Check for updates

RESEARCH NOTE

Transcription factor regulation as a mechanism of confounding effects between distinct human traits [version 1; referees: 1 not approved]

Milos Pjanic*, Clint L. Miller*, Thomas Quertermous

Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, 94305, USA

* Equal contributors

Abstract

Genome-wide association studies (GWAS) to date have discovered thousands of genetic variants linked to human diseases and traits, which hold the potential to unravel the mechanisms of complex phenotypes. However, given that the majority of these associated variants reside in non-coding genomic regions, their predicted *cis* and *trans*-regulatory functions remain largely undefined. Here we show that correlation between human diseases and traits can follow geographical distribution of human populations, and that the underlying mechanism is at least partly genetically based. We report two Type 2 Diabetes (T2D) GWAS variants (rs7903146 and rs12255372) in the *TCF7L2* locus that regulate expression in skin tissues but not lymphoblastoid or adipose tissues, of the *KITLG* gene that encodes an important regulator of melanogenesis and light hair color in European populations. We also report extensive binding events of TCF7L2 protein in the promoter region, immediate upstream region and first intron of the *KITLG* gene, which supports a *trans*-interaction between *TCF7L2* and *KITLG*. We further show that both light hair color and T2D genetic variants are correlated with geographic latitude. Taken together, our observations suggest that natural variation in transcription factor loci in European human populations may be an underlying and confounding factor for the geographical correlation between human phenotypes, such as type 2 diabetes and light hair color. We postulate that transcription factor regulation may confound the correlation between seemingly diverse human traits. Furthermore, our findings demonstrate the importance of dissecting the genomic architecture of GWAS loci using multiple genetic and genomic datasets.

Open Peer Review

Referee Status: ✗

	Invited Referees
	1
version 1 published 25 Nov 2015	✗ report

1 **Gregory Gibson**, Georgia Institute of Technology USA, **Urko M. Marigorta**, Georgia Institute of Technology USA

Discuss this article

Comments (0)

Corresponding author: Milos Pjanic (mpjanic@stanford.edu)

Competing interests: The authors have no conflicts of interest or competing interests to disclose.

Introduction

A recent publication[1] has demonstrated the potential causative mechanism of a genome-wide association study (GWAS) locus for the development of blond hair. Through a series of elegant *in vivo* experiments in mice, the study's findings strengthen association of single nucleotide polymorphism (SNP) rs12821256, initially discovered as one of the top GWAS hits in European populations[2], with light hair color development. This work implicates a mechanism of long-range regulation of a gene on chromosome 12, termed *KITLG* that encodes the ligand for a receptor-type protein-tyrosine kinase, and is located 350kb away from the variant. Further, using data generated by the ENCODE consortium, the study reveals a molecular mechanism by which SNP rs12821256 confers the blond hair phenotype via directly altering a canonical binding site for transcription factor TCF7L2[3]. This may shed light on possible *cis*- and *trans*-acting mechanisms responsible for the association of rs12821256 with the quantitative trait of light hair color.

On the other hand, the *TCF7L2* locus on chromosome 10 is well-known for its strong association with type 2 diabetes (T2D) and glycemic traits from several GWAS studies[4,5]. It confers the strongest effect on T2D to date, with a per-allele odds ratio of 1.39[6]. Lead risk-associated SNPs from the *TCF7L2* locus include two intronic SNPs (rs7903146 and rs4506565). The majority of SNPs from the *TCF7L2* locus are non-coding and may alter the levels of expression or affect alternative splicing of *TCF7L2*, while SNPs located in *TCF7L2* exons give rise to alternate protein isoforms. In addition, numerous SNPs from this locus that are in linkage disequilibrium (LD) with GWAS lead SNPs could be candidates for the causal variant(s). Given these reports, it seems likely that specific *TCF7L2* expression levels or the composition of its 13 or more transcripts (UCSC annotation) and isoforms in pancreatic beta cells confer risk for T2D, while in melanocytes the composition of *TCF7L2* variants and levels may influence trans TCF7L2 protein binding to SNP rs12821256 to alter expression of the downstream *KITLG* gene, an important regulator of melanogenesis.

TCF7L2 is expressed in a variety of human tissues, where it plays a critical role in the Wnt signaling pathway. In skin tissues *TCF7L2* reaches moderate expression levels with RPKM (Reads Per Kilobase of transcript per Million mapped reads) values between 10 and 20, which are higher than that observed in pancreas (<10)[7].

Main body

In addition to binding to rs12821256, we report here that TCF7L2 binds to the promoter region of the *KITLG* gene (as shown in the ENCODE ChIP-Seq data sets), as well as throughout the first intron and immediate upstream region, and overlaps the active enhancer histone modification mark H3K27ac (Figure 1A), which further implicates its role in the regulation of *KITLG* expression. When we queried the Genotype-Tissue Expression (GTEx) database or eQTL resources from the Gilad/Pritchard group there were no SNPs from the *TCF7L2* locus detected as expression quantitative trait loci (eQTL SNPs) for KITLG (search terms in Supplementary Table S1), nor when we investigated HapMap data through the GENEVAR (GENe Expression VARiation) platform. However, a significant eQTL association between *TCF7L2* SNPs (rs7903146 and rs12255372) and *KITLG* was observed in skin tissues in data from the MuTHER (Multiple Tissue Human Expression Resource) healthy female twin studies[8] (Figure 1B, p=0.0089 and 0.0349, respectively), implicating a strong *trans*-eQTL interaction in skin tissues compared to Lymphoblastoid cell lines (LCL) or adipose tissues where either the absence of, or weak eQTL association was found.

As demonstrated by the International Diabetes Federation data for 2014[9], T2D is less prevalent in northern European populations (compared to, e.g. southern Europeans). Data of the frequency of T2D patients in Europe shows that southern European countries, i.e. Spain (7.9%), Portugal (9.6%), Balkan countries (9.8%) and Turkey (14.8%) have the highest percentage of T2D patients in Europe (average 10.2%). On the other hand, northern European countries like Britain (3.9%), Sweden (4.5%), Norway (5.2), Baltic countries (3.8%, 5.0% and 5.7%), and Iceland (3.2%) have lower percentage of diabetics compared to the rest of Europe (average 3.9%). This difference in disease prevalence could be attributed to differences in dietary or other environmental factors, but also could reflect differences in allele frequency of disease-associated alleles. In fact, the frequencies of SNP rs12821256 and light hair color are more common in northern European populations, e.g., blond and light brown hair reaching 75% in Icelandic populations and rs12821256 MAF reaching its frequency maximum of 0.19 in Iceland (Supplementary Figure S1)[1,2]. Similarly, using data from ALFRED (ALlele FRequency Database), we found an inverse correlation of population's geographic latitude and frequency of *TCF7L2* SNP rs7903146 (Figure 2A), and another *TCF7L2* SNP rs12255372 also showed a similar trend (Figure 2B). Thus, it is intriguing to speculate whether TCF7L2 protein isoforms may give rise to light hair color via binding to rs12821256 and regulating the *KITLG* gene in one cell type (melanocytes), while in pancreatic beta cells they may act as risk factors for the development of diabetes (Figure 3), through *TCF7L2* gene regulation and potential cross-composition of *TCF7L2* isoforms.
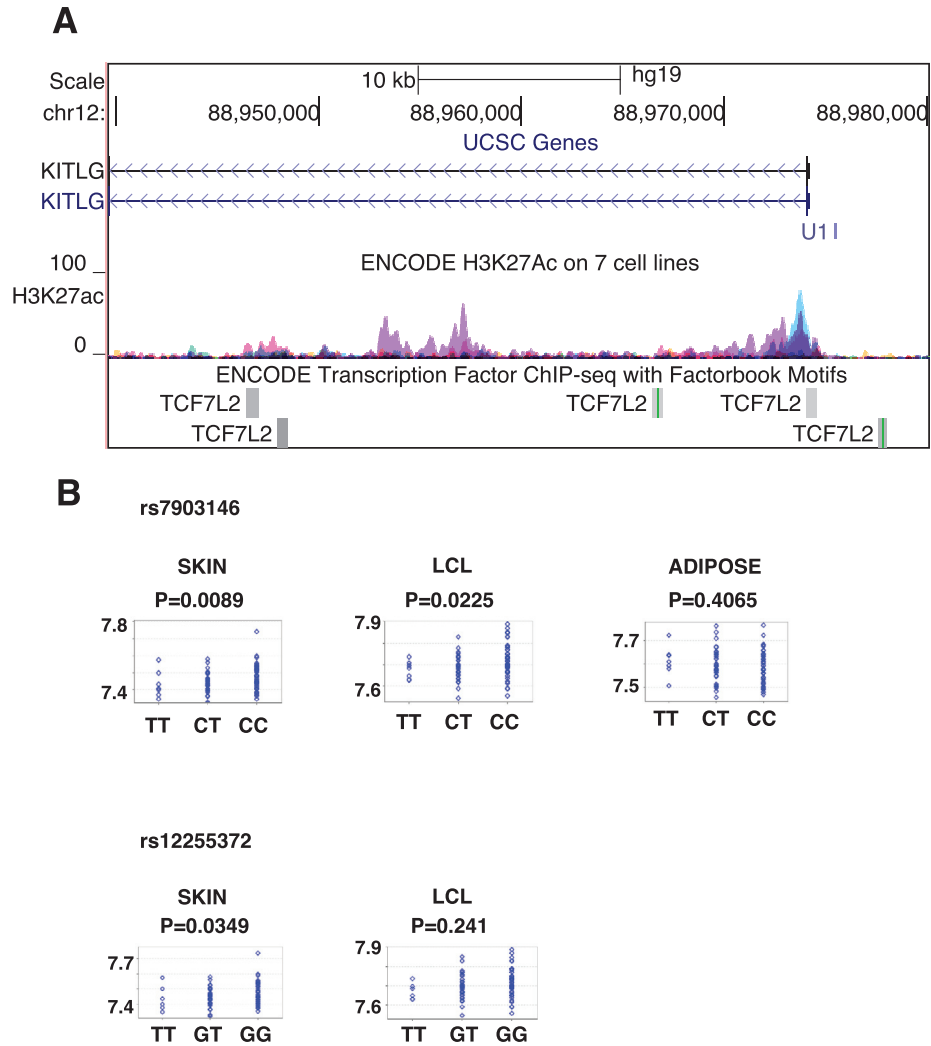
Figure 1. TCF7L2 locus variation and protein occupancy implicated in regulation of KITLG gene. A. TCF7L2 protein binding at the KITLG promoter, upstream of the KITLG promoter and multiple binding events in the first intron of KITLG gene. TCF7L2 binding sites overlap regulatory histone mark H3K27ac, implying their functionality in gene regulation. Data were taken from the ENCODE consortium. **B**. eQTL analysis of two T2D SNPs from TCF7L2 locus (rs7903146 and rs12255372) and KITLG gene in multiple tissues: skin, lymphoblastoid cell line (LCL) and adipose. Data from MuTHER healthy female twin studies.
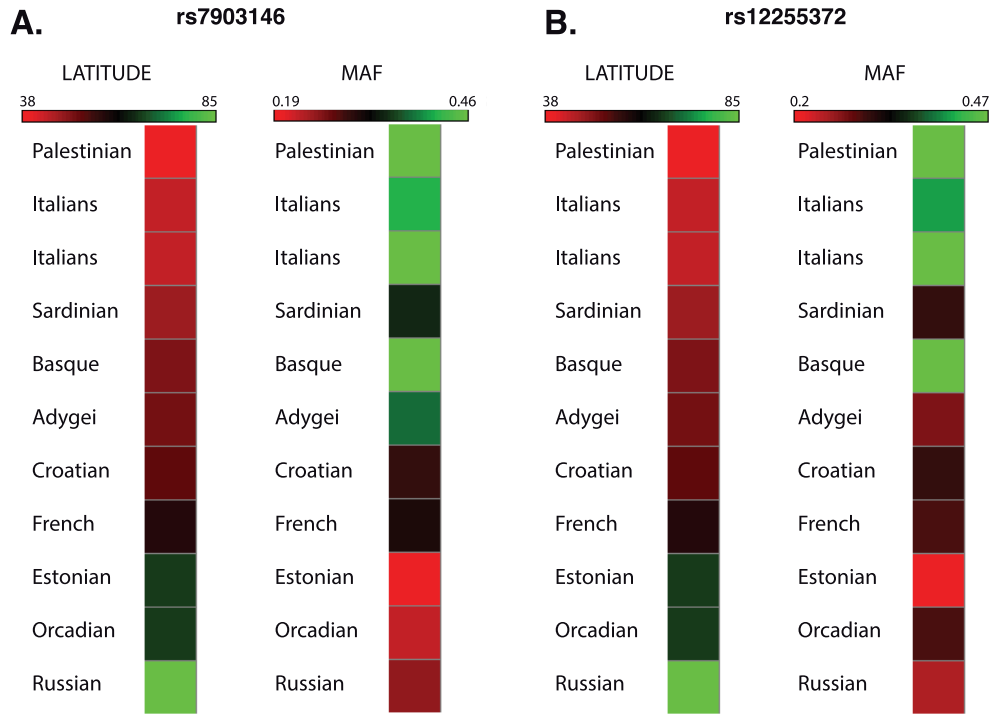
Figure 2. Inverse correlation of geographic latitude and T2D SNP minor allele frequency. Maximal geographical latitude of the population and T2D SNP minor allele frequency (MAF) were taken from Alfred (Allele Frequency Database) and plotted as a heatmap. **A**. SNP rs7903146 **B**. SNP rs12255372.
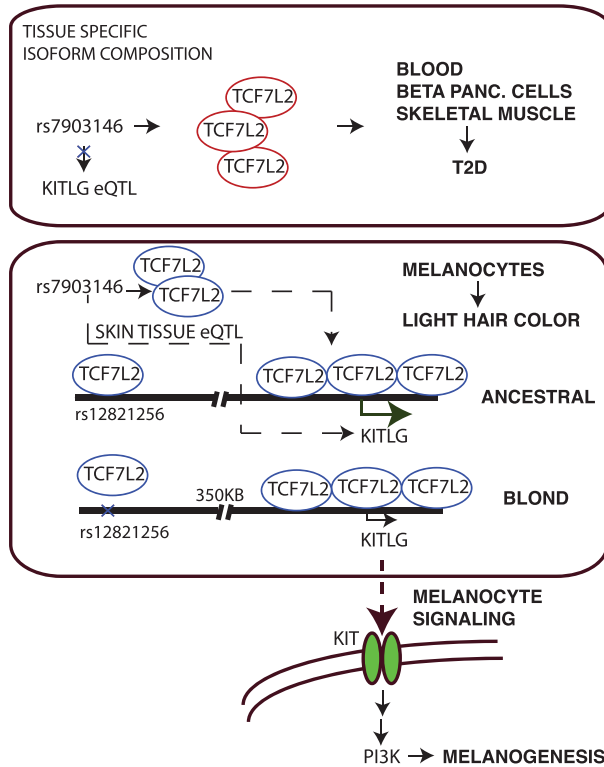

Figure 3. Schematic representation of transcription factor regulation as basis for confounding effects between diseases and traits. T2D SNP rs7903146 from TCF7L2 locus is shown as eQTL SNP for KITLG gene in skin tissues. eQTL association is lost in other tissues, indicating regulation of KITLG gene by TCF7L2 isoforms explicitly in skin tissues.

Conclusion

In summary, the putative *trans*-eQTL interaction in skin tissues we report here implicates natural genetic variation in the T2D locus, *TCF7L2*, to regulate expression of *KITLG*, a gene linked to light hair color development. We postulate that this could be the underlying genetic mechanism accounting for the association between hair color and T2D risk in European populations. Our observations here strengthen the hypothesis of a genetically determined correlation between diseases and traits in human population, as also demonstrated in a recent publication with the inversely correlated height and coronary artery disease (CAD) phenotypes, where height-associated variants were associated with an increase of 13.5% in the risk of CAD[10]. Furthermore, these observations illustrate how investigating the genetic architecture underlying complex traits and diseases may inform appropriate risk stratification in diverse human populations.

Author contributions

MP conceived the study. MP, CLM, and TQ jointly analyzed the results and wrote and revised the manuscript. All authors have seen and agreed to the final content of the manuscript.

Competing interests

The authors have no conflicts of interest or competing interests to disclose.

Supplementary information

Supplementary Figure S1. Correlation of geographic latitude and rs12821256 minor allele frequency.

Maximal geographic latitude of the population and rs12821256 minor allele frequency (MAF) were taken from Alfred (Allele Frequency Database) and plotted as a heatmap.

Click here to access the data.

Supplementary Table S1. Query results from the Genotype-Tissue Expression (GTEx) database and eQTL resources from the Gilad/Pritchard group.

There were no SNPs from the *TCF7L2* locus that could be detected as expression quantitative trait loci SNPs (eQTL SNPs) for KITLG gene using GTEx database, neither TCF7L2 locus (500Kb from chr10:114553649-115062602) could be linked to any trans eQTL SNP-gene interaction using eQTL resources from the Gilad/Pritchard group.

Click here to access the data.

References

1. Guenther CA, Tasic B, Luo L, *et al.*: **A molecular basis for classic blond hair color in Europeans.** *Nat Genet.* 2014; **46**(7): 748–52.
   PubMed Abstract | **Publisher Full Text**

2. Sulem P, Gudbjartsson DF, Stacey SN, *et al.*: **Genetic determinants of hair, eye and skin pigmentation in Europeans.** *Nat Genet.* 2007; **39**(12): 1443–52.
   PubMed Abstract | **Publisher Full Text**

3. ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome.** *Nature.* 2012; **489**(7414): 57–74.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. McCarthy MI, Zeggini E: **Genome-wide association studies in type 2 diabetes.** *Curr Diab Rep.* 2009; **9**(2): 164–71.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Grant SF, Thorleifsson G, Reynisdottir I, *et al.*: **Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes.** *Nat Genet.* 2006; **38**(3): 320–3.
   PubMed Abstract | **Publisher Full Text**

6. Palomaki GE, Melillo S, Marrone M, *et al.*: **Use of genomic panels to determine risk of developing type 2 diabetes in the general population: a targeted evidence-based review.** *Genet Med.* 2013; **15**(8): 600–11.
   PubMed Abstract | **Publisher Full Text**

7. GTEx Consortium: **The Genotype-Tissue Expression (GTEx) project.** *Nat Genet.* 2013; **45**(6): 580–5.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Nica AC, Parts L, Glass D, *et al.*: **The architecture of gene regulatory variation across multiple human tissues: the MuTHER study.** *PLoS Genet.* 2011; **7**(2): e1002003.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Federation ID: **IDF Diabetes Atlas.** 6th edn. Brussels, Belgium: International Diabetes Federation, 2013.
   Reference Source

10. Nelson CP, Hamby SE, Saleheen D, *et al.*: **Genetically determined height and coronary artery disease.** *N Engl J Med.* 2015; **372**(17): 1608–18.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Referee Status: ✖

Version 1

Referee Report 07 January 2016

✖   **Gregory Gibson**, **Urko M. Marigorta**

Center for Integrative Genomics, School of Biology, Georgia Institute of Technology, Atlanta, GA, USA

Pjanic, Miller and Quertermous suggest a genetic mechanism that may account for the correlation of latitude, frequency of blond hair color and prevalence of type 2 diabetes (T2D) in European populations. The mechanism involves TCF7L2 and its regulation of KITLG in skin tissue. Specifically, they describe (i) that rs12821256 (SNP associated with blond hair color) alters a binding site for TCF7L2, (ii) that KITLG contains other binding sites for TCF7L2 in its vicinity (Fig1A in the manuscript) and (iii) that rs7903146 in TCF7L2 (the strongest known risk SNP for T2D) acts as a trans-eQTL for KITLG in skin tissue (Fig 1B in the manuscript).

The manuscript does not provide new evidence, but connects disparate evidence from GWAS, ENCODE, eQTL databases and molecular studies to put forward a very interesting hypothesis. However, and even if the pieces do fit together, more clarifying evidence is needed to validate this hypothesis:

1. Fig 1B shows that C-allele increases expression of KITLG in skin tissue (this is T2D protective allele, but this is not indicated anywhere in the text). According to the authors' mechanism, C-allele should also lead to higher expression of TCF7L2 and consequently lower expression of KITLG in individuals with "blond" allele at rs12821256. However, and here lies the main caveat of the study, the evidence for a trans-eQTL is minuscule. The authors need to explain how much of the variance is explained at each step (for instance, variance in expression levels of KITLG in skin tissue that can be explained by rs7903146). Even if the proposed mechanism is interesting, a minor contribution by rs7903146 could not "explain" the core hypothesis.

2. The authors should go beyond SNP associations and validate this mechanism by looking at TCF7L2 expression levels (and its correlation with KITLG levels among genotypes) in skin samples from GTEx. There are >200 such individuals available, which should be enough to validate this point.

3. Moreover, other association evidence from GWAS could strengthen the hypothesis. For instance, rs7903146 should have a suggestive p-value for being a "blond hair" SNP and rs12821256 blond allele should in turn have a suggestive p-value for being a T2D protective allele (as blond people should be overrepresented among T2D cases). The authors could check publicly available GWAS p-value files to check this point (preferably within cohorts of a very homogeneous genetic background, e.g. Iceland).

4. Are there studies on the extent to which the lower prevalence of T2D in Northern Europeans is genetic in origin? Is there any evidence from less heterogeneous sources at the environmental level? For instance, do European Americans of Northern European ancestry have less T2D than other European Americans? Prevalence data is suggestive, but more evidence would reinforce the authors' point about a pleiotropic mechanism between blond hair color and genetic protection to T2D. At the very least, the authors should discuss more about this part of the hypothesis.

5. Related to the previous point, do the authors think this molecular mechanism explains the lower prevalence of T2D in Northern Europeans, or it is one of many other causal factors? This aspect should be made more clear for readers.

6. Fig 2B is redundant

We have read this submission. We believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Competing Interests: No competing interests were disclosed.