Candidate gene analysis of the fibrinogen phenotype reveals the importance of polygenic co-regulation

H. Toinét Cronjé †, Cornelie Nienaber-Rousseau †, Lizelle Zandberg, Tinashe Chikowore, Zelda de Lange, Tertia van Zyl and Marlien Pieters

Centre of Excellence for Nutrition, *North-West University, Potchefstroom, South Africa*

Correspondence to Cornelie Nienaber-Rousseau: *Private bag x6001, Nutrition, Box 594, Potchefstroom 2520, South Africa.*
cornelie.nienaber@nwu.ac.za
http://dx.doi.org/10.1016/j.matbio.2016.10.005

Abstract

Fibrinogen and its functional aspects have been linked to cardiovascular disease. There is vast discrepancy between the heritability of fibrinogen concentrations observed in twin studies and the heritability uncovered by genome wide association studies. We postulate that some of the missing heritability might be explained by the pleiotropic and polygenic co-regulation of fibrinogen through multiple targeted genes, apart from the fibrinogen genes themselves. To this end we investigated single nucleotide polymorphisms (SNPs) in genes coding for phenotypes associated with total and γ′ fibrinogen concentrations and clot properties. Their individual and accumulative associations with the fibrinogen variables were explored together with possible co-regulatory processes as a result of the gain and loss of transcription factor binding sites (TFBS). Seventy-eight SNPs spanning the *APOB*, *APOE*, *CBS*, *CRP*, *F13A1*, *FGA*, *FGB*, *FGG*, *LDL-R*, *MTHFR*, *MTR, PCSK-9* and *SERPINE-1* genes were included in the final analysis. A novel *PCSK-9* SNP (rs369066144) was identified in this population, which associated significantly ($p = 0.04$) with clot lysis time (CLT). Apart from SNPs in the fibrinogen (*FGA*, *FGB* and *FGG*) and FXIII (*F13A1*) genes, the fibrinogen phenotypes were also associated with SNPs in genes playing a role in lipid homeostasis (*LDL-R, PCSK-9*) together with *CBS* and *CRP* polymorphisms (particularly, *CRP*-rs3093068). The genetic risk scores, presenting accumulative genetic risk, were significantly associated ($p \leq 0.007$) with total and γ′ fibrinogen concentrations, lag time, slope and CLT, highlighting the importance of a polygenetic approach in determining complex phenotypes. SNPs significantly associated with the fibrinogen phenotypes, resulted in a total of 75 TFBS changes, of which 35 resulted in a loss and 40 in a gain of TFBS. In terms of co-regulation, V$IRF4.02, V$E2FF and V$HIFF were of particular importance. The investigation into TFBS provided valuable insight as to how sequence divergences in seemingly unrelated genes can result in transcriptional co-regulation of the fibrinogen phenotypes. The observed associations between the identified SNPs and the fibrinogen phenotypes therefore do not imply direct effects on cardiovascular disease outcomes, but may prove useful in explaining more of the genetic regulation of the investigated fibrinogen phenotypes.

Introduction

Fibrinogen is a hexameric glycoprotein consisting of two alpha (α), two beta (β) and two gamma (γ) polypeptide chains, and is encoded for by the fibrinogen α, β and γ chain genes (*FGA*, *FGB* and *FGG*) on the q-arm of chromosome four [1]. Several variants of the fibrinogen molecule exist, of which fibrinogen gamma prime (γ′) is considered to be a common splice variant contributing to between 8 and 11% of total plasma fibrinogen concentrations [2,3]. Fibrinogen γ′ arises in response to alternative splicing of the carboxyl-terminal region of γ-mRNA, resulting in a higher molecular weight γ′ than the γA chain [2,3]. Fibrinogen is an essential haemostatic protein, as the precursor of fibrin in the final stages of

Table 1. Descriptive characteristics of the study participants.

Variable	Median (25–75th percentiles)
Age (years)	48 (41–56)
Gender [n (%)]	
Male	628 (37.4)
Female	1049 (62.6)
HIV status [n (%)]	
Positive	269 (16.1)
Negative	1401 (83.9)
Total fibrinogen (g/L)	2.9 (2.2–5.0)
Fibrinogen γ′ (%)	10.2 (1.2–14.7)
Lag time (min)	6.5 (5.1–7.8)
Slope ($\times 10^{-3}$ au/s)	9.0 (6.5–12.0)
Maximum absorbance (nm)	0.4 (0.3–0.5)
CLT (min)	57.1 (50.9–63.9)

CLT = clot lysis time; HIV = human immunodeficiency virus

blood coagulation [4]. Both total and γ′ fibrinogen influence clot structure [5,6], and increased levels have been related to denser blood clots resistant to clot lysis [7]. Denser clots that remain in the vasculature for a longer time are associated with cardiovascular disease (CVD) outcomes such as myocardial infarction, stroke and coronary artery disease [8]. Increased fibrinogen concentration, irrespective of its functional properties, is also associated with CVD [9].

Family and twin studies investigating the genetics of fibrinogen have reported the heritability of fibrinogen concentrations to be between 30 and 51% [10–16]. Genome wide association studies (GWAS) have only allocated 3.7% of this possible 51% fibrinogen heritability to common single nucleotide polymorphisms (SNPs; single base pair variations occurring at frequencies of >1% in the population) to date [17]. As these GWAS included millions of SNPs without significant results, investigating pleiotropic and polygenic sequence divergences and the possible co-regulation thereof among candidate genes chosen based on association analyses of phenotypes related to fibrinogen, might prove valuable. There is growing evidence that a significant proportion of the heritability of complex phenotypes, such as fibrinogen, may be explained by a combination of genetic variants, and their combined effects can be calculated in the form of polygenic risk scores [18], making use of a biological filtering approach by taking into account the mechanistic pathways associated with the complex phenotypes. Furthermore, the combination of genetics and molecular biology has greatly facilitated the identification of candidate genes [19,20]. Multifactorial phenotypes can now be represented as complex interactive networks, which consist of a combination of genetic and non-genetic factors. Therefore, genetic variations in multiple genes in one particular pathway or disease network could lead to synergistic heterozygosity [21]. We have incorporated the concept of synergistic heterozygosity into our hypothesis, through genetic risk score (GRS) analyses, to observe the polygenic effects of harbouring several risk alleles concurrently. In addition, gene expression is generally controlled by transcriptional enhancers, which consist of a cluster of transcription factor binding sites (TFBS) spaced by spacer sequences and enhancers. SNPs in these regions have potential functional significance, which is not necessarily obvious when observed independently from a functional context [22–24]. Being in regulatory regions, these SNPs may, therefore, associate with the outcome phenotype through polygenic co-regulation rather than being in the causal pathway.

In agreement with international research [25,26], ethnic differences in fibrinogen concentrations have been reported in the South African population. Black South Africans present with higher total fibrinogen concentrations than their counterparts from European ancestry [27,28]. A pre-disposition to hypercoagulability has also been observed in the black South African population [29]. Studies on the genetic variation in African individuals have revealed vast genetic diversity in Africa [30,31]. Furthermore, research conducted in the South African Prospective Urban and Rural Epidemiology (PURE) study population revealed less linkage disequilibrium (LD; neighbouring polymorphisms that are inherited together) in the fibrinogen genes than what has been observed in European populations [32]. The higher variance in fibrinogen concentrations, together with a unique genetic diversity and low LD, presents a promising opportunity to unravel the missing heritability observed for fibrinogen.

To this end, we conducted a candidate gene association study by investigating SNPs of variables associated with fibrinogen-related phenotypes. Candidate genes to be included were further verified through an *in silico* pathway and network analysis to increase our understanding of the polygenic regulation of the fibrinogen phenotype. Polygenetic and pleiotropic co-regulation were explored *via* GRS and TFBS analyses. Protein concentration (both total and γ′ fibrinogen) and functionality (plasma clot properties) were used as outcome variables.

Results and discussion

The total study population consisted of 1677 participants included after quality control (QC). Table 1 presents the descriptive characteristics of the study population. Total and γ′ fibrinogen concentrations, together with lag time, slope, maximum absorbance and clot lysis time (CLT), were used as phenotype outcomes for all further analyses.

This study is the first exploration of the pleiotropic and polygenetic co-regulation of candidate SNPs associated with the fibrinogen protein and its functional phenotypes in a large black South African