Institut d'Investigació Biomèdica de Bellvitge, Institut Catala d'Oncologia Hospitalet, 08908 Barcelona, Spain. [43]St Mark's Hospital, Middlesex HA1 3UJ, UK.

Published OnlineFirst December 26, 2014; DOI: 10.1158/1055-9965.EPI-14-1118

Research Article

Cancer Epidemiology, Biomarkers & Prevention

Plasma Isoflavones and Risk of Primary Liver Cancer in Japanese Women and Men with Hepatitis Virus Infection: A Nested Case–Control Study

Takehiro Michikawa[1,2], Manami Inoue[2,3], Norie Sawada[2], Yasuhito Tanaka[4], Taiki Yamaji[2], Motoki Iwasaki[2], Taichi Shimazu[2], Shizuka Sasazuki[2], Masashi Mizokami[5], and Shoichiro Tsugane[2], for the Japan Public Health Center-based Prospective Study Group

Abstract

Background: Evidence suggests that estrogen plays a preventive role in primary liver cancer development, and it might be thought that isoflavones, which are structurally similar to estrogens and bind to estrogen receptors, are associated with the risk of liver cancer. We investigated this suspected association by measuring plasma concentrations of isoflavones in a nested case–control study of a population-based prospective cohort in Japan.

Methods: From 18,628 target participants ages 40 to 69 years who returned the baseline questionnaire and provided blood samples, we selected those with either hepatitis B or hepatitis C virus infection at baseline (n = 1,544). Among these, 90 (28 women and 62 men) were newly diagnosed with primary liver cancer from 1993 through 2006; they were matched with 175 controls (54 women and 121 men). Plasma concentrations of isoflavones (genistein, daidzein, glycitein, and equol) were measured using triple quadrupole tandem liquid chromatography-mass spectrometry. The ORs of liver cancer development based on plasma concentrations were estimated with a conditional logistic regression model.

Results: Basically, distributions of plasma isoflavone concentrations did not differ between the cases and controls. No statistically significant associations of genistein, daidzein, glycitein, and equol with primary liver cancer risk were found in either women or men.

Conclusions: In middle-aged Japanese women and men with hepatitis virus infection, plasma isoflavones were unassociated with the occurrence of primary liver cancer.

Impact: The role of isoflavones in liver carcinogenesis merits further study using both biomarkers and data on dietary intake of isoflavones. *Cancer Epidemiol Biomarkers Prev; 24(3); 532–7. ©2014 AACR.*

Introduction

Women worldwide have a lower incidence of primary liver cancer, respond better to treatment, and show better survival (1). These data indicate that estrogen plays a preventive role in liver cancer development. In sex hormone–related cancers such as breast cancer, an association is suspected between isoflavones and cancer risk, because isoflavones are structurally similar to 17β-estradiol, have the ability to bind to estrogen receptors, and act not only as estrogen agonists but also as antagonists (2). We previously examined the association between dietary intake of isoflavones and primary liver cancer, and found that isoflavone consumption was positively associated with liver cancer risk among Japanese women (3). Given the preventive effects of estrogen against liver cancer, we thought this positive association might be partly explained by the antiestrogenic effects of isoflavones. The main exposure variable in our previous study was isoflavone consumption as assessed via a food-frequency questionnaire, so clearly a more objective measure was required to confirm the association epidemiologically. Isoflavone concentrations in the blood are superior to dietary assessments as markers reflecting *in vivo* absorption and metabolism (4).

In this study, we examined the effects of plasma isoflavone concentrations on primary liver cancer risk among women and men with hepatitis virus infection, using a nested case–control design based on data from a large-scale population-based prospective cohort study in Japan. As far as we know, no previous prospective studies have examined liver cancer using biomarkers to assess isoflavone exposure.

Materials and Methods

Study population

The study design of the Japan Public Health Center-based Prospective Study (JPHC Study), which began in 1990 for

[1]Environmental Epidemiology Section, Center for Environmental Health Sciences, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan. [2]Epidemiology and Prevention Group, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan. [3]Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [4]Department of Virology and Liver Unit, Nagoya City University Graduate School of Medical Sciences, Mizuho-ku, Nagoya, Japan. [5]The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Note: Study group members are listed in the Acknowledgments section.

Corresponding Author: Manami Inoue, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Phone: 81-3-5841-3688; Fax: 81-3-5841-3637; E-mail: mnminoue@m.u-tokyo.ac.jp

doi: 10.1158/1055-9965.EPI-14-1118

Published OnlineFirst December 26, 2014; DOI: 10.1158/1055-9965.EPI-14-1118

cohort I and in 1993 for cohort II, has been published elsewhere (5). Cohort I included all registered Japanese residents ages 40 to 59 years of 5 public health center areas, and in cohort II, all residents ages 40 to 69 years of 6 other areas. The present study was approved by the Institutional Review Board of the National Cancer Center, Tokyo, Japan.

In this study, we used cohort II data. In cohort II (1993–1994), 56,542 participants (response rate, 82%) answered a baseline questionnaire on sociodemographic characteristics, medical history, smoking and drinking habits, diet, and so on. Of these, 37% voluntarily provided 10 mL of blood at health checkups during the baseline survey (1993–1995). The blood samples were divided into plasma and buffy layers and preserved at −80°C until analysis. We measured hepatitis B surface antigen (HBsAg) by reversed passive hemagglutination with a commercial kit (Institute of Immunology Co., Ltd.) and anti-hepatitis C virus antibody (anti-HCV) with a third-generation immunoassay (Lumipulse II Ortho HCV; Ortho-Clinical Diagnosis K.K.). Study participants were informed of the objectives and methods of the study in writing, and those who answered the questionnaire and donated blood were regarded as having given informed consent to participate. Of these, we selected only those who had no history of cancer at baseline and had provided data on basic characteristics, leaving us with a total of 18,628 participants (6,401 men and 12,227 women).

Follow-up

Participants were followed up from the date of blood collection until December 31, 2006. Information on residence status and survival was obtained annually through residential registries. With a follow-up rate of 99.7%, selection bias due to lost to follow-up was negligible.

Data on primary liver cancer incidence were collected for the JPHC cancer registry from two data sources: major local hospital records and population-based cancer registries. Cases were coded using the International Classification of Diseases for Oncology, Third Edition (code: C22.0; ref. 6). Death certificates were used as a supplementary information source. The proportion of cases for which information was available from death certificates only was 4.7%, indicating satisfactory cancer registry system quality during the study period.

Selection of cases and controls

From the 18,628 participants, we selected 1,544 infected either with hepatitis B virus (HBV; positivity for HBsAg) or hepatitis C virus (HCV; positivity for anti-HCV) at baseline. Up to the end of the study period after blood collection, we identified 91 new cases of primary liver cancer among these 1,544 participants. For each case, we selected 2 controls at random from among the participants with no history of liver cancer when the case was diagnosed. Controls were matched to each patient with respect to age (within 5 years), sex, public health center area, fasting status at blood collection, hepatitis virus infection status (HBV or HCV), and baseline menopausal status (for women). We could not find appropriate matched controls for 1 patient, and found only 1 matched control with a sufficient quantity of plasma for each of 5 other patients. Finally, a total of 90 patients (28 women and 62 men) and 175 controls (54 women and 121 men) were included in the present analysis.

Laboratory assay for isoflavones

From the blood samples collected at baseline, plasma concentrations of isoflavones (genistein, daidzein, glycitein, and equol) were assessed using triple quadrupole tandem liquid chromatography-mass spectrometry (7). All samples were analyzed at a single laboratory (SRL). Laboratory technicians performed the analyses under the mask of case–control status, and samples from matched sets were assayed together. The detection limit for all of the isoflavones was 1.0 ng/mL. For quality control, a pooled blood sample from healthy volunteers was used, and interassay and intraassay coefficients of variation were <6.2% and <3.0% for all isoflavones, respectively.

Statistical analysis

Using our previous findings (3), we performed sex-specific analyses. Comparisons of the baseline characteristics between the cases and controls were performed with the χ^2 or the Mann–Whitney test, as appropriate. In the controls, Spearman rank correlation coefficients were calculated for plasma concentrations and dietary intakes of genistein and diadzein. The dietary genistein and daidzein intake as assessed with the food-frequency questionnaire has been described in detail previously (3).

Using a conditional logistic regression model, we calculated ORs and 95% confidence intervals (CI) of primary liver cancer development for plasma genistein and diadzein concentrations divided into sex-specific tertiles according to the frequency of distribution among the controls. For glycitein and equol concentrations, three categories were defined: participants with concentrations below the detection limit, and lower half and upper half of those above the detection limit. The trend was tested by assigning ordinal values for categorical variables. In a multivariable model, we adjusted for the following variables previously associated with liver cancer risk (8): alcohol consumption (never, past, or regular for women, and never, past, <150, 150 to <450, or ≥450 g/week ethanol for men); body mass index (BMI; <25.0, ≥25.0 kg/m^2); diabetes defined as a self-reported history of diabetes, and/or antidiabetic medication use, and/or blood glucose ≥5.55 mmol/L (100 mg/dL) fasting or ≥7.77 mmol/L (140 mg/dL) nonfasting (yes, no); and coffee consumption (almost never, once a week to <1 cup/day, ≥1 cup/day). An additional model was further adjusted for serum alanine aminotransferase (ALT) levels (<30, 30–69, ≥70 IU/L; ref. 9). We also entered the following variables in the model: smoking status (a suspected risk factor for liver cancer), vegetable intake, fish intake, and plasma concentrations of total adiponectin (factors associated with liver cancer risk in the JPHC Study; refs. 10–12). However, the inclusion of these factors did not change the risk estimates substantially.

Subgroup analyses were performed for 79 patients with cancer with HCV infection (26 women and 53 men) and 27 female patients after menopause at baseline. To examine the effect modification of exposure to isoflavones by BMI (<25.0, ≥25.0 kg/m^2) and diabetes (yes or no), factors associated with both isoflavones (13) and liver cancer (8), we conducted stratified analyses using an unconditional logistic regression model adjusted for matching factors and variables in the multivariable model. In addition, stratified analyses of equol producers (defined as participants with equol concentrations above the detection limit of ≥ 1.0 ng/mL) were performed, because the beneficial health effects of isoflavones were likely to differ between equol producers with specific intestinal bacteria and nonproducers (14). In these stratified analyses, we dichotomized participants into low and

Published OnlineFirst December 26, 2014; DOI: 10.1158/1055-9965.EPI-14-1118

Table 1. Selected baseline characteristics of cases and controls

Variables	Women			Men		
	Cases ($n = 28$) Prevalence (%)	Controls ($n = 54$) Prevalence (%)	P value[a]	Cases ($n = 62$) Prevalence (%)	Controls ($n = 121$) Prevalence (%)	P value[a]
Age, y						
40–49	3.6	3.7	Matching variable	1.6	3.3	Matching variable
50–59	25.0	27.8		29.0	28.9	
60–69	71.4	68.5		69.4	67.8	
Hepatitis virus infectious status[b]						
HBV	7.1	5.6	Matching variable	14.5	14.0	Matching variable
HCV	92.9	94.4		85.5	86.0	
Menopausal status, premenopausal	3.6	3.7	Matching variable			
Alcohol consumption, regular drinker	10.7	25.9	0.14	50.0	66.9	0.08
Smoking status, current smoker	14.3	7.4	0.08	48.4	47.9	0.81
BMI, $\geq$25 kg/m^2	46.4	22.2	0.02	33.9	14.9	<0.01
Diabetes, yes[c]	17.9	11.1	0.40	40.3	27.3	0.07
Coffee consumption, $\geq$1 cup/day	25.0	35.2	0.35	22.6	40.5	0.02
ALT level, $\geq$70 IU/L	53.6	4.0	< 0.01	42.4	7.0	<0.01
Vegetable intake (g/day)[d]	43.3 (32.5–66.0)	49.1 (30.9–71.5)	0.65	48.1 (25.3–75.8)	48.7 (30.5–75.3)	0.48
Fish intake (g/day)[d]	37.7 (22.8–57.4)	38.1 (21.0–53.5)	0.93	58.9 (37.5–76.0)	52.5 (32.7–73.2)	0.40
Dietary intake of genistein (mg/day)[d]	14.0 (10.5–20.5)	10.3 (6.4–18.7)	0.01	11.9 (6.6–21.2)	13.6 (8.2–20.4)	0.63
Dietary intake of daidzein (mg/day)[d]	8.4 (6.3–12.3)	6.1 (3.8–10.0)	0.01	7.1 (3.9–12.9)	8.1 (4.9–12.2)	0.63

[a]Calculated using the χ^2 test and the Mann–Whitney test.
[b]Positive for hepatitis B surface antigen was regarded as indicating HBV infection and positive for anti-hepatitis C virus antibody as indicating HCV infection.
[c]Diabetes was defined as a self-reported history of diabetes, and/or antidiabetic medication use, and/or blood glucose $\geq$5.55 mmol/L (100 mg/dL) fasting or $\geq$7.77 mmol/L (140 mg/dL) nonfasting.
[d]Energy-adjusted by using the residual method, median (interquartile range).

high groups, based on the median concentrations of genistein and daidzein in the controls. For glycitein and equol, participants were categorized into low (not detected) and high (detected) groups. We used a likelihood ratio test to examine the potential effect modifications according to the stratified variables.

All analyses were performed with STATA version 11 (STATA Corporation, College Station). All P values reported are two sided, and differences at $P < 0.05$ were considered significant.

Results

The baseline characteristics of the case and control groups are shown in Table 1. Among the women, the proportions of overweight and high ALT levels and dietary intakes of genistein and daidzein were higher in the case group than in the control group. Among the men, there were statistically different distributions of BMI, coffee consumption, and ALT levels between the case and control groups. The groups showed no differences in median plasma concentrations of isoflavones, except for equol in women: median concentrations of equol tended to be lower in the cases than in the controls (0 vs. 2.8 ng/mL, $P = 0.04$; Table 2). Spearman rank correlation coefficients between plasma concentrations and dietary intake of genistein were 0.12 for women and 0.27 for men, and those of diadzein were 0.09 for women and 0.29 for men.

We found no consistent association in either sex between plasma isoflavone concentrations and primary liver cancer risk (Table 3). The multivariable ORs of primary liver cancer for the high versus low tertiles of genistein, daidzein, glycitein, and equol were 1.31 (95% CI, 0.28–6.05), 0.55 (0.10–3.19), 1.97 (0.35–10.93), and 0.44 (0.13–1.49), respectively, for women; for men, they were 1.33 (0.58–3.09), 1.23 (0.47–3.21), 2.14 (0.82–5.59), and 1.35 (0.60–3.04). Further adjustment for ALT levels did not change the tendency of the results (data not shown). No material change was seen when the premenopausal women were excluded: multivariable ORs for the high versus low group are 1.12 (95% CI, 0.24–5.35) for genistein, 0.50 (0.09–2.89) for diadzein, 1.79 (0.32–10.14) for glycitein, and 0.37 (0.10–1.34) for equol. We also observed a null association when we restricted analyses to participants with HCV infection (data not shown). In addition, there was no statistical evidence of any effect modification across the strata of BMI, diabetes, and equol producers (Supplementary Table S1).

Discussion

Our previous cohort analysis showed that dietary intake of isoflavones increased the risk of primary liver cancer in women (3). Multivariable HRs for the high versus low tertiles of genistein intake and daidzein intake were 3.19 (95% CI, 1.13–9.00, P for trend = 0.03) and 3.90 (1.30–11.69, P for trend = 0.01), respectively. Even when analysis was restricted to participants infected with hepatitis virus (25 cases of liver cancer), this positive

Table 2. Plasma concentrations of isoflavones in cases and controls

Isoflavones	Women			Men		
	Cases ($n = 28$) Median (interquartile range)	Controls ($n = 54$) Median (interquartile range)	P value[a]	Cases ($n = 62$) Median (interquartile range)	Controls ($n = 121$) Median (interquartile range)	P value[a]
Genistein (ng/mL)	46.5 (16.2–111.4)	44.6 (17.1–164.2)	0.71	94.8 (30.2–201.9)	64.7 (31.3–162.5)	0.25
Diadzein (ng/mL)	18.8 (3.3–43.9)	21.7 (5.3–65.0)	0.30	31.8 (7.3–81.2)	27.4 (10.0–66.8)	0.76
Glycitein[b] (ng/mL)	1.5 (0–3.5)	1.1 (0–3.5)	0.77	2.4 (0–7.8)	2.1 (0–5.4)	0.22
Equol[b] (ng/mL)	0 (0–3.3)	2.8 (0–14.6)	0.04	7.1 (0–26.9)	3.7 (0–16.7)	0.25

[a]Calculated using the Mann–Whitney test.
[b]Values below the detection limit (< 1.0 ng/mL) were regarded as zero.

Published OnlineFirst December 26, 2014; DOI: 10.1158/1055-9965.EPI-14-1118

Table 3. ORs and 95% CIs of primary liver cancer according to plasma concentrations of isoflavones

	Women				Men			
Plasma concentration	**Low**	**Middle**	**High**	***P* for trend[a]**	**Low**	**Middle**	**High**	***P* for trend[a]**
Genistein (ng/mL)	<28.0	28.0–100.9	>100.9		<41.9	41.9–111.0	>111.0	
Number of cases/controls	11/18	9/18	8/18		21/41	14/40	27/40	
Matching variables adjusted OR (95% CI)[b]	1.00 (reference)	0.79 (0.27–2.33)	0.65 (0.18–2.29)	0.50	1.00 (reference)	0.70 (0.31–1.56)	1.36 (0.64–2.87)	0.41
Multivariable OR (95% CI)[c]	1.00 (reference)	1.41 (0.37–5.41)	1.31 (0.28–6.05)	0.75	1.00 (reference)	0.61 (0.25–1.62)	1.33 (0.58–3.09)	0.47
Daidzein (ng/mL)	<12.0	12.0–62.8	>62.8		<13.1	13.1–47.8	>47.8	
Number of cases/controls	12/19	11/19	5/16		22/43	15/38	25/40	
Matching variables adjusted OR (95% CI)[b]	1.00 (reference)	0.83 (0.28–2.43)	0.33 (0.07–1.58)	0.19	1.00 (reference)	0.75 (0.32–1.77)	1.32 (0.57–3.03)	0.53
Multivariable OR (95% CI)[c]	1.00 (reference)	1.21 (0.29–5.13)	0.55 (0.10–3.19)	0.50	1.00 (reference)	0.68 (0.25–1.82)	1.23 (0.47–3.21)	0.70
Glycitein (ng/mL)	Not detected	1.0–3.4	>3.4		Not detected	1.0–3.2	>3.2	
Number of cases/controls	12/26	9/14	7/14		16/39	17/42	29/40	
Matching variables adjusted OR (95% CI)[b]	1.00 (reference)	1.43 (0.47–4.31)	1.09 (0.26–4.50)	0.81	1.00 (reference)	1.02 (0.44–2.36)	1.95 (0.84–4.53)	0.10
Multivariable OR (95% CI)[c]	1.00 (reference)	2.96 (0.60–14.61)	1.97 (0.35–10.93)	0.48	1.00 (reference)	1.05 (0.41–2.68)	2.14 (0.82–5.59)	0.11
Equol (ng/mL)	Not detected	1.0–12.3	>12.3		Not detected	1.0–11.2	>11.2	
Number of cases/controls	18/20	4/16	6/18		22/46	13/35	27/40	
Matching variables adjusted OR (95% CI)[b]	1.00 (reference)	0.33 (0.10–1.14)	0.40 (0.13–1.25)	0.06	1.00 (reference)	0.76 (0.33–1.79)	1.44 (0.70–2.94)	0.33
Multivariable OR (95% CI)[c]	1.00 (reference)	0.26 (0.06–1.17)	0.44 (0.13–1.49)	0.11	1.00 (reference)	0.74 (0.27–2.00)	1.35 (0.60–3.04)	0.47

[a]Linear trends were tested using the exposure categories as ordinal variables.

[b]Matching variables were age, sex, public health center area, fasting status at blood collection, hepatitis virus infectious status, and baseline menopausal status (for women).

[c]Adjusted for alcohol consumption (never, past, or regular drinker for women; never, past, <150, 150 to <450, ≥450 g/week ethanol for men), BMI (<25.0, ≥25.0 kg/m^2), diabetes (yes, no), and coffee consumption (almost never, once a week to <1 cup/day, ≥1 cup/day).

association remained essentially unchanged. We thought, therefore, that plasma concentrations of isoflavones would tend to be positively associated with the occurrence of primary liver cancer in women with hepatitis virus infection, although we suspected sufficient statistical power might not be obtained due to the relatively small sample size. In the present study, however, we observed no apparent association.

One possible explanation for this inconsistency is that plasma concentrations of isoflavones might not actually reflect dietary intake of isoflavones among participants with hepatitis virus infection. The half-lives of genistein and daidzein in the blood are reported to be 8.4 hours and 5.8 hours, respectively (15). Plasma concentrations of isoflavones reflect the intake of isoflavones in a dose-dependent manner (16, 17), and these concentrations are known to depend on the time elapsed since the last meal. Therefore, we matched fasting times in the cases and controls to minimize any exposure misclassification caused by differences in fasting times. However, plasma isoflavone concentrations are markers of short-term isoflavone exposure, whereas the results of our food-frequency questionnaire on dietary intake of isoflavones reflect personal dietary habits over long periods of time. Short-term exposure does not necessarily correlate with long-term exposure. Even so, in the general population of our validation study, isoflavone concentrations in the blood correlated reasonably well with isoflavone intake as estimated from the questionnaire (Spearman correlation coefficient = 0.33 for genistein and 0.31 for daidzein in the combined data on both sexes; ref. 18): plasma isoflavone concentrations seemed to be maintained in Japanese who like isoflavone-rich foods. Our earlier work within the JPHC Study on the associations between isoflavones and cancers in other sites supports this assumption, with similar associations observed between the results of a cohort study using a food-frequency questionnaire and those of a nested case–control study using plasma concentrations (19–26). Therefore, the present results indicate a possibility that plasma concentrations of isoflavones do not reflect dietary intake of isoflavones in people infected with hepatitis virus.

The correlation coefficients for plasma concentrations and dietary intake (as estimated from the food-frequency questionnaire) of genistein and daidzein in the present study population infected with hepatitis virus (Spearman correlation coefficient for genistein = 0.12 in women and 0.27 in men, and for diadzein = 0.09 in women and 0.29 in men) tended to be lower than those in the general population of our validation study (18). Although these findings might be the result of chance, another possible explanation lies in the metabolism of isoflavones: isoflavones are absorbed in the upper small intestine and conjugated in the liver (27); conjugated metabolites are excreted in the bile, are deconjugated in the lower bowel, and are absorbed again (27), meaning that an enterohepatic circulation is formed. In the present study, we targeted people infected with hepatitis virus. We hypothesize, therefore, that the conjugation metabolism of isoflavones in the liver is delayed when liver function is impaired by virus-related liver disease, decreasing the enterohepatic circulation volume of isoflavones in patients with virus-related liver disease, and thereby reducing isoflavone concentrations in the blood. We also hypothesize that the metabolism of isoflavones gradually declines as virus-related liver diseases progress. If this hypothesis is correct, we might observe a positive association between plasma isoflavone concentrations and liver cancer in subgroup analyses excluding participants with severe hepatitis and liver cirrhosis, or after

Published OnlineFirst December 26, 2014; DOI: 10.1158/1055-9965.EPI-14-1118

adjustment for liver disease stage. However, a limitation of this study is that we had no information on the clinical severity of liver disease related to HBV or HCV infection. Because of the relatively small number of cases of liver cancer, it was also difficult to find any meaningful association after excluding cases diagnosed in the first several years of follow-up. Careful consideration to this hypothesis leads us to believe that the null association we observed in women might be partially explained by the variance between plasma isoflavone concentrations and dietary isoflavone intake caused by changes in the metabolism of isoflavones related to liver disease with hepatitis virus. Further epidemiologic and experimental investigations are needed to examine the association between biomarkers of isoflavones at the early stage of virus-related liver disease and liver cancer.

As in our previous study based on dietary isoflavone intake (3), we found no association between plasma isoflavone concentrations and primary liver cancer in men. Testosterone is reportedly associated with the risk of hepatocellular carcinoma (the most common form of primary liver cancer; refs. 28, 29). However, the earlier studies did not address concerns about the association between soy isoflavones and the male sex hormone (30). We found no evidence that isoflavones play any role in the etiology of liver cancer in men, and we believe that even if they do have a role, it is small.

The major strength of our study is that it is, to our knowledge, the first prospective study to evaluate the association between plasma isoflavones and primary liver cancer. We used objective measures that reflected dietary intake of isoflavones and individual differences in absorption and metabolism (4), and attempted to elucidate the influences of exposure to isoflavones in the liver carcinogenesis. In addition, using blood samples for exposure assessment made it possible to examine the role of equol, which cannot be assessed from food-frequency questionnaires. Another strength is its nested case–control design. Blood samples were collected before cancer diagnosis, and the cases and controls were selected from the same population participating in the JPHC Study. ORs estimated in the nested case–control design represent a better approximation of risk ratios (31), allowing our study to overcome the disadvantages inherent in the case–control design. However, caution is necessary in generalizing the results, because our participants were limited to those who answered the questionnaire and provided blood samples (32). Also, primary liver cancer that was unrelated to HBV or HCV infection was not considered in this study.

In conclusion, we found no apparent association between plasma concentrations of isoflavones and the risk of primary liver cancer in participants of either sex with hepatitis virus infection. To clarify the role of isoflavones in the etiology of liver cancer, further studies using both biomarkers and data on dietary intake of isoflavones are required.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: T. Michikawa, M. Inoue, T. Yamaji, M. Iwasaki, M. Mizokami, S. Tsugane

Development of methodology: T. Michikawa, M. Inoue, Y. Tanaka

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): T. Michikawa, M. Inoue, N. Sawada, Y. Tanaka, T. Yamaji, M. Iwasaki, T. Shimazu, M. Mizokami, S. Tsugane

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): T. Michikawa, M. Inoue, N. Sawada, T. Yamaji, M. Iwasaki, S. Sasazuki, S. Tsugane

Writing, review, and/or revision of the manuscript: T. Michikawa, M. Inoue, T. Yamaji, M. Iwasaki, T. Shimazu, S. Sasazuki, M. Mizokami, S. Tsugane

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Inoue, S. Tsugane

Study supervision: M. Inoue, M. Mizokami, S. Tsugane

Acknowledgments

The authors thank all staff members in each study area and in the central offices for their cooperation and technical assistance. They also thank the Aomori, Iwate, Ibaraki, Niigata, Osaka, Kochi, Nagasaki, and Okinawa Cancer Registries for providing incidence data.

Members of the JPHC Study (principal investigator: S. Tsugane) group are as follows: S. Tsugane, N. Sawada, S. Sasazuki, M. Iwasaki, T. Shimazu, T. Yamaji, and T. Hanaoka, National Cancer Center, Tokyo; J. Ogata, S. Baba, T. Mannami, A. Okayama, and Y. Kokubo, National Cerebral and Cardiovascular Center, Osaka; K. Miyakawa, F. Saito, A. Koizumi, Y. Sano, I. Hashimoto, T. Ikuta, Y. Tanaba, H. Sato, Y. Roppongi, and T. Takashima, Iwate Prefectural Ninohe Public Health Center, Iwate; Y. Miyajima, N. Suzuki, S. Nagasawa, Y. Furusugi, N. Nagai, Y. Ito, S. Komatsu, and T. Minamizono, Akita Prefectural Yokote Public Health Center, Akita; H. Sanada, Y. Hatayama, F. Kobayashi, H. Uchino, Y. Shirai, T. Kondo, R. Sasaki, Y. Watanabe, Y. Miyagawa, Y. Kobayashi, M. Machida, K. Kobayashi, and M. Tsukada, Nagano Prefectural Saku Public Health Center, Nagano; Y. Kishimoto, E. Takara, T. Fukuyama, M. Kinjo, M. Irei, and H. Sakiyama, Okinawa Prefectural Chubu Public Health Center, Okinawa; K. Imoto, H. Yazawa, T. Seo, A. Seiko, F. Ito, F. Shoji, and R. Saito, Katsushika Public Health Center, Tokyo; A. Murata, K. Minato, K. Motegi, T. Fujieda, and S. Yamato, Ibaraki Prefectural Mito Public Health Center, Ibaraki; K. Matsui, T. Abe, M. Katagiri, M. Suzuki, and K. Matsui, Niigata Prefectural Kashiwazaki and Nagaoka Public Health Center, Niigata; M. Doi, A. Terao, Y. Ishikawa, and T. Tagami, Kochi Prefectural Chuo-higashi Public Health Center, Kochi; H. Sueta, H. Doi, M. Urata, N. Okamoto, F. Ide, and H. Goto, Nagasaki Prefectural Kamigoto Public Health Center, Nagasaki; H. Sakiyama, N. Onga, H. Takaesu, M. Uehara, T. Nakasone, and M. Yamakawa, Okinawa Prefectural Miyako Public Health Center, Okinawa; F. Horii, I. Asano, H. Yamaguchi, K. Aoki, S. Maruyama, M. Ichii, and M. Takano, Osaka Prefectural Suita Public Health Center, Osaka; Y. Tsubono, Tohoku University, Miyagi; K. Suzuki, Research Institute for Brain and Blood Vessels Akita, Akita; Y. Honda, K. Yamagishi, S. Sakurai, and N. Tsuchiya, University of Tsukuba, Ibaraki; M. Kabuto, National Institute for Environmental Studies, Ibaraki; M. Yamaguchi, Y. Matsumura, S. Sasaki, and S. Watanabe, National Institute of Health and Nutrition, Tokyo; M. Akabane, Tokyo University of Agriculture, Tokyo; T. Kadowaki and M. Inoue, The University of Tokyo, Tokyo; M. Noda and T. Mizoue, National Center for Global Health and Medicine, Tokyo; Y. Kawaguchi, Tokyo Medical and Dental University, Tokyo; Y. Takashima and Y. Yoshida, Kyorin University, Tokyo; K. Nakamura and R. Takachi, Niigata University, Niigata; J. Ishihara, Sagami Women's University, Kanagawa; S. Matsushima and S. Natsukawa, Saku General Hospital, Nagano; H. Shimizu, Sakihae Institute, Gifu; H. Sugimura, Hamamatsu University School of Medicine, Shizuoka; S. Tominaga, Aichi Cancer Center, Aichi; N. Hamajima, Nagoya University, Aichi; H. Iso and T. Sobue, Osaka University, Osaka; M. Iida, W. Ajiki, and A. Ioka, Osaka Medical Center for Cancer and Cardiovascular Disease, Osaka; S. Sato, Chiba Prefectural Institute of Public Health, Chiba; E. Maruyama, Kobe University, Hyogo; M. Konishi, K. Okada, and I. Saito, Ehime University, Ehime; N. Yasuda, Kochi University, Kochi; S. Kono, Kyushu University, Fukuoka; S. Akiba, Kagoshima University, Kagoshima.

Grant Support

This study was supported by National Cancer Center Research and Development Fund (to M. Inoue, N. Sawada, T. Yamaji, M. Iwasaki, T. Shimazu, S. Sasazuki, and S. Tsugane), a Grant-in-Aid for Cancer Research (to M. Inoue and S. Tsugane), and by a Health and Labour Sciences Research Grant—Research on Hepatitis from the Ministry of Health, Labour, and Welfare, Japan (to M. Inoue).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 30, 2014; revised December 2, 2014; accepted December 5, 2014; published OnlineFirst December 26, 2014.