

RESEARCH

Open Access



Weighted gene co-expression network analysis to explain the relationship between plasma total carotenoids and lipid profile

Bénédicte L. Tremblay^{1,2}, Frédéric Guénard^{1,2}, Benoît Lamarche^{1,2}, Louis Pérusse^{1,3} and Marie-Claude Vohl^{1,2*}

Abstract

Background: Variability in circulating carotenoids may be attributable to several factors including, among others, genetic variants and lipid profile. However, relatively few studies have considered the impact of gene expression in the inter-individual variability in circulating carotenoids. Most studies considered expression of genes individually and ignored their high degree of interconnection. Weighted gene co-expression network analysis (WGCNA) is a systems biology method used for finding gene clusters with highly correlated expression levels and for relating them to phenotypic traits. The objective of the present observational study is to examine the relationship between plasma total carotenoid concentrations and lipid profile using WGCNA.

Results: Whole blood expression levels of 533 probes were associated with plasma total carotenoids. Among the four WGCNA distinct modules identified, turquoise, blue, and brown modules correlated with plasma high-density lipoprotein cholesterol (HDL-C) and total cholesterol. Probes showing a strong association with HDL-C and total cholesterol were also the most important elements of the brown and blue modules. A total of four and 29 hub genes associated with total carotenoids were potentially related to HDL-C and total cholesterol, respectively.

Conclusions: Expression levels of 533 probes were associated with plasma total carotenoid concentrations. Using WGCNA, four modules and several hub genes related to lipid and carotenoid metabolism were identified. This integrative analysis provides evidence for the potential role of gene co-expression in the relationship between carotenoids and lipid concentrations. Further studies and validation of the hub genes are needed.

Keywords: Carotenoids, French Canadians, Hub genes, Lipid profile, WGCNA

Background

Carotenoids are a reliable biomarker of fruit and vegetable consumption [1, 2]. Over 90% of the daily carotenoid intakes are provided by fruits and vegetables [3]. Carotenoids are a family composed of more than 700 fat-soluble pigments, but α -carotene, β -carotene, β -cryptoxanthin, lutein, lycopene, and zeaxanthin represent over 95% of total circulating carotenoids in human plasma or serum [3, 4]. Inter-individual variability in circulating carotenoids has

been observed and may be attributable to several factors including age, sex, body weight, physical activity, genetic, and lipid profile [2]. Accordingly, lower total cholesterol (TC), LDL-cholesterol (LDL-C), and HDL-cholesterol (HDL-C) concentrations were associated with lower circulating carotenoids [5]. Plasma HDL-C concentrations also mediated the observed sex difference in serum carotenoids in Caucasian individuals [5]. Correlations were observed between α - and β -carotene and HDL-C and triglyceride (TG) concentrations [6] and between β -cryptoxanthin and zeaxanthin and TG in a previous study by our group [7]. The NHANES 2003–2006 study also reported that several serum carotenoids positively correlated with HDL-C concentrations [8]. Thus, the link

* Correspondence: marie-claude.vohl@fsaa.ulaval.ca

¹Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada

²School of Nutrition, Laval University, 2425 rue de l'Agriculture, Quebec City, QC G1V 0A6, Canada

Full list of author information is available at the end of the article



between circulating carotenoids and plasma lipid profile has been observed in several studies and is plausible considering that the majority of circulating carotenoids are associated with lipoproteins [9].

Carotenoid intakes are inversely associated with the risk of cardiovascular diseases and certain cancers [10]. They may mediate their effects mainly via antioxidant properties but also via other mechanisms such as gap junction communication, cell growth regulation, immune response, and modulation of gene expression [11, 12]. Several genome-wide association studies have identified genetic variants that influence circulating carotenoid concentrations [13–16]. Genetic variants may cause differences in the absorption, assimilation, distribution, metabolism, and excretion of carotenoids [13, 17, 18]. Carotenoids have also been shown to regulate the expression of genes protective against carcinogenesis and inflammation [19]. Carotenoids and their derivatives (e.g., retinoid) may exert their effect on gene expression via several transcriptional systems such as the retinoid receptors, the nuclear factor-kappa B, and the peroxisome proliferator-activated receptors [20]. However, relatively few studies have considered the impact of gene expression in the inter-individual variability in circulating carotenoid concentrations. Gene expression may represent a potential mechanism that links carotenoids to lipid profile. In addition, no previous study has examined the interconnection between plasma carotenoids, lipid profile, and genome-wide gene expression levels. Most studies considered expression of genes individually and ignored the high degree of interconnection between genes. Weighted gene co-expression network analysis (WGCNA) is a widely used systems biology approach used to identify gene clusters (modules) with highly correlated expression levels, to relate the modules to phenotypic traits, and to identify key hub genes within modules that are related to the phenotypic traits [21].

Thus, the objective of the present observational study is to examine the relationship between plasma total carotenoid concentrations and lipid profile using WGCNA. The hypothesis was that genome-wide gene expression levels are associated with plasma total carotenoids and that clusters of genes associated with carotenoids are correlated to lipid profile. To test this hypothesis, genes associated with total carotenoids were identified using regressions and WGCNA was used to identify specific modules and key hub genes related to lipid profile.

Results

Characteristics of study participants

Characteristics and biochemical parameters of participants are presented in Table 1. Fathers and mothers had significant differences in HDL-C concentrations. Concentrations of all six main carotenoids (α -carotene, β -carotene, β -cryptoxanthin, lutein, lycopene, and zeaxanthin) and total carotenoids measured in the fasting state are presented in Additional file 1: Table S1.

Associations between total carotenoids and genome-wide gene expression levels

Normalized gene expression levels of 18,160 probes were tested for associations with plasma total carotenoid concentrations. A total of 533 probes were associated ($P \leq 0.05$) with total carotenoids (Additional file 2: Table S2). WGCNA was then conducted on this subset of probes, which were associated with total carotenoids in order to put them in relationship with lipid profile.

Weighted gene co-expression network analysis (WGCNA)

A total of four distinct modules were identified from gene expression levels of the 533 probes using a dynamic tree cutting algorithm (Fig. 1). The blue, brown, turquoise, and gray modules contained 117, 64, 184, and 168 genes,

Table 1 Characteristic and biochemical parameters of study subjects

Biochemical parameters	Fathers ($n = 6$)	Mothers ($n = 16$)	Boys ($n = 18$)	Girls ($n = 8$)
Age (years)	42.0 \pm 2.8	42.4 \pm 6.0	11.9 \pm 3.8	10.1 \pm 2.0
BMI (kg/m ²)	24.8 \pm 1.3	23.5 \pm 3.4	–	–
BMI percentile	–	–	48.9 \pm 32.7	52.1 \pm 29.8
TC (mmol/L)	4.70 \pm 0.61	4.67 \pm 0.55	4.35 \pm 0.53	4.15 \pm 0.48
HDL-C (mmol/L) ¹	1.36 \pm 0.33	1.73 \pm 0.35	1.56 \pm 0.28	1.54 \pm 0.31
LDL-C (mmol/L)	2.83 \pm 0.67	2.53 \pm 0.49	2.36 \pm 0.42	2.20 \pm 0.52
ApoB100 (g/L)	0.89 \pm 0.22	0.77 \pm 0.11	0.72 \pm 0.12	0.67 \pm 0.15
TG (mmol/L)	1.13 \pm 0.35	0.88 \pm 0.34	0.94 \pm 0.39	0.89 \pm 0.41
Total carotenoids (μ mol/L)	5.77 \pm 2.93	6.57 \pm 2.23	5.88 \pm 2.37	5.30 \pm 1.09

All values are means \pm SD

ApoB100 apolipoprotein B-100, BMI body mass index, HDL-C high-density lipoprotein cholesterol, LDL-C low-density lipoprotein cholesterol, SD standard deviation, TC total cholesterol, TG triglycerides

¹Means are significantly different ($p \leq 0.05$) between fathers and mothers

respectively. The 168 uncorrelated genes were assigned to the gray module, which was excluded from further analysis. Using a cutoff height of 0.25 for the clustering of module eigengenes (MEs), none of the four modules merged. Thus, the merged dynamic yielded same modules as the dynamic tree cut (Fig. 1). To find modules of interest, correlations between MEs and lipid profile traits (TC, LDL, HDL, TG, apolipoprotein B-100 (ApoB100)) were computed. According to the heatmap of module-trait correlations (Fig. 2), ME of the turquoise module (184 genes) correlated inversely with HDL-C ($r = -0.32$, $p = 0.03$) and TC ($r = -0.35$, $p = 0.01$), while ME of the blue module (117 genes) correlated inversely with HDL-C ($r = -0.36$, $p = 0.01$) and TC ($r = -0.41$, $p = 0.004$), and ME of the brown module (64 genes) correlated positively with HDL-C ($r = 0.31$, $p = 0.03$).

Gene significance (GS), defined as the correlation between gene expression and lipid profile traits, was put in relation with module membership (MM), defined as the correlation of the ME and the gene expression profile. GS for HDL-C correlated with MM in the brown module ($r = 0.29$, $p = 0.02$), while GS for TC correlated with MM in the blue module ($r = 0.42$, $p = 2.4 \times 10^{-6}$) (Fig. 3). This suggests that genes highly significantly associated with HDL-C and TC were also the most important elements of the brown and blue modules, respectively.

Thus, brown and blue modules were selected as modules of interest in subsequent analyses.

Enrichment analysis

Enrichment analysis was performed for genes in brown and blue modules in order to elucidate biological mechanisms. In the GO database, a total of 73 GO terms from all three main categories (biological process, cellular component, and molecular function) were significantly enriched (corrected $p \leq 0.05$) in the brown module (Additional file 3: Table S3). In the GO database, a total of 253 GO terms also from all three main categories were significantly enriched (corrected $p \leq 0.05$) in the blue module (Additional file 4: Table S4). In the Reactome database, the most significant terms were used to better define modules of interest. Thus, the function representing the brown module was “apoptosis,” while the function representing the blue module was “mRNA splicing” (Additional file 3: Table S3 and Additional file 4: Table S4). Moreover, transcription factor (TF) enrichment analysis was conducted to identify TFs that are enriched in genes of the brown and blue modules. In the brown module, *NFE2L2* was the more significantly enriched TF ($p = 0.016$) with target genes *MAF* and *MAFF*. In the blue module, *ESR1* was the most significantly enriched TF ($p = 0.0047$) with target genes *NOP56*, *ATF1*, *LDHA*,

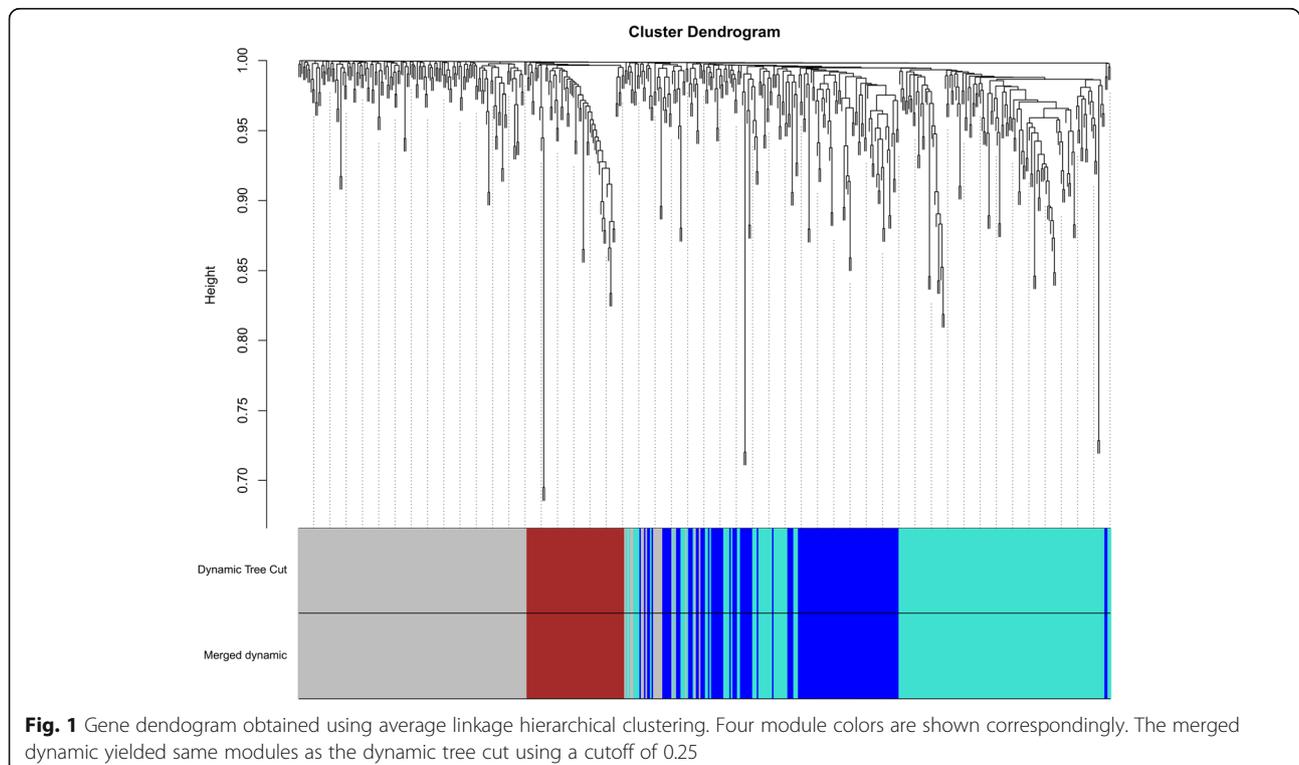
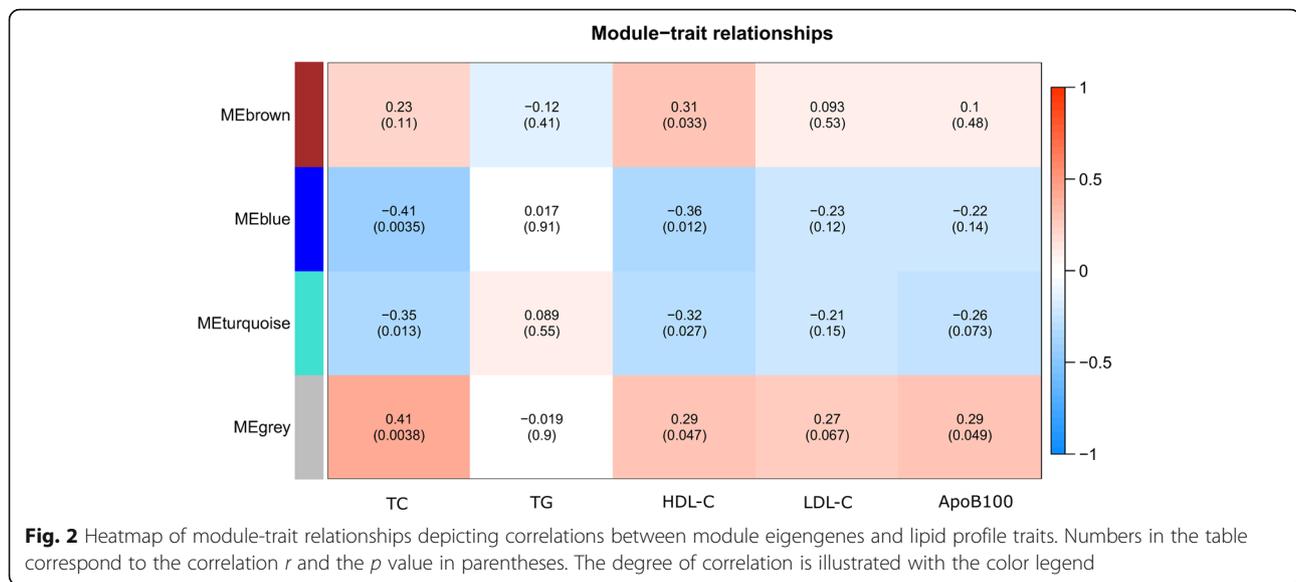


Fig. 1 Gene dendrogram obtained using average linkage hierarchical clustering. Four module colors are shown correspondingly. The merged dynamic yielded same modules as the dynamic tree cut using a cutoff of 0.25



SMARCB1, *HNRNPA2B1*, *HNRNPR*, *SLC3A2*, *DNM1L*, *CCT7*, *HNRNPA1*, *DBN1*, and *MARK2*.

Hub gene analysis

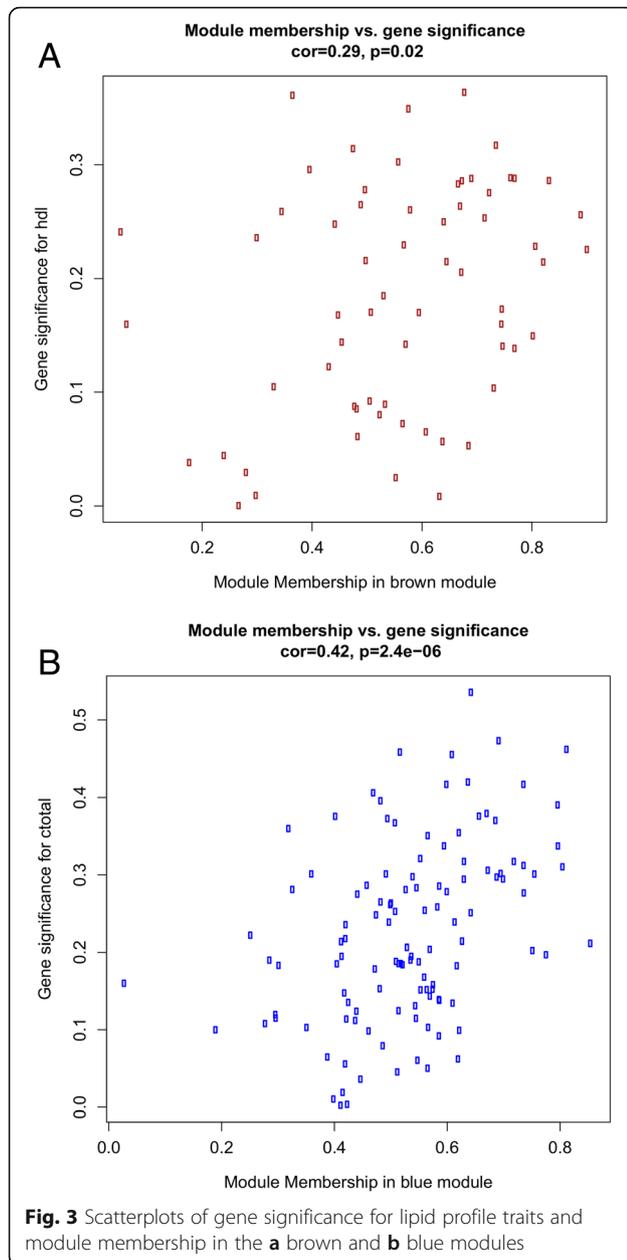
Hub gene analysis in brown and blue modules was conducted in order to refine the analysis of potential mechanisms. A total of seven transcripts (four hub genes: *CISH*, *FAM123C*, *ZSWIM4*, *FAM13A*) were identified in the brown module (Table 2). In the blue module, 32 transcripts (29 hub genes: *FEZ1*, *MYBPC3*, *OPRL1*, *SIRPB1*, *SLC16A3*, *SLC4A1*, *GPR89C*, *PSMD6*, *RANBP1*, *AP3M1*, *LOC221710*, *SIRT5*, *DNM1L*, *CR2*, *LBH*, *HNRNPA1*, *BIVM*, *NOP56*, *LOC728026*, *FBXO5*, *SNRPN*, *CCT7*, *TIAL1*, *SLC25A4*, *PCNA*, *PNN*, *S100BBP*, *UBE1DC1*, and *ZCCHC11*) were identified (Table 3). Classification terms (unknown, likely regulated, and likely regulator) were added in the Tables 2 and 3 to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations.

Discussion

We first tested the association between whole blood genome-wide gene expression levels and plasma total carotenoid concentrations in order to identify probes influenced by carotenoids. A total of 533 probes were significantly associated with total carotenoids. To the best of our knowledge, this is the first study that considers the impact of plasma total carotenoids on genome-wide gene expression levels. Up to now, studies focused mainly on the effect of genetic variants on the inter-individual variability in β -carotene, lycopene, and lutein bioavailability in response to dietary interventions [22–24]. SNPs were located in genes mainly related to

fatty acids and cholesterol absorption including *SR-B1*, *CD36*, *NPC1L1*, and *ABCA1*. Since several proteins are involved in the absorption and metabolism of carotenoids, several papers have suggested that variations in the expression of the genes encoding for these proteins might also impact carotenoid bioavailability [25]. Moreover, carotenoids modulate the mechanisms of cell proliferation, growth factor signaling, and gap junction communication, and lead to changes in the expression of many proteins involved in these processes [20]. The effect of carotenoids on gene expression may result from direct interaction with ligand-activated nuclear receptors such as retinoid receptors, the nuclear factor-kappa B, and the peroxisome proliferator-activated receptors [20, 26, 27]. In accordance, two important TFs were enriched among genes associated with total carotenoids. *NFE2L2*, which encodes a TF regulating genes with antioxidant response elements, was enriched in the brown module. Interestingly, some carotenoids have been showed to protect against light-induced cell damage through *NFE2L2* activation [28]. *ESR1*, a ligand-activated TF, was enriched in the blue module. Genetic variations in this gene were associated with serum TC and LDL-C concentrations [29].

Second, WGCNA was used to identify modules and key genes related to lipid profile in the subset of 533 probes associated with carotenoids. MEs of the turquoise and blue modules had a significant correlation with HDL-C and TC, while ME of the brown module had a significant correlation with HDL-C. This suggests that highly co-expressed genes in these modules had potential interaction or consistent biological effects on HDL-C and TC. Only the brown and the blue modules showed a significant correlation between the GS for HDL-C and



TC, respectively, and the MM of the module. Functional enrichment analysis revealed several enrichments of categories related to cell metabolism and processes. However, functional enrichment analysis yields large amount of GO terms and Reactome terms non-specific to lipid profile. Thus, it was not precise enough to identify specific pathways or genes related to lipid or carotenoid metabolism.

The analysis with hub genes allowed refining the analysis of potential mechanisms linking carotenoids and lipid profile. A total of four hub genes were identified in the brown module: *FAM123C*, *ZSWIM4*, *FAM13A*, and *CISH*. *FAM13A* and *CISH* are of interest in the context

of lipid profile. *FAM13A* encodes for a family with sequence similarity 13 member A. *FAM13A* variants have been associated with HDL-C in individuals from various descents [30, 31]. *FAM13A* promoted fatty acid oxidation, possibly by interacting with an activating sirtuin 1 and increasing expression of *CPT1A* [32]. Interestingly, another study using WGCNA identified *FAM13A* as a hub gene related to hyperlipidemia [33]. *CISH* encodes for a cytokine-inducible SH2 containing protein. It controls the signaling of a variety of cytokines, in particular interleukin-2, and seems to be critical for T cell proliferation and survival in response to infection [34]. Despite the fact that expression of *CISH* was not inhibited in the presence of delipidated HDL lipoproteins in a cell study, the link is plausible since HDL specifically inhibits the production of pro-inflammatory cytokines, which is also controlled by *CISH* [35]. A total of 29 hub genes were identified in the blue module. Relevant genes related to lipid and carotenoid metabolism include *MYBPC3*, *OPRL1*, *SLC16A3*, *SIRT5*, *HNRNPA1*, *SLC25A4*, and *PCNA*. *MYBPC3* encodes for a myosin binding protein C, cardiac. Mutation in this gene (*MYBPC3*-Q1061X) results in hypertrophic cardiomyopathy and cardiac oxidative stress with elevated TG and branched-chain amino acid levels [36, 37]. *OPRL1* encodes for an opioid-related nociceptin receptor 1 involved in many biological functions including stress, inflammatory, and immune responses. A study by our group identified interaction effects between 29 SNPs, including rs2229205 in *OPRL1*, and total fat intake on LDL peak particle diameter [38]. *SLC16A3* encodes for a proton-linked monocarboxylate transporter designated solute carrier family 16 member 3. Monocarboxylate transporters are involved in the transport of short-chain fatty acids and may also be involved in the transport of cholesterol-lowering agents [39]. *SLC16A3* was significantly upregulated in 102 men receiving an antioxidant-rich diet compared to controls [40]. *SIRT5* encodes for sirtuin 5, which is involved in the regulation of mitochondrial metabolism, oncogenesis, and oxidative stress [41, 42]. Obese *Sirt5*^{-/-} mice showed increased serum cholesterol concentrations compared to *Sirt5*^{+/+} mice [43]. *HNRNPA1* encodes for a heterogeneous nuclear ribonucleoprotein A1, which has been shown to reduce HMGCR enzyme activity and increase LDL-C uptake [44]. *SLC25A4* encodes for a solute carrier family 25 member 4. A variation in this gene is associated with hypertrophic cardiomyopathy [45]. Linoleic acid also increased *ANTI* (*SLC25A1*) expression as a compensatory response to an increase in intracellular ROS [46]. Finally, *PCNA* encodes for a proliferating cell nuclear antigen, which has high expression in tumor tissues [47]. Interestingly, carotenoids present various suppressive abilities against *PCNA* expressions in cell proliferation [48]. Lycopene also decreased the positive

Table 2 Hub genes identified in the brown module

Probe ID	Gene (Chr, accession number)	GS HDL-C	<i>p</i> value GS	MM	<i>p</i> value MM	Potential causal relationships of the hub genes ¹
ID_4920064	<i>CISH</i> (Chr3, NM_145071)	0.29	0.049	0.67	1.44×10^{-7}	Likely regulator
ID_6480731	<i>AMER3</i> (Chr2, NM_001105193)	0.36	0.011	0.68	1.12×10^{-7}	Likely regulated
ID_3870603	<i>ZSWIM4</i> (Chr19, NM_023072)	0.29	0.047	0.69	5.12×10^{-8}	Unknown
ID_290093	<i>LOC647322</i> (Chr2, withdrawn)	0.32	0.028	0.74	2.39×10^{-9}	Unknown
ID_4540376	<i>FAM13A</i> (Chr4, NM_001015045)	0.29	0.047	0.76	2.88×10^{-10}	Likely regulator
ID_2260315	<i>HS.572752</i>	0.29	0.047	0.77	1.48×10^{-10}	Unknown
ID_3420706	<i>HS.566359</i>	0.29	0.049	0.83	2.03×10^{-13}	Unknown

Chr chromosome, GS gene significance, HDL-C high-density lipoprotein cholesterol, MM module membership

¹Classification terms have been added to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations

rate of PCNA and protein expressions of PCNA in lung tissue [49]. In summary, several hub genes were related to lipid metabolism, oxidative stress, or antioxidant action of carotenoids. Thus, the plausible link between carotenoids and lipid profile does not seem to be entirely due to the fact that carotenoids are transported by lipoproteins. Indeed, circulating carotenoids have been showed to impact TFs involved in lipid metabolism [50, 51]. Accordingly, in the present study, plasma carotenoids modulated expression of several genes related, among others, to lipid metabolism. Moreover, classification terms were used to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations. A total of seven transcripts were classified as “likely regulator” as they have been associated with or shown to influence lipid levels in the literature and thus represent potential regulators of lipid concentrations. For example, *FAM13A*, which is associated with total carotenoids, also demonstrated a regulatory effect on HDL-C via its genetic variations and effect on fatty acid oxidation [30–32]. A total of 20 transcripts were classified as “likely regulated” meaning that they were associated with both total carotenoids and lipid concentrations in the present study. Finally, a total of 12 transcripts were classified as “unknown” considering they showed only associations with total carotenoids.

The present study has strengths but also some limitations. The main strength results from the study of the combination of both genome-wide gene expression levels and total carotenoids, representing six predominant plasma carotenoids. To the best of our knowledge, this is the first study that considers the impact of plasma total carotenoids on genome-wide expression levels. Another strength is the use of WGCNA that adds important information about the effect of co-expression

network of genes, which is useful to detect biological pathways related to lipid profile. On the other hand, the study’s main limitation resides in the small sample size, which limits statistical power. However, the study of a founder population with relatively homogeneous genetics and shared environment is a new aspect in this field [52]. Nonetheless, this limits the generalization of results to other populations. Finally, our study did not account for diet, physical activity, smoking, and alcohol consumption of participants, all of which may affect circulating carotenoid concentrations [53, 54].

Conclusions

In conclusion, whole blood expression levels of 533 probes were associated with plasma total carotenoid concentrations. Using WGCNA, four modules were identified. A total of four and 29 hub genes associated with total carotenoids were potentially related to HDL-C and TC, respectively. This integrative analysis provides evidence for the potential role of gene co-expression in the relationship between carotenoids and lipid concentrations. Further studies and validation of the hub genes are needed. Finally, this article may also serve as an example of how to include a wide range of omics data in nutrition studies, using systems biology methods.

Methods

Patients and design

A total of 48 Caucasian French-Canadian subjects from 16 families were recruited in the greater Quebec City metropolitan area, in Canada, as part of the GENERATION Study. The GENERATION Study was designed to evaluate familial resemblances in omics (DNA methylation [55] and gene expression [56]) and metabolic (metabolites [57] and carotenoids [7]) profiles in healthy families and to test the impact of these profiles on

Table 3 Hub genes identified in the blue module

Probe ID	Gene (Chr, accession number)	GS TC	<i>p</i> value GS	MM	<i>p</i> value MM	Potential causal relationships of the hub genes ¹
ID_2350019	<i>FEZ1</i> (Chr11, NM_022549)	0.31	0.032	-0.81	4.55×10^{-12}	Unknown
ID_6940768	<i>MYBPC3</i> (Chr11, NM_000256)	0.34	0.019	-0.80	1.06×10^{-11}	Likely regulator
ID_1440754	<i>OPRL1</i> (Chr20, NM_000913)	0.30	0.037	-0.70	3.76×10^{-8}	Likely regulator
ID_5870341	<i>LOC642691</i> (Chr1, withdrawn)	0.30	0.040	-0.69	5.85×10^{-8}	Unknown
ID_3780017	<i>SIRPB1</i> (Chr20, NM_006065)	0.40	0.0049	-0.68	1.01×10^{-7}	Likely regulated
ID_110719	<i>SLC16A3</i> (Chr17, NM_004207)	0.30	0.037	-0.62	2.45×10^{-6}	Likely regulator
ID_2120152	<i>SLC4A1</i> (Chr17, NM_000342)	0.31	0.030	-0.60	6.55×10^{-6}	Likely regulated
ID_2630739	<i>HS.154948</i>	-0.42	0.0032	0.60	6.37×10^{-6}	Likely regulated
ID_6040072	<i>GPR89C</i> (Chr1, NM_001097616)	-0.46	0.0012	0.61	4.10×10^{-6}	Likely regulated
ID_5810070	<i>PSMD6</i> (Chr3, NM_014814)	-0.36	0.012	0.62	3.19×10^{-6}	Likely regulated
ID_7000735	<i>RANBP1</i> (Chr22, NM_002882)	-0.38	0.0072	0.62	2.63×10^{-6}	Likely regulated
ID_6200437	<i>AP3M1</i> (Chr10, NM_207012)	-0.47	0.00076	0.62	2.35×10^{-6}	Likely regulated
ID_4670703	<i>SMIM13</i> (Chr6, XM_927609)	-0.35	0.013	0.62	2.29×10^{-6}	Likely regulated
ID_1510270	<i>SIRT5</i> (Chr6, NM_012241)	-0.29	0.042	0.63	1.48×10^{-6}	Likely regulator
ID_3890632	<i>DNM1L</i> (Chr12, NM_012063)	-0.32	0.028	0.63	1.46×10^{-6}	Unknown
ID_4810292	<i>CR2</i> (Chr1, NM_001877)	-0.42	0.0030	0.64	1.03×10^{-6}	Likely regulated
ID_2810246	<i>LBH</i> (Chr2, NM_030915)	-0.46	0.0011	0.64	9.22×10^{-7}	Likely regulated
ID_10209	<i>HNRNPA1</i> (Chr12, NM_031157)	-0.53	8.78×10^{-5}	0.64	7.83×10^{-7}	Likely regulator
ID_1090367	<i>BIVM</i> (Chr13, NM_017693)	-0.38	0.0085	0.66	3.64×10^{-7}	Likely regulated
ID_5560465	<i>NOP56</i> (Chr20, NM_006392)	-0.38	0.0079	0.67	1.70×10^{-7}	Likely regulated
ID_1990154	<i>PTMAP12</i> (Chr9, XM_001126659)	-0.31	0.035	0.67	1.51×10^{-7}	Unknown
ID_4010097	<i>FBXO5</i> (Chr6, NM_012177)	-0.42	0.0027	0.68	1.37×10^{-7}	Likely regulated
ID_7100367	<i>HS.571502</i>	-0.39	0.0057	0.69	7.78×10^{-8}	Likely regulated
ID_4850133	<i>SNRPN</i> (Chr15, NM_022806)	-0.37	0.0096	0.69	6.75×10^{-8}	Likely regulated
ID_7160719	<i>CCT7</i> (Chr2, NM_006429)	-0.47	0.00069	0.69	4.71×10^{-8}	Likely regulated
ID_5560504	<i>TIAL1</i> (Chr10, NM_003252)	-0.29	0.042	0.70	2.85×10^{-8}	Unknown
ID_6840747	<i>SLC25A4</i> (Chr4, NM_001151)	-0.32	0.028	0.72	7.82×10^{-9}	Unknown

Table 3 Hub genes identified in the blue module (*Continued*)

Probe ID	Gene (Chr, accession number)	GS TC	<i>p</i> value GS	MM	<i>p</i> value MM	Potential causal relationships of the hub genes ¹
ID_6900079	<i>PCNA</i> (Chr20, NM_182649)	-0.42	0.0032	0.74	2.27×10^{-9}	Likely regulated
ID_7150132	<i>PNN</i> (Chr14, NM_002687)	-0.31	0.031	0.74	2.23×10^{-9}	Unknown
ID_3170189	<i>S100PBP</i> (Chr1, NM_022753)	-0.30	0.038	0.76	4.91×10^{-10}	Unknown
ID_110059	<i>UBA5</i> (Chr3, NM_024818)	-0.39	0.0061	0.80	1.12×10^{-11}	Likely regulated
ID_4260093	<i>TUT4</i> (Chr1, NM_001009882)	-0.46	0.00096	0.81	2.10×10^{-12}	Likely regulated

Chr chromosome, *GS* gene significance, *MM* module membership, *TC* total cholesterol

¹Classification terms have been added to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations

cardiometabolic health. Families were composed of 16 mothers, 6 fathers, and 26 children. Families living under the same roof comprised at least the mother and one child aged between 8 and 18. Parents had to be the biological parents of their child (or children), in good general health, non-smokers, with body mass index (BMI) ranging between 18 and 35 kg/m², and free of any metabolic conditions requiring treatment, although the use of Synthroid® (levothyroxine) or oral contraceptive was tolerated. Children also had to be non-smokers, in good general health, and not using psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Blood samples were taken from both parents and children during their visit at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University. All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document.

Anthropometric and cardiometabolic measurements

Body weight, waist girth, and height were measured according to the procedures recommended by the Airlie Conference [58]. Blood samples were collected from an antecubital vein into vacutainer tubes containing EDTA after 12-h overnight fast and 48-h alcohol abstinence. Plasma was separated by centrifugation (2500 g for 10 min at 4 °C), and samples were aliquoted and frozen (-80 °C) for subsequent analyses. Enzymatic assays were used to measure plasma TC and TG concentrations [59, 60]. Precipitation of very-low-density lipoprotein (VLDL) and LDL particles in the infranantant with heparin-manganese chloride generated the HDL-C fraction [61]. LDL-C was calculated with the Friedewald formula [62]. ApoB100 concentrations were measured in plasma by the rocket immunoelectrophoretic method [63]. Using a sensitive assay, plasma C-reactive protein

(CRP) was measured by nephelometry (Prospec equipment Behring) [64].

RNA extraction and gene expression analysis

Total RNA was isolated and purified from whole blood using PAXgene Blood RNA Kit (Qiagen). Quantification and verification of the purified RNA was assessed using both the NanoDrop (Thermo Scientific, Wilmington, DE, USA) and the 2100 Bioanalyzer (Agilent Technologies, Cedar Creek, TX, USA). The HumanHT-12 v4 Expression BeadChip (Illumina Inc., San Diego, CA) was used to measure expression levels of ~47,000 probes (>31,000 annotated genes). This was performed at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada). The FlexArray software (version 1.6) [65] and the lumi R package were used to analyze and normalize gene expression levels. Probes with a detection *p* value ≤ 0.05 in at least 25% of all subjects were considered in the analysis. A total of 18,160 probes among the 47,323 probes on the microarray (38.4%) showed significant gene expression in the blood.

Plasma carotenoid measurements

Samples and standards used for the measurement of carotenoid concentrations were prepared as reported previously [7]. Briefly, 100 μL of fasting plasma samples were thawed a few hours before analysis. Plasma samples were transferred to Eppendorf tubes with 20 μL of 2-propanol and 20 μL of carotenoid standard. Samples were transferred on a 400-μL fixed well plate (ISOLUTE® SLE+, Biotage, Charlotte, NC) with 900 μL of hexane to isopropanol (90/10, v/v) in each well. Each extracted sample was evaporated under nitrogen and reconstituted with 300 μL of methanol to dichloromethane (65/35, v/v). Plates were shaken for 10 min and samples were transferred into high-performance liquid chromatography glass vials for analysis.

High-performance liquid chromatography (HPLC)-UV analysis was performed using an Agilent 1260 liquid handling system (Agilent, Mississauga, Ontario, Canada) as described previously [7]. Carotenoids were separated with a mobile phase consisting of methanol to water (98/2, v/v; Eluent A) and methyl-tert-butyl ether (MTBE; Eluent B; VWR, Mississauga, Ontario, Canada). Flow-rate was set at 1 mL/min and the gradient elution was as follows: 2% Eluent B (initial), 2.0–80% Eluent B (0.0–27.0 min), isocratic 80% Eluent B (27.0–31.0 min), 80.0–2.0% Eluent B (31.0–31.1 min), and isocratic 2% Eluent B (31.1–34.0 min). UV detector was set at 450 nm, and identification of each compound was confirmed using retention time and UV spectra (190–640 nm) of the pure compounds. Data acquisition was carried out with the Chemstation software (Agilent, Mississauga, Ontario, Canada). For all carotenoids, the concentrations are reported in micromoles per liter of plasma. One outlier in β -cryptoxanthin, defined as value falling outside of the mean \pm 4 standard deviations, was excluded from analyses.

Association between total carotenoids and gene expression levels

Total plasma carotenoid (micromoles per liter of plasma) concentrations were calculated as the sum of α -carotene, β -carotene, β -cryptoxanthin, lutein, lycopene, and zeaxanthin concentrations. Concentrations of plasma carotenoids are available in Additional file 1: Table S1. R software v2.14.1 (R Foundation for Statistical Computing; <http://www.r-project.org>) [66] was used to compute regressions between normalized gene expression levels of all 18,160 probes and total carotenoids adjusted for the family ID. Weighted gene co-expression network analysis (WGCNA) was performed with gene expression levels of 533 probes showing a significant association (p value \leq 0.05, obtained from the linear model function) with total carotenoids.

Weighted gene co-expression network analysis (WGCNA)

WGCNA was performed with the WGCNA package [21, 67] in R software [66]. First, co-expression was measured between each gene pair using Pearson correlation coefficients (varying from -1 to 1). Considering the small sample size of the present study, Pearson correlations measuring linear relationships were selected to avoid the pitfall of overfitting [68]. In order to transform the correlation coefficients into a weighted adjacency matrix (values ranging from 0 to 1), the co-expression similarity was raised to a power $\beta = 6$ [68]. The adjacency matrix allows measuring the connection strengths between nodes. From this matrix, we can build the topological overlap matrix (TOM) that considers the topological similarity. The TOM was then used to calculate the

corresponding dissimilarity (1-TOM) in order to form clusters. Average linkage hierarchical clustering coupled with the TOM-based dissimilarity was used to group genes with coherent expression profiles into modules [68]. More specifically, the dynamic tree cutting algorithm (deep split = 2, minimum number of genes per modules = 30, cut height = 0.25) was used to detect gene modules (clusters of densely interconnected genes in terms of co-expression). The Partitioning Around Medoids (PAM) method was also used to allow the assignment of outlying genes to modules. Colors are randomly assigned to modules except for the gray color, which is reserved for the module with unassigned genes. To identify modules that were significantly associated with lipid profile traits (TC, LDL-C, HDL-C, TG, ApoB100), correlations between module eigengenes (MEs) (i.e., the first principle component of the module, which represents the overall expression level of the module) [69] and traits were computed. Gene significance (GS), defined as the absolute correlation between the gene and the trait, was used to quantify associations of individual genes with lipid profile traits. To quantify the similarity of all genes to every module, a quantitative measure of module membership (MM) was defined as the correlation of the ME and the gene expression profile. Genes with the highest MM and highest GS were those with high significance (hub genes) [70]. The hub genes within a module were chosen based on $GS > 0.2$ and $MM > 0.6$, with a p value \leq 0.05. All these analyses were computed using commands implemented in the WGCNA package.

Functional enrichment analysis and hub genes classification

Reactome and Gene Ontology (GO) enrichment analysis was conducted with KEGG Orthology-Based Annotation System (KOBAS) on the genes of modules of interest [71]. Hypergeometric test/Fisher's exact test was used to identify significant pathways. FDR correction method (Benjamini and Hochberg) was used to account for multiple testing in enrichment analysis. The most significant Reactome terms were used to characterize brown and blue modules. TFs enrichment analysis was conducted using Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>), an online biological information database that integrates several biological databases [72]. The most significant TFs in each module were retained. Moreover, analyses were carried out in order to suggest potential implications of hub genes in the relationship between carotenoids and lipids. First of all, regressions between expression levels of all hub genes and plasma lipid concentrations adjusted for the family ID were computed in Statistical Analysis Software (SAS) version 9.4. More specifically, the seven transcripts of the brown module

were tested for associations with HDL-C and the 32 transcripts of the blue module were tested for associations with TC. Significant associations ($p \leq 0.05$) are presented in Additional file 5: Table S5. Considering that all hub genes initially showed associations with total carotenoids, potential causal relationships of hub genes were classified as follows: (1) unknown: genes that showed association with carotenoids but not with lipids, (2) likely regulated: genes that showed association with both carotenoids and lipids, and (3) likely regulator: genes that showed association with carotenoids and either have been associated with lipids in literature or showed association with lipids in addition to being shown to influence lipid levels in the literature. The literature of associations of genes with lipids is detailed in the discussion. These classification terms were used to quickly have an idea of the potential implications of hub genes in the relationship between carotenoids and lipids.

Statistical analysis

SAS version 9.4 was used to compute sex differences in cardiometabolic parameters between fathers and mothers and between daughters and sons using an unpaired *t* test.

Additional files

- Additional file 1: Table S1.** Concentrations of plasma carotenoids (micromoles per liter of plasma). (XLSX 15 kb)
- Additional file 2: Table S2.** Probes showing significant association with plasma total carotenoids ($n = 533$) (XLSX 42 kb)
- Additional file 3: Table S3.** Functional enrichment analysis in the brown module using GO and Reactome databases (XLSX 15 kb)
- Additional file 4: Table S4.** Functional enrichment analysis in blue module using Gene Ontology and Reactome databases. (XLSX 28 kb)
- Additional file 5: Table S5.** Significant *p* values of regressions between expression levels of hub genes and plasma lipid concentrations. (XLSX 11 kb)

Abbreviation

ApoB100: Apolipoprotein B100; GO: Gene Ontology; GS: Gene significance; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; ME: Module eigengenes; MM: Module membership; SD: Standard deviation; TC: Total cholesterol; TF: Transcription factor; TG: Triglycerides; TOM: Topological overlap matrix; WGCNA: Weighted gene co-expression network analysis

Acknowledgments

We would like to thank Christian Couture, Véronique Garneau, Catherine Raymond, and Véronique Richard who contributed to the success of this study.

Funding

This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. MCV is Tier 1 Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. BLT is a recipient of a scholarship from the Canadian Institutes of Health Research (CIHR).

Availability of data and materials

Expression datasets supporting the conclusion of this article are available in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (GSE114620).

Authors' contributions

Each author contributed to the work. BL and MCV designed the research. BLT, FG, and LP conducted the research and performed statistical analyses. BLT wrote the paper. BLT and MCV have primary responsibility for the final content. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document. The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada. ²School of Nutrition, Laval University, 2425 rue de l'Agriculture, Quebec City, QC G1V 0A6, Canada. ³Department of Kinesiology, Laval University, 2300 rue de la Terrasse, Quebec City, QC G1V 0A6, Canada.

Received: 15 January 2019 Accepted: 15 April 2019

Published online: 08 May 2019

References

- Souverain OW, de Vries JH, Freese R, Watzl B, Bub A, Miller ER, et al. Prediction of fruit and vegetable intake from biomarkers using individual participant data of diet-controlled intervention studies. *Br J Nutr*. 2015; 113(9):1396–409.
- Couillard C, Lemieux S, Vohl MC, Couture P, Lamarche B. Carotenoids as biomarkers of fruit and vegetable intake in men and women. *Br J Nutr*. 2016;116(7):1206–15.
- Maiani G, Caston MJ, Catasta G, Toti E, Cambrodon IG, Bysted A, et al. Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. *Mol Nutr Food Res*. 2009; 53(Suppl 2):S194–218.
- Eroglu A, Harrison EH. Carotenoid metabolism in mammals, including man: formation, occurrence, and function of apocarotenoids. *J Lipid Res*. 2013; 54(7):1719–30.
- Allore T, Lemieux S, Vohl MC, Couture P, Lamarche B, Couillard C. Correlates of the difference in plasma carotenoid concentrations between men and women. *Br J Nutr*. 2019;121(2):172–181.
- Farook VS, Reddivari L, Mummidi S, Puppala S, Arya R, Lopez-Alvarenga JC, et al. Genetics of serum carotenoid concentrations and their correlation with obesity-related traits in Mexican American children. *Am J Clin Nutr*. 2017;106(1):52–8.
- Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Genetic and common environmental contributions to familial resemblances in plasma carotenoid concentrations in healthy families. *Nutrients*. 2018;10(8):1002.
- Wang Y, Chung SJ, McCullough ML, Song WO, Fernandez ML, Koo SI, et al. Dietary carotenoids are associated with cardiovascular disease risk biomarkers mediated by serum carotenoid concentrations. *J Nutr*. 2014; 144(7):1067–74.
- Clevidence BA, Bieri JG. Association of carotenoids with human plasma lipoproteins. *Methods Enzymol*. 1993;214:33–46.
- Rao AV, Rao LG. Carotenoids and human health. *Pharmacol Res*. 2007;55(3): 207–16.

11. Paiva SA, Russell RM. Beta-carotene and other carotenoids as antioxidants. *J Am Coll Nutr.* 1999;18(5):426–33.
12. Bertram JS. Carotenoids and gene regulation. *Nutr Rev.* 1999;57(6):182–91.
13. D'Adamo CR, D'Urso A, Ryan KA, Yerges-Armstrong LM, Semba RD, Steinle NI, et al. A common variant in the SETD7 gene predicts serum lycopene concentrations. *Nutrients.* 2016;8(2):82.
14. Beydoun MA, Nalls MA, Canas JA, Evans MK, Zonderman AB. Gene polymorphisms and gene scores linked to low serum carotenoid status and their associations with metabolic disturbance and depressive symptoms in African-American adults. *Br J Nutr.* 2014;112(6):992–1003.
15. Ferrucci L, Perry JR, Matteini A, Perola M, Tanaka T, Silander K, et al. 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.* 2009;84(2):123–33.
16. Zubair N, Kooperberg C, Liu J, Di C, Peters U, Neuhauser ML. Genetic variation predicts serum lycopene concentrations in a multiethnic population of postmenopausal women. *J Nutr.* 2015;145(2):187–92.
17. Borel P, de Edelenyi FS, Vincent-Baudry S, Malezet-Desmoulin C, Margotat A, Lyan B, et al. Genetic variants in BCMO1 and CD36 are associated with plasma lutein concentrations and macular pigment optical density in humans. *Ann Med.* 2011;43(1):47–59.
18. Bohn T, Desmarchelier C, Dragsted LO, Nielsen CS, Stahl W, Ruhl R, et al. Host-related factors explaining interindividual variability of carotenoid bioavailability and tissue concentrations in humans. *Mol Nutr Food Res.* 2017;61(6):1600685.
19. Hix LM, Lockwood SF, Bertram JS. Bioactive carotenoids: potent antioxidants and regulators of gene expression. *Redox report.* 2004;9(4):181–91.
20. Sharoni Y, Danilenko M, Dubi N, Ben-Dor A, Levy J. Carotenoids and transcription. *Arch Biochem Biophys.* 2004;430(1):89–96.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 2008;9:559.
22. Borel P, Desmarchelier C, Nowicki M, Bott R. A combination of single-nucleotide polymorphisms is associated with interindividual variability in dietary beta-carotene bioavailability in healthy men. *J Nutr.* 2015;145(8):1740–7.
23. Borel P, Desmarchelier C, Nowicki M, Bott R. Lycopene bioavailability is associated with a combination of genetic variants. *Free Radic Biol Med.* 2015;83:238–44.
24. Borel P, Desmarchelier C, Nowicki M, Bott R, Morange S, Lesavre N. Interindividual variability of lutein bioavailability in healthy men: characterization, genetic variants involved, and relation with fasting plasma lutein concentration. *Am J Clin Nutr.* 2014;100(1):168–75.
25. Borel P, Desmarchelier C. Bioavailability of fat-soluble vitamins and phytochemicals in humans: effects of genetic variation. *Annu Rev Nutr.* 2018;38:69–96.
26. Seo JY, Kim H, Seo JT, Kim KH. Oxidative stress induced cytokine production in isolated rat pancreatic acinar cells: effects of small-molecule antioxidants. *Pharmacology.* 2002;64(2):63–70.
27. Mueller E, Smith M, Sarraf P, Kroll T, Aiyer A, Kaufman DS, et al. Effects of ligand activation of peroxisome proliferator-activated receptor gamma in human prostate cancer. *Proc Natl Acad Sci U S A.* 2000;97(20):10990–5.
28. Inoue Y, Shimazawa M, Nagano R, Kuse Y, Takahashi K, Tsuruma K, et al. Astaxanthin analogs, adonixanthin and lycopene, activate Nrf2 to prevent light-induced photoreceptor degeneration. *J Pharmacol Sci.* 2017;134(3):147–57.
29. Huang Q, Wang TH, Lu WS, Mu PW, Yang YF, Liang WW, et al. Estrogen receptor alpha gene polymorphism associated with type 2 diabetes mellitus and the serum lipid concentration in Chinese women in Guangzhou. *Chin Med J.* 2006;119(21):1794–801.
30. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45(11):1274–83.
31. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum Mol Genet.* 2017;26(9):1770–84.
32. Jiang Z, Knudsen NH, Wang G, Qiu W, Naing ZCC, Bai Y, et al. Genetic control of fatty acid beta-oxidation in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol.* 2017;56(6):738–48.
33. Miao L, Yin RX, Pan SL, Yang S, Yang DZ, Lin WX. Weighted gene co-expression network analysis identifies specific modules and hub genes related to hyperlipidemia. *Cellular physiology and biochemistry.* 2018;48(3):1151–63.
34. Khor CC, Vannberg FO, Chapman SJ, Guo H, Wong SH, Walley AJ, et al. CISH and susceptibility to infectious diseases. *N Engl J Med.* 2010;362(22):2092–101.
35. Gruaz L, Delucinge-Vivier C, Descombes P, Dayer JM, Burger D. Blockade of T cell contact-activation of human monocytes by high-density lipoproteins reveals a new pattern of cytokine and inflammatory genes. *PLoS One.* 2010;5(2):e9418.
36. TLT L, Sivaguru M, Velayutham M, Cardounel AJ, Michels M, Barefield D, et al. Oxidative stress in dilated cardiomyopathy caused by MYBPC3 mutation. *Oxidative Med Cell Longev.* 2015;2015:424751.
37. Jorgenrud B, Jalanko M, Helio T, Jaaskelainen P, Laine M, Hilvo M, et al. The metabolome in Finnish carriers of the MYBPC3-Q1061X mutation for hypertrophic cardiomyopathy. *PLoS One.* 2015;10(8):e0134184.
38. Rudkowska I, Perusse L, Bellis C, Blangero J, Despres JP, Bouchard C, et al. Interaction between common genetic variants and total fat intake on low-density lipoprotein peak particle diameter: a genome-wide association study. *J Nutrigenet Nutrigenomics.* 2015;8(1):44–53.
39. Lean CB, Lee EJ. Genetic variations of the MCT4 (SLC16A3) gene in the Chinese and Indian populations of Singapore. *Drug Metab Pharmacokinet.* 2012;27(4):456–64.
40. Bohn SK, Myhrstad MC, Thoresen M, Holden M, Karlsen A, Tunheim SH, et al. Blood cell gene expression associated with cellular stress defense is modulated by antioxidant-rich food in a randomised controlled clinical trial of male smokers. *BMC Med.* 2010;8:54.
41. Bringman-Rodenbarger LR, Guo AH, Lyssiotis CA, Lombard DB. Emerging roles for SIRT5 in metabolism and cancer. *Antioxid Redox Signal.* 2018;28(8):677–90.
42. Liang F, Wang X, Ow SH, Chen W, Ong WC. Sirtuin 5 is anti-apoptotic and anti-oxidative in cultured SH-EP neuroblastoma cells. *Neurotox Res.* 2017;31(1):63–76.
43. Yu J, Sadhukhan S, Noriega LG, Moullan N, He B, Weiss RS, et al. Metabolic characterization of a Sirt5 deficient mouse model. *Sci Rep.* 2013;3:2806.
44. Yu CY, Theusch E, Lo K, Mangravite LM, Naidoo D, Kutilova M, et al. HNRNPA1 regulates HMGR alternative splicing and modulates cellular cholesterol metabolism. *Hum Mol Genet.* 2014;23(2):319–32.
45. Echaniz-Laguna A, Chassagne M, Ceresuela J, Rouvet I, Padet S, Acquaviva C, et al. Complete loss of expression of the ANT1 gene causing cardiomyopathy and myopathy. *J Med Genet.* 2012;49(2):146–50.
46. Won JC, Park JY, Kim YM, Koh EH, Seol S, Jeon BH, et al. Peroxisome proliferator-activated receptor-gamma coactivator 1-alpha overexpression prevents endothelial apoptosis by increasing ATP/ADP translocase activity. *Arterioscler Thromb Vasc Biol.* 2010;30(2):290–7.
47. Wang X, Wang D, Yuan N, Liu F, Wang F, Wang B, et al. The prognostic value of PCNA expression in patients with osteosarcoma: a meta-analysis of 16 studies. *Medicine.* 2017;96(41):e8254.
48. Cheng HC, Chien H, Liao CH, Yang YY, Huang SY. Carotenoids suppress proliferating cell nuclear antigen and cyclin D1 expression in oral carcinogenic models. *J Nutr Biochem.* 2007;18(10):667–75.
49. Donaldson MS. A carotenoid health index based on plasma carotenoids and health outcomes. *Nutrients.* 2011;3(12):1003–22.
50. Gervois P, Torra IP, Fruchart JC, Staels B. Regulation of lipid and lipoprotein metabolism by PPAR activators. *Clin Chem Lab Med.* 2000;38(1):3–11.
51. Keller H, Dreyer C, Medin J, Mahfoudi A, Ozato K, Wahli W. Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. *Proc Natl Acad Sci U S A.* 1993;90(6):2160–4.
52. Moreau C, Lefebvre JF, Jomphe M, Bherer C, Ruiz-Linares A, Vezina H, et al. Native American admixture in the Quebec founder population. *PLoS One.* 2013;8(6):e65507.
53. Gruber M, Chappell R, Millen A, LaRowe T, Moeller SM, Iannaccone A, et al. Correlates of serum lutein + zeaxanthin: findings from the Third National Health and Nutrition Examination Survey. *J Nutr.* 2004;134(9):2387–94.
54. Al-Delaimy WK, van Kappel AL, Ferrari P, Slimani N, Steghens JP, Bingham S, et al. Plasma levels of six carotenoids in nine European countries: report from the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr.* 2004;7(6):713–22.
55. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in blood leukocyte DNA methylation levels. *Epigenetics.* 2016;11(11):831–8.
56. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human whole blood transcriptome. *BMC Genomics.* 2018;19(1):300.

57. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects. *Nutr Res.* 2019;61:22–30.
58. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropometric measurements: The Airlie (VA) Consensus Conference. Champaign, IR, USA: 1988.
59. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. *ClinChimActa.* 1987;166(1):1–8.
60. Burstein M, Samaille J. On a rapid determination of the cholesterol bound to the serum alpha- and beta-lipoproteins. *ClinChimActa.* 1960;5:609.
61. Albers JJ, Warnick GR, Wiebe D, King P, Steiner P, Smith L, et al. Multi-laboratory comparison of three heparin-Mn²⁺ precipitation procedures for estimating cholesterol in high-density lipoprotein. *ClinChem.* 1978;24(6): 853–6.
62. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *ClinChem.* 1972;18(6):499–502.
63. Laurell CB. Quantitative estimation of proteins by electrophoresis in agarose gel containing antibodies. *AnalBiochem.* 1966;15(1):45–52.
64. Pirro M, Bergeron J, Dagenais GR, Bernard PM, Cantin B, Despres JP, et al. Age and duration of follow-up as modulators of the risk for ischemic heart disease associated with high plasma C-reactive protein levels in men. *ArchInternMed.* 2001;161(20):2474–80.
65. Blazejczyk MMM, Nadon R. FlexArray. In: A statistical data analysis software for gene expression microarrays. Montreal: Genome Quebec; 2007.
66. Team RC. R: a language and environment for statistical computing. Vienna: R Foundation for statistical computing; 2013.
67. Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw.* 2012;46(11).
68. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* 2005;4:Article17.
69. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol.* 2007;1:54.
70. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A.* 2006;103(46):17402–7.
71. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39(Web Server issue):W316–22.
72. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90–7.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

