Identification of the Functional Variant(s) that Explain the Low-Density Lipoprotein Receptor (*LDLR*) GWAS SNP rs6511720 Association with Lower LDL-C and Risk of CHD

Roaa Hani Fairoozy[1☯]*, Jon White[2☯], Jutta Palmen[1], Anastasia Z. Kalea[1], Steve E. Humphries[1☯]

1 Centre for Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Science, University College London, London, United Kingdom, 2 University College London Genetics Institute, Department of Genetics, Environment and Evolution, London, United Kingdom

☯ These authors contributed equally to this work.
* roaa.fairoozy.12@ucl.ac.uk

Abstract

Background

The Low-Density Lipoprotein Receptor (*LDLR*) SNP rs6511720 (G>T), located in intron-1 of the gene, has been identified in genome-wide association studies (GWAS) as being associated with lower plasma levels of LDL-C and a lower risk of coronary heart disease (CHD). Whether or not rs6511720 is itself functional or a marker for a functional variant elsewhere in the gene is not known.

Methods

The association of *LDLR* SNP rs6511720 with incidence of CHD and levels of LDL-C was determined by reference to CARDIoGRAM, C4D and Global lipids genetics consortium (GLGC) data. SNP annotation databases were used to identify possible SNP function and prioritization. Luciferase reporter assays in the liver cell line Huh7 were used to measure the effect of variant genotype on gene expression. Electrophoretic Mobility Shift Assays (EMSAs) were used to identify the Transcription Factors (TFs) involved in gene expression regulation.

Results

The phenotype-genotype analysis showed that the rs6511720 minor allele is associated with lower level of LDL-C [beta = -0.2209, p = 3.85 x10^-262], and lower risk of CHD [log (OR) = 0.1155, p = 1.04 x10^-7]. Rs6511720 is in complete linkage. Rs6511720 is in complete linkage disequilibrium (LD) with three intron-1 SNPs (rs141787760, rs60173709, rs57217136). Luciferase reporter assays in Huh7 cells showed that the rare alleles of both rs6511720 and rs57217136 caused a significant increase in *LDLR* expression compared to the common alleles (+29% and +24%, respectively). Multiplex Competitor-EMSAs (MC-

Abbreviations: CHD, coronary heart disease; EMSAs, electrophoretic mobility shift essays; FH, familial hypercholesterolemia; GWAS, genome-wide association study; LD, linkage disequilibrium; *LDLR*, Low-density lipoprotein receptor; LDL-C, Low-density lipoprotein cholesterol; MAF, minor allele frequency; MC-EMSA, multiplex competitor-EMSAs; RAR, retinoic acid receptor; SNP, single nucleotide polymorphism; SRE, serum response element; STAT1, signal transducer and activator of transcription 1.

EMSA) identified that the transcription factor serum response element (SRE) binds to rs6511720, while retinoic acid receptor (RAR) and signal transducer and activator of transcription 1 (STAT1) bind to rs57217136.

Conclusion

Both *LDLR* rs6511720 and rs57217136 are functional variants. Both these minor alleles create enhancer-binding protein sites for TFs and may contribute to increased *LDLR* expression, which is consequently associated with reduced LDL-C levels and 12% lower CHD risk.

Introduction

Elevated plasma lipid levels promote atherosclerosis and increase the risk of coronary heart disease (CHD). Low-density lipoprotein cholesterol (LDL-C) is taken up from the blood by the LDL-Receptor (LDL-R). *LDLR* is located on chromosome 19 at p13.1-p13.3 and it encodes a cell surface glycoprotein predominantly expressed in hepatocytes. LDL-R mediates the removal of cholesterol-carrying LDL-C particles from the blood via ApoB-100 [1–3]. The 45kb gene comprises 18 exons and 17 introns [4]. Mutation in the *LDLR* gene leads to the monogenic disorder, familial hypercholesterolemia (FH), and to date over 1,200 mutations have been reported in the *LDLR* gene that cause FH [5]. The vast majority of these mutations are located in the exonic regions, and thus affect protein structure and function, while 10% are in the intronic region (exon boundary), and these are predicted to affect correct splicing, and 2% in the promoter region, which are predicted to prevent gene transcription. A single nucleotide polymorphism (SNP) within *LDLR* exon 12, rs688 is associated with both LDL-C and CHD in a gender-independent mode [6, 7]. It also acts as a modulator of alterative exon splicing, which can lead to a shift in the reading frame and an altered gene transcript [8–11]. Non-coding SNPs in *LDLR* have also been reported to be functional, for example, in the promoter region c.-139C>G [12], c.-101T>C, c.-121T>C [13], and -49C>T [14], and rs17248720 in the intergenic region [15] are involved in regulation of gene expression and have been reported to cause FH.

In the last decade, genome-wide association studies (GWAS) have identified numerous loci that harbor common signal nucleotide polymorphisms (SNPs) which have relatively small effects on lipid traits, including at the *LDLR* locus where SNPs are associated with LDL-C levels. The majority of common variants that have been discovered in GWAS are in non-coding regions and their functional implications are unknown [16]. Interpretation of the molecular mechanisms of non-coding variants is a huge challenge because of linkage disequilibrium (LD) and the diversity of non-coding functions, including transcriptional, mRNA splicing and control of translation [17, 18]. The T allele of the *LDLR* SNP rs6511720 (G>T) [MAF = 0.10 in a European population, (1000 Genomes Project Phase 3)] has been identified as being associated with lower plasma levels of LDL-C (size effect: -0.15 to -0.26 mmol/L) and a lower risk of CAD [19–21], myocardial infarction (MI) [22] and abdominal aortic aneurysm (AAA) [23]. Between-study similarities have provided confidence that the *LDLR* SNP rs6511720 is either functional or may be a marker for a functional variant elsewhere in the gene. In addition, Talmud et al. (2013) constructed a weighted LDL-C-raising gene score of 12 common LDL-C-raising SNPs previously identified by the Global Lipids Genetics consortium, including the *LDLR* SNP rs6511720 [19] in two patient groups (FH without an identified mutation, FH with an identified mutation) and one control group. They found that the mean weighted SNP score

for both mutation-negative and mutation-positive FH patients was significantly higher than in control subjects. The difference between mutation-negative and mutation-positive also was significant. They proposed that these common LDL-C-raising SNPs explained the hypercholesterolemia phenotype in at least 80% of patients with a clinical diagnosis of FH but with no identified mutation [24], however, the functional roles of many of these SNPs are unknown.

The rs6511720 SNP is located in intron-1 of the *LDLR* gene, where *cis*-acting gene regulatory sites are commonly found [25]. *Cis*-regulatory elements physically interact with the promoter region of a gene to initiate DNA transcription [26–28]. Such SNP could be a part of an enhancer element due to a modification of the transcription factor-binding site (TFBS) that in turn recruits co-activators and chromatin regulators to facilitate the transcription of the *LDLR* gene [29]. However, the analysis of the genetic function of such variants is complex because of the LD between SNPs, which are co-inherited with a causal variant. Thus, all SNPs in LD with the functional variant may carry some or all of the associations with the trait of interest, although they have may have not any relevant function.

This study uses a range of methods to identify functional variants and their role in gene regulation. We employed SNP selection and prioritization methods based on the data from bioinformatics databases. Three SNPs (rs57217136, rs141787760 and rs60173709) in addition to GWAS hit SNP rs6511720 were selected for functional analyses. A number of functional study techniques were performed including conventional electrophoretic mobility shift assays (EMSA), reporter assay and multiplex competitor electrophoretic mobility shift assays (MC-EMSA) to identify the regulatory role of selected SNPs.

Methods

Statistical analysis

The summary estimates for the log odds Ratio (OR) and its standard error of rs6511720 on CHD were taken from CARDIoGRAM [30] and C4D [31] (http://www.cardiogramplusc4d.org/downloads/) and were combined by fixed effects meta-analysis. We obtained an estimate of the regression coefficient for rs6511720 on LDL-C from Global lipids genetics consortium (GLGC) [32] (http://csg.sph.umich.edu/abecasis/public/lipids2013/). All analyses were conducted using the statistical computing environment R version 3.2.0.

SNP annotation

Several SNP annotation databases were used to identify possible SNP functions. The UCSC genome browser [33] was used as the source of genome-wide maps of the chromatin state of the region of interest of the gene. Locuszoom [34] (http://csg.sph.umich.edu/locuszoom/), SNAP [35] (https://www.broadinstitute.org/mpg/nsnap/ldplot.php) and HaploReg V2-V4 [36, 37] (http://compbio.mit.edu/HaploReg) were used to identify LD SNPs with the *LDLR* GWAS hit SNP rs6511720. HaploReg and MatInspector [38] (Genomatix Software GmbH, Germany) were used to create the SNP profile. Three SNPs in complete LD ($r^2 = 1$) with GWAS hit SNP rs6511720 (rs57217136, rs141787760 and rs60173709) were selected for further analyses.

Functional assessment of *LDLR* intron-1 SNPs

Electrophoretic mobility shift assays (EMSAs) were used to investigate the effect of variants' genotype on DNA-protein binding. Probes of 31 bp sequences that encompassed the common or rare variant of each of the four SNPs were employed (probe sequences are available upon request). Probes were labeled using the biotin 3'-end DNA labeling kit (Pierce, Rockford, IL, USA) as recommended by the manufacturer. Each binding reaction consisted of 10μl of 2X

binding buffer (16% Ficol, 40mM HEPES, 100mM KCl, 2mM EDTA, and 1mM DTT), 0.8ng of poly(dI.dC), 50mM of $MgCl_2$, 0.8mg of BSA and 2nmol of Huh7 (a human hepatocellular carcinoma cell line) nuclear extract, corrected with dH_2O to a final volume of 20μl. The reaction mixture was incubated at 25°C for 30 min, followed by the addition of 6X loading buffer. Samples were electrophoresed on a 6% polyacrylamide gel for 180–240 min at 120V and then transferred to a nylon membrane using Southern transfer. The images were obtained using a chemiluminescent nucleic acid detection module (Pierce, Rockford, IL, USA) according to the manufacturer's instructions. Multiplex competitor electrophoretic mobility shift assay (MC-EMSAs) were carried out to identify the TFs that are involved in expression differentiation between alleles of rs6511720 and rs57217136 SNPs. The MC-EMSAs were performed using seven sets of cocktails, each with ten unlabeled dsDNA consensus sequences for well-characterized TFBS [39]. These TF cocktails were incubated with the binding reaction mix comprising Huh7 nuclear extract for 15 min and then a labeled probe was added and incubated at 25°C for 30 min. When a particular set of TFs eliminated the band shift, as a result of binding the unlabeled competitor, the individual competitor from this set was examined by an additional EMSA, in order to specify the TF that bound to the allele of the variant.

Luciferase reporter assays were performed to determine whether the four *LDLR* intron-1 SNPs influenced gene expression. The Phusion® High-Fidelity PCR Kit (New England Bio-Labs Inc) was used to amplify DNA fragments of interest including the *LDLR* promoter (594bp), and the *LDLR* SNPs rs6511720 (883bp), rs57217136 (814bp), rs141787760, as well as rs60173709 (643bp), as recommended by the manufacturer (primer sequence available upon request).

For cloning, the In-Fusion® HD Cloning Kit (Clontech Laboratories, Inc.) was used following the manufacturer's instructions. The *LDLR* promoter DNA fragment was inserted into a pGL3-basic luciferase reporter vector to generate a LDLR-luciferase-construct (promoter only) reporter plasmid. The insertion was upstream of the Luciferase gene (*luc*+) at the promoter site. Then, the *LDLR* intron-1 SNP sequences encompassing the SNP allele were individually inserted into the enhancer site of the LDLR-luciferase-construct after the SV40 polyadenylation signal. These constructs were then transformed in *E-coli* cells.

Using a quikChange Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA, USA), the rs6511720 G>T variant was created at position 79, rs141787760 (C>-) and rs60173709 (T>-) deletions were generated individually at positions 41 and 247, respectively, and rs57217136 C>T mutation was created at position 401. Both original and mutated LDLR-luciferase-constructs were transfected into Huh7 cells along with the pRL-TK plasmid as a renilla luciferase control reporter vector. When the Huh7 cells were 80% confluent, they were plated into a 96-well plate ($2x10^4$ cells/well) and incubated for 24 hours. Transfection was carried out using Opti-MEM® reduced serum medium and lipofectamine 2000 (Invitrogen). Luciferase activity was measured using the dual Luciferase Reporter Assay System kit (Promega), following the manufacturer's instructions. The mean relative expression difference between variants' genotype was determined by a two-sample *t*-test.

Results

Association of rs6511720 genotype with LD-C and CHD

The meta-analysis of the CARDIoGRAM and the C4D data showed that the rs6511720 minor allele was associated with lower risk of CHD [log (OR) = −0.1155; SE = 0.0217; p = 1.04 x10^{-7}] (S1 Table). The GLGC (N = 170,607) showed that rs6511720 minor allele is associated with lower levels of LDL-C [beta = −0.2209; SE = 0.0061; p = 3.85 x10^{-262}] (S2 Table). This suggests

that individuals carrying one or more copies of the minor LDL-C lowering T allele of rs6511720 would be at lower lifetime risk of CHD.

SNPs annotation and prioritization

To identify potentially functional SNPs at the *LDLR* locus, we considered variants with strong LD ($r^2 = 0.8$) with the *LDLR* GWAS hit SNP rs6511720. This SNP is in strong LD with 48 other SNPs: three SNPs are located in intron-1 and the others are located ≥1.5kb upstream of *LDLR* locus (see LD plot, S1 Fig). To further prioritize variants for a functional analysis follow-up, genome-wide maps of the chromatin state of relevant cell types (HepG2, human liver hepatocytes) were examined. Variant position was evaluated for evidence of the histone modification mark H3K4me1, H3K27Ac and H3K4me3, as well as for DNase I hypersensitivity sites and for formaldehyde assisted isolation of regulatory elements (FAIRE). The post-translational chromatin mark H3K4me1, and H3K27Ac are often associated with enhancer regions [40], while H3K4me3 associated with promoter regions [41, 42]. DNase I and FAIRE are established methods for the identification of nucleosome regulatory regions [43] (Fig 1). In addition, TFs occupancy in chromatin was assessed using genome wide ChIP-seq data sets (UCSC genome browser accessed 01-03-2016). Of the 48 variants meeting the LD threshold ($r^2<0.8$), four SNPs were found to have strong chromatin signals and interesting TF binding profiles (S1 Table) including three SNPs in complete LD ($r^2 = 1$) with rs6511720 (rs57217136 (T>C), rs141787760 (C>deletion) and rs60173709 (T>deletion). The four selected SNPs are located in intron-1 of *LDLR* within ≈ 1200bp of each other.

In-silico tools, MatInspector and HaploReg, showed different DNA-protein binding profiles of these SNPs (S1 Table). The MatInspector software identified a sequence around the minor allele of rs6511720 that was predicted to bind to GATA and the snRNA activating protein/ proximal sequence element (SNAP/PSE) complex, but HaploReg did not predict any protein binding around this SNP. SNAP/PSE has a role in gene transcription initiation [44–46]. MatInspector did not find any binding site for either rs141787760 and rs60173709, while HaploReg software predicted that the minor allele of rs141787760 would bind to TFs that have roles in chromatin modulation through general transcription regulation and a known *LDLR* transcription regulator SP1 (specificity protein 1). Rs60173709 was also predicted as a binding
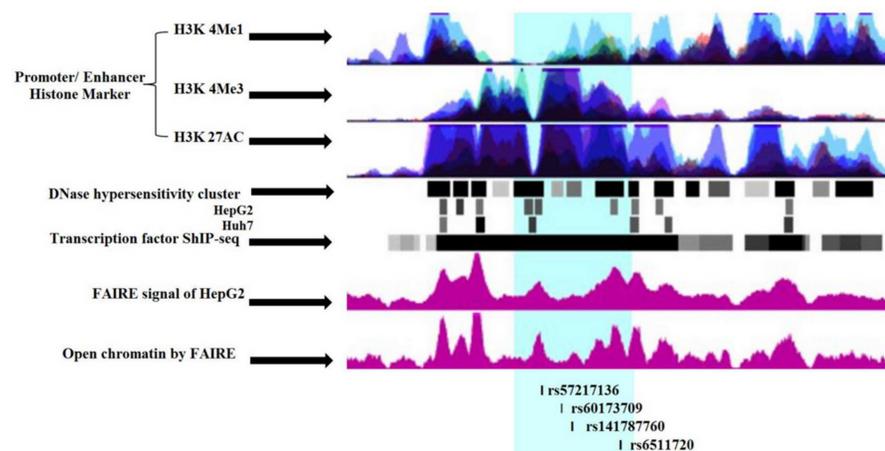

Fig 1. Genome-wide maps of chromatin state of LDLR intron-1. Schematic presentation of the LDLR intron-1 chromatin status (https://genome-euro.ucsc.edu). The area of interest in intron-1 is highlighted in light-blue color. Promoter/ Enhancer histone marker of seven cell lines (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, and NHLF). FAIRE: formaldehyde assisted isolation of regulatory elements.
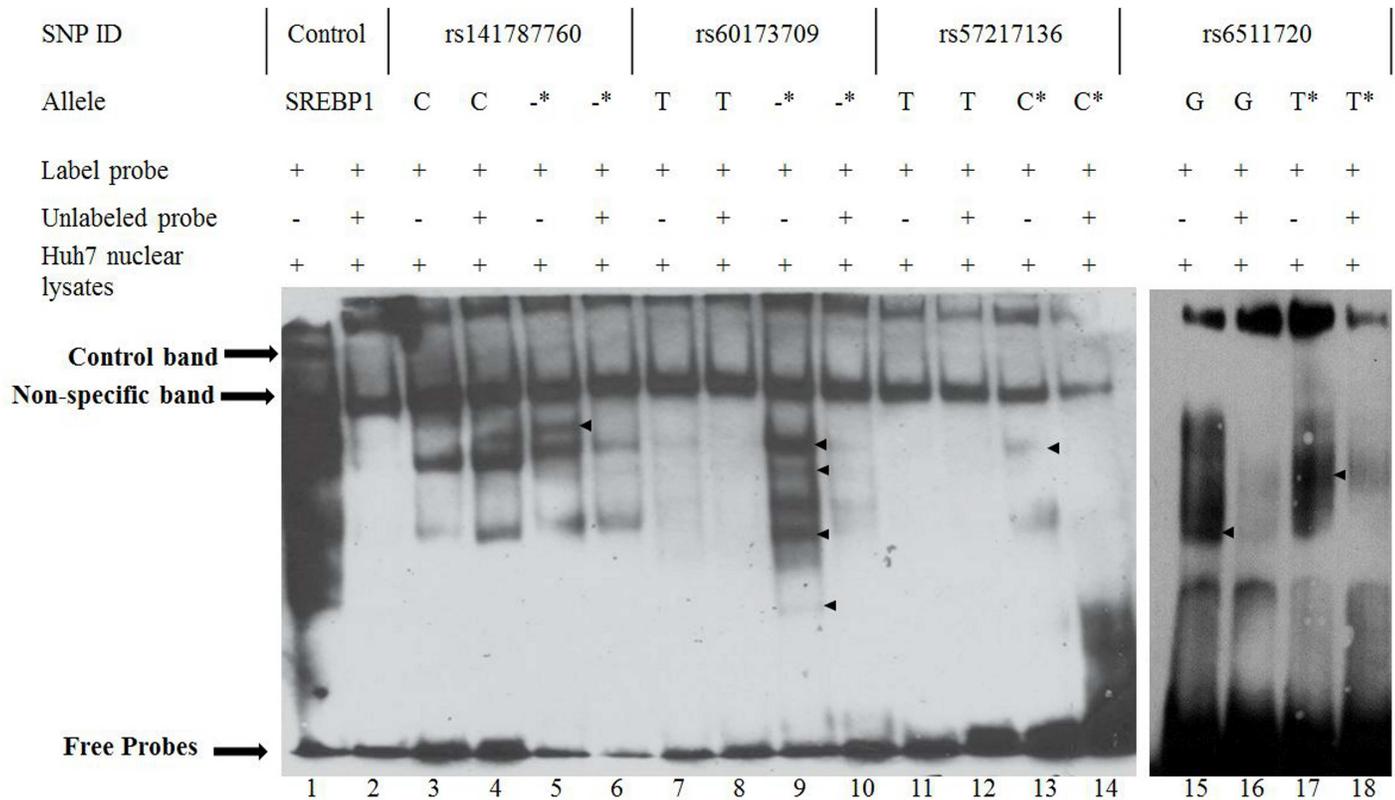
doi:10.1371/journal.pone.0167676.g001

Fig 2. DNA binding properties of *LDLR* intron-1 SNPs. Conventional EMSA analysis of the *LDLR* intron-1 SNPs (rs6511720, rs141687760, rs60173709, and rs57217136). Binding of SREBP1 was used as the control (lane 1 and 2). The lanes with a labeled probe showed a specific band indicated by arrows, while when the unlabeled probe was added the band disappeared. These four SNPs have allele-specific binding, indicated by arrows. (-) = deletion and (*) = minor allele.

site for some proteins, but not any that have a clear role in *LDLR* gene regulation. For rs57217136 HaploReg predicted that the major allele would bind to TFs such as forkhead box protein A (FOXA1, FOXA2), which are liver transcription activators, and sterol regulatory element-binding protein 1 (SREBP1), while the minor allele would bind to SP1.

Allele-specific protein binding of *LDLR* intron-1 SNPs in Huh7 cells

The *in-silico* analysis suggested that the *LDLR* intron-1 selected SNPs are sites for TF binding. To assess whether the alleles of the SNPs differentially affect protein-DNA binding *in vitro*, EMSAs for the four intron-1 SNPs were carried out using Huh7 nuclear lysates. As shown in Fig 2, all four SNPs demonstrated allele-specific binding. The rare alleles of rs57217136, rs141787760 and rs60173709 demonstrated a new DNA-protein complex, while the rs6511720 T allele formed a different protein complex, which moved more slowly than the major G allele complex.

Allele-specific enhancer activity of *LDLR* intron-1 SNPs in Huh7 cells

Luciferase reporter assays were carried out to determine the influence of the four SNPs on the transcriptional activity of the *LDLR* promoter. Generating a fragment comprising all four SNPs was not possible because intron-1 of the *LDLR* has 20 Alu repeats [47]. Therefore, three fragments were generated to exclude Alu repeats; the first fragment contained rs6511720, the
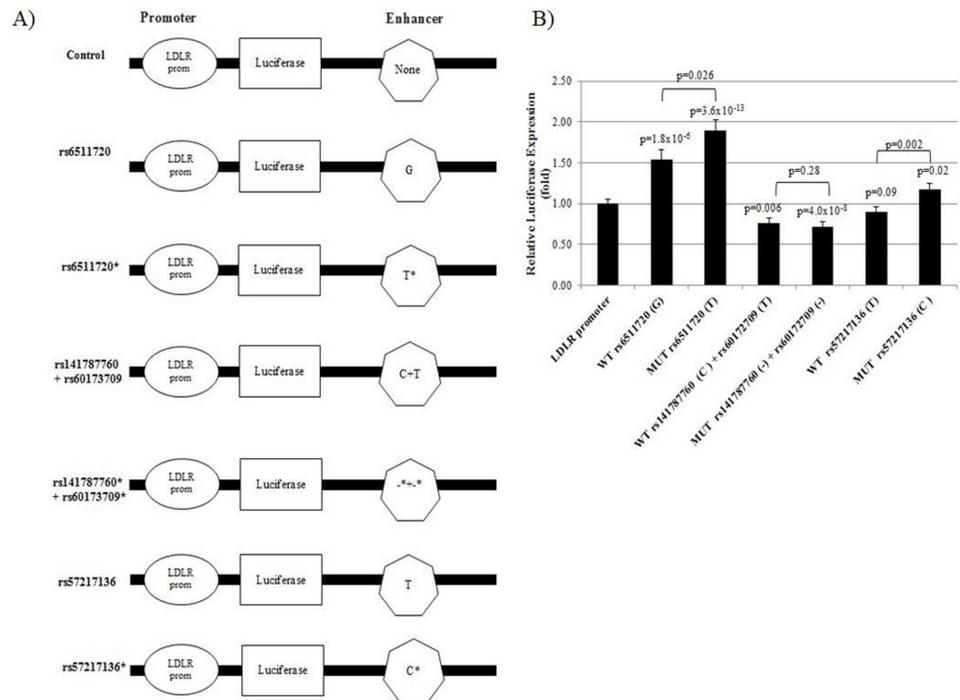
Fig 3. *LDLR* luciferase constructs and SNP luciferase activity in Huh7 cell line. A) Schematic presentation of LDLR-luciferase-construct (promoter only) and LDLR-luciferase-enhancer-constructs. The constructs were transfected into Huh7 cells. B) Results of luciferase reporter assays showing relative expression of LDLR-luciferase-enhancer constructs of *LDLR* SNPs relative to the LDLR-luciferase (no enhancer) construct. (-) = deletion and (*) = minor allele.

doi:10.1371/journal.pone.0167676.g003

second fragment rs141787760/rs60173709, and the third fragment rs57217136. These fragments were inserted individually into the enhancer site of the LDLR-luciferase-construct and transfected into Huh7 cells (Fig 3A). Luciferase activity measurements showed that the rs141787760/rs60173709 transfected construct led to a significant reduction in *LDLR* promoter expression when compared to the promoter alone: -24% (p = 0.006) for the major allele and -29% (p = 4.0×10^{-8}) for the minor allele, but with no significant difference between the alleles. In contrast, the rare allele of rs6511720 functioned as an enhancer, whilst the rare alleles of both rs6511720 and rs57217136 increased *LDLR* promoter expression activity significantly by +29% and +24% (p = 0.026, and p = 0.002), respectively, when compared to the common alleles (Fig 3B).

Identification of allele-specific transcription factor binding

MC-EMSAs were carried out to identify the specific TFs that bind to rs6511720 and rs57217136 variants. The results showed that the serum response element (SRE) is bound to a sequence around the protective allele (T) of rs6511720 (S2B Fig), while the retinoic acid receptor (RAR) and signal transducer and activator of transcription-1 (STAT1) are bound to a sequence around the C allele of rs57217136 (S2D Fig).

Discussion

Several GWAS have reported that the *LDLR* SNP rs6511720 (G>T) is associated with lower plasma levels of LDL-C and a lower risk of CHD [19–21]. Data from the GLGC consortium

showed that The *LDLR* rs6511720 minor (T, Forward strand) allele is carried by approximately 10% of the population and have confirmed the allele is protective, being associated with lower levels of LDL-C. This association is consistent and of sufficient strength to suggest that common variation in *LDLR* has implications for health, and that determining the precise molecular mechanism of the effect is relevant. The lead hit SNP has strong LD with 48 SNPs, thus to prioritize candidate variants, genome-wide maps of regulatory elements were used, which are useful resources to identify variants differentially affecting transcriptional activity. We found four promoter-proximal intron-1 variants (rs6511720, rs57217136, rs141787760 and rs60173709), which had strong chromatin signals in liver cells: these SNPs are in complete LD (r^2 = 1) with the GWAS hit SNP. *In-silico* tools, MatInspector and HaploReg, showed different DNA-protein binding profiles of these SNPs, The *LDLR* SNP rs6511720 was predicted to be a strong enhancer, and the other three SNPs were predicted to be promoter-activators [48]. Although, the inconsistent prediction between *in-silico* tools prevented finding an interesting TF to test, the data shed light on candidate SNPs that have a strong regulatory profile, and which were worthy of further investigation using functional assays.

The *in-silico* findings were confirmed by EMSAs, where the four SNPs all showed allele-specific protein binding to the rare alleles. This finding suggested that these proteins may up-regulating *LDLR* expression. To determine whether the different genotypes have an influence on *LDLR* expression, gene reporter assays were used. Data from the luciferase reporter assays showed that *LDLR* intron-1 SNPs are indeed affecting *LDLR* gene expression, as the minor alleles of rs6511720 and rs57217136 showed significant higher expression, while rs141787760 and rs60173709 showed significant lower expression but no difference between the alleles. It is important to consider the combined effect on expression of these four SNPs, given that the rare alleles of all four are always present together. It is likely that the lowering effect of the minor alleles of rs141787760 and rs60173709 partially repress the raising effect of the rare alleles of rs6511720 and rs57217136 SNPs. If the combined effect of the minor alleles of four *LDLR* intron-1 SNPs was estimated by summation, we would predict that the haplotype will show ~ 29% higher expression. This would lead to up-regulation of LDL-R numbers on the surface of hepatocytes and this in turn would explain the lower levels of LDL-C in individuals carrying the rs6511720 minor allele haplotype.

The actual mechanism of this effect is only partially unraveled by our work. Both minor alleles of SNPs rs6511720 and rs57217136 are predicted to create enhancer-binding protein sites for TFs that would contribute to increased *LDLR* expression. MC-EMSAs showed rs6511720 minor allele (T) was bound to the serum response element (SRE) transcription factor. The SRE contains a binding site for serum response factors (SRF), which have a role in *LDLR* gene expression stimulation [49]. It also showed that rs57217136 was bound to RAR and STAT1. The Retinoic Acid receptor is a member of a family of nuclear receptor proteins actively involved in retinoic acid mediated transcriptional regulation of genes that controls lipid metabolism, through dimerization with other proteins to initiate transcription [50]. STAT1 is involved in lipid metabolism via JAK/STATs through several pathways. A phosphorylated STAT1 is inter-nuclear and binds to *sis*-inducible element (SIE) sequence in the promoter region and regulates gene expression [51, 52].

Limitations

Our results suggest that a cis-regulatory element near rs6511720 and rs57217136 SNPs acts in the liver cell line. However, in our study we did not examine other LD distal SNPs that may have a role in transcription regulation and therefore further studies are needed in this direction. In this study, we used luciferase reporter assays and EMSAs, which are techniques used

to measure the difference of allelic expression and to determine DNA-protein interaction *in vitro*. An *in vitro* study may not fully represent what is occurring *in vivo*, where open chromatin structure and epigenetics have a potential role in gene regulation, and where transfection of a small fragment of DNA into a cell line cannot accurately reflect the natural situation. Finally, since we studied the effect of these SNPs only in liver cells we cannot determine whether or not they also influence LDLR gene expression in other tissues. However since the major site of expression of the LDLR is the liver, where clearance of LDL-C from the plasma occurs, this is not a major limitation.

In conclusion, integration of bioinformatics with GWAS disease-associated variants helps to elucidate gene-regulatory variants underlying association signals. Both rs6511720 and rs57217136 were identified as part of a cis-regulatory complex in a liver cell line that altered transcriptional activity through binding SRE, RAR, and STAT1. However, more studies are needed to define the spatial organization of the gene, which has a fundamental role in controlling gene expression [53–56], by the chromatin looping that brings enhancers and promoters into close spatial proximity to interact and initiate transcription. To identify the interaction between the functional SNPs and the *LDLR* promoter, chromosome conformation capture (3C) would be a useful technique [57, 58] or chromosome conformation capture carbon copy (5C) [59].

Supporting Information

S1 Fig. *LDLR* rs6511720 LD plot. A LD plot was generated using Locuszoom (http://csg.sph.umich.edu/locuszoom/). SNPs are plotted with the meta-analysis p value of LDL-C association (as–$\log_{10}$ values) as a function of genomic position. The lead SNP (rs6511720) is represented by a diamond, while LD SNPs are represented by circles. The LD SNPs are color coded to represent the r-squared between SNP and the putative associated variant, where red indicates a strong LD $r^2 \geq 0.8$ and dark blue indicates a weak LD $r^2 \leq 0.2$. A blue line indicates estimated recombination rates and dark blue arrows indicate gene annotations. LD and recombination rates are based on HapMap Phase II (CEU, YRI and JPT+CHB) or 1000 Genomes (CEU) and gene information from the UCSC browser.
(TIF)

S2 Fig. DNA binding and expression of the transcription factors of the *LDLR* SNPs: rs6511720 and rs57217136. MC-EMSA analysis. Nuclear proteins from the Huh7 cell line were incubated with 7 cocktails of unlabelled DNA competitors (70 well-characterized DNA-binding proteins) for 15 minutes, then a 5' end-biotinylated allele-specific probe was added. The multiplex competitors compete out any specific interactions with a labeled probe, eliminating or reducing any positive shift result. A) *LDLR* rs6511720 MC-EMSA for both alleles of the SNP, T allele (rare) specific bands were eliminated by cocktail 4. B) The single competitors from cocktail 4 (a) were run individually in a further EMSA, showing SRE resulted in competition. C) *LDLR* rs57217136 MC-EMSA for C allele, the C allele (rare) specific bands were eliminated by cocktail 4. D) The single competitors of the cocktail 4 (c) were run individually in a further EMSA, showing competitors RAR and STAT1 resulted in competition.
(TIF)

S1 Table. Association of rs6511720 genotype and CHD risk in CARDIoGRAM and C4D.
(PDF)

S2 Table. Association of rs6511620 genotype and lipid treats from data in Global lipids genetics consortium (GLGC).
(PDF)

S3 Table. Predicted regulatory element and protein binding of LDLR selected SNPs. * minor allele.
(PDF)

Author Contributions

Conceptualization: SEH RHF.

Data curation: SEH.

Formal analysis: RHF JW.

Funding acquisition: SEH RHF.

Investigation: RHF JP.

Methodology: RHF.

Project administration: SEH.

Resources: RHF SEH.

Software: RHF JW.

Supervision: SEH.

Validation: SEH.

Visualization: SEH RHF.

Writing – original draft: RHF.

Writing – review & editing: SEH AZK.

References

1. Goldstein JL, Brown MS. Binding and Degradation of Low Density Lipoproteins by Cultured Human Fibroblasts: COMPARISON OF CELLS FROM A NORMAL SUBJECT AND FROM A PATIENT WITH HOMOZYGOUS FAMILIAL HYPERCHOLESTEROLEMIA. Journal of Biological Chemistry. 1974; 249 (16):5153–62. PMID: 4368448

2. Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG. Hyperlipidemia in Coronary Heart Disease II. GENETIC ANALYSIS OF LIPID LEVELS IN 176 FAMILIES AND DELINEATION OF A NEW INHERITED DISORDER, COMBINED HYPERLIPIDEMIA. The Journal of Clinical Investigation. 1973; 52(7):1544–68. doi: 10.1172/JCI107332 PMID: 4718953

3. Brown MS, Goldstein JL. Receptor-mediated endocytosis: insights from the lipoprotein receptor system. Proceedings of the National Academy of Sciences of the United States of America. 1979; 76 (7):3330–7. PMID: 226968

4. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Human Mutation. 1992; 1(6):445–66. doi: 10.1002/humu.1380010602 PMID: 1301956

5. Usifo E, Leigh SEA, Whittall RA, Lench N, Taylor A, Yeats C, et al. Low-Density Lipoprotein Receptor Gene Familial Hypercholesterolemia Variant Database: Update and Pathological Assessment. Annals of Human Genetics. 2012; 76(5):387–401. doi: 10.1111/j.1469-1809.2012.00724.x PMID: 22881376

6. Boright AP, Connelly PW, Brunt JH, Morgan K, Hegele RA. Association and linkage of LDLR gene variation with variation in plasma low density lipoprotein cholesterol. J Hum Genet. 1998; 43(3):153–9. doi: 10.1007/s100380050060 PMID: 9747026

7. Martinelli N, Girelli D, Lunghi B, Pinotti M, Marchetti G, Malerba G, et al. Polymorphisms at LDLR locus may be associated with coronary artery disease through modulation of coagulation factor VIII activity and independently from lipid profile. Blood. 2010; 116(25):5688–97. doi: 10.1182/blood-2010-03-277079 PMID: 20810930

8. Zhu H, Tucker HM, Grear KE, Simpson JF, Manning AK, Cupples LA, et al. A Common Polymorphism Decreases Low-Density Lipoprotein Receptor Exon 12 Splicing Efficiency and Associates with

Increased Cholesterol. Human Molecular Genetics. 2007; 16(14):1765–72. doi: 10.1093/hmg/ddm124 PMID: 17517690

9. Zou F, Gopalraj RK, Lok J, Zhu H, Ling IF, Simpson JF, et al. Sex-dependent Association of a Common Low Density Lipoprotein Receptor Polymorphism with RNA Splicing Efficiency in the Brain and Alzheimers Disease. Human Molecular Genetics. 2008; 17(7):929–35. doi: 10.1093/hmg/ddm365 PMID: 18065781

10. Gao F, Ihn HE, Medina MW, Krauss RM. A common polymorphism in the LDL receptor gene has multiple effects on LDL receptor function. Human Molecular Genetics. 2013; 22(7):1424–31. doi: 10.1093/hmg/dds559 PMID: 23297366

11. Lee J-D, Hsiao K-M, Wang T-C, Lee T-H, Kuo Y-W, Huang Y-C, et al. Mutual effect of rs688 and rs5925 in regulating low-density lipoprotein receptor splicing. DNA and cell biology. 2014; 33(12):869–75. doi: 10.1089/dna.2014.2577 PMID: 25188588

12. Smith AJP, Ahmed F, Nair D, Whittall R, Wang D, Taylor A, et al. A functional mutation in the LDLR promoter (-139C>G) in a patient with familial hypercholesterolemia. Eur J Hum Genet. 2007; 15(11):1186–9. doi: 10.1038/sj.ejhg.5201897 PMID: 17625505

13. Khamis A, Palmen J, Lench N, Taylor A, Badmus E, Leigh S, et al. Functional analysis of four LDLR 5′UTR and promoter variants in patients with familial hypercholesterolaemia. European Journal of Human Genetics. 2015; 23(6):790–5. doi: 10.1038/ejhg.2014.199 PMID: 25248394

14. Mozas P, Galetto R, Albajar M, Ros E, Pocoví M, Rodríguez-Rey JC. A mutation (−49C>T) in the promoter of the low density lipoprotein receptor gene associated with familial hypercholesterolemia. Journal of lipid research. 2002; 43(1):13–8. PMID: 11792717

15. De Castro-Oros I, Perez-Lopez J, Mateo-Gallego R, Rebollar S, Ledesma M, Leon M, et al. A genetic variant in the LDLR promoter is responsible for part of the LDL-cholesterol variability in primary hypercholesterolemia. BMC Medical Genomics. 2014; 7(1):17.

16. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337(6099):1190–5. doi: 10.1126/science.1222794 PMID: 22955828

17. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008; 9 (5):356–69. doi: 10.1038/nrg2344 PMID: 18398418

18. Donnelly P. Progress and challenges in genome-wide association studies in humans. Nature. 2008; 456(7223):728–31. doi: 10.1038/nature07631 PMID: 19079049

19. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466(7307):707–13. http://www.nature.com/nature/journal/v466/n7307/abs/nature09270.html#supplementary-information. doi: 10.1038/nature09270 PMID: 20686565

20. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008; 40(2):189–97. http://www.nature.com/ng/journal/v40/n2/suppinfo/ng.75_S1.html. doi: 10.1038/ng.75 PMID: 18193044

21. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nature genetics. 2008; 41 (1):47–55. doi: 10.1038/ng.269 PMID: 19060911

22. Anand SS, Xie C, Paré G, Montpetit A, Rangarajan S, McQueen MJ, et al. Genetic variants associated with myocardial infarction risk factors in over 8000 individuals from five ethnic groups: The INTERHEART Genetics Study. Circulation Cardiovascular genetics. 2009; 2(1):16–25. doi: 10.1161/CIRCGENETICS.108.813709 PMID: 20031563

23. Bradley DT, Hughes AE, Badger SA, Jones GT, Harrison SC, Wright BJ, et al. A variant in LDLR is associated with abdominal aortic aneurysm. Circulation: Cardiovascular Genetics. 2013; 6(5):498–504.

24. Talmud PJ, Shah S, Whittall R, Futema M, Howard P, Cooper JA, et al. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. The Lancet. 2013; 381(9874):1293–301. http://dx.doi.org/10.1016/S0140-6736(12)62127-8.

25. Rose A. Intron-mediated regulation of gene expression. Nuclear pre-mRNA processing in plants: Springer; 2008. p. 277–90.

26. Schleif R. DNA looping. Annual Review of Biochemistry. 1992; 61(1):199–223.

27. Calhoun VC, Stathopoulos A, Levine M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proceedings of the National Academy of Sciences. 2002; 99(14):9243–7.

28. Su W, Jackson S, Tjian R, Echols H. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes & development. 1991; 5(5):820–6.

29. Kadonaga JT. Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. Cell. 2004; 116(2):247–57. http://dx.doi.org/10.1016/S0092-8674(03)01078-X. PMID: 14744435

30. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analyses identifies 13 new susceptibility loci for coronary artery disease. Nature genetics. 2011; 43 (4):333–8. doi: 10.1038/ng.784 PMID: 21378990

31. TheCoronaryArteryDisease(C4D)GeneticsConsortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet. 2011; 43(4):339–44. http://www.nature.com/ng/journal/v43/n4/abs/ng.782.html#supplementary-information. doi: 10.1038/ng.782 PMID: 21378988

32. Global Lipids Genetics C. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013; 45(11):1274–83. http://www.nature.com/ng/journal/v45/n11/abs/ng.2797.html#supplementary-information. doi: 10.1038/ng.2797 PMID: 24097068

33. Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. UCSC genome browser tutorial. Genomics. 2008; 92(2):75–84. http://dx.doi.org/10.1016/j.ygeno.2008.02.003. PMID: 18514479

34. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26(18):2336–7. doi: 10.1093/bioinformatics/btq419 PMID: 20634204

35. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24(24):2938–9. doi: 10.1093/bioinformatics/btn564 PMID: 18974171

36. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012; 40(D1):D930–D4.

37. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic acids research. 2015.

38. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 2005; 21(13):2933–42. doi: 10.1093/bioinformatics/bti473 PMID: 15860560

39. Smith AJP, Humphries SE. Characterization of DNA-Binding Proteins Using Multiplexed Competitor EMSA. Journal of Molecular Biology. 2009; 385(3):714–7. http://dx.doi.org/10.1016/j.jmb.2008.11.035. PMID: 19059416

40. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107(50):21931–6. doi: 10.1073/pnas.1016071107 PMID: 21106759

41. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007; 39 (3):311–8. http://www.nature.com/ng/journal/v39/n3/suppinfo/ng1966_S1.html. doi: 10.1038/ng1966 PMID: 17277777

42. Calo E, Wysocka J. Modification of enhancer chromatin: what, how and why? Molecular cell. 2013; 49 (5).

43. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee B-K, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome research. 2011; 21 (10):1757–67. doi: 10.1101/gr.121541.111 PMID: 21750106

44. Mittal V, Ma B, Hernandez N. SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain. Genes & development. 1999; 13(14):1807–21.

45. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. Genes & development. 1996; 10(21):2657.

46. Sadowski CL, Henry RW, Kobayashi R, Hernandez N. The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein. Proceedings of the National Academy of Sciences of the United States of America. 1996; 93(9):4289–93. PMID: 8633057

47. Amsellem S, Briffaut D, Carrié A, Rabès J, Girardet J, Fredenrich A, et al. Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. Hum Genet. 2002; 111(6):501–10. doi: 10.1007/s00439-002-0813-4 PMID: 12436241

48. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473(7345):43–9.

http://www.nature.com/nature/journal/v473/n7345/abs/10.1038-nature09906-unlocked. html#supplementary-information. doi: 10.1038/nature09906 PMID: 21441907

49. Pak YK. Serum response element-like sequences of the human low density lipoprotein receptor promoter: possible regulation sites for sterol-independent transcriptional activation. Biochem Mol Biol Int. 1996; 38(1):31–6. PMID: 8932516

50. Chawla A, Repa JJ, Evans RM, Mangelsdorf DJ. Nuclear receptors and lipid physiology: opening the X-files. Science. 2001; 294(5548):1866–70. doi: 10.1126/science.294.5548.1866 PMID: 11729302

51. Planas AM, Berruezo M, Justicia C, Barrón S, Ferrer I. Stat3 Is Present in the Developing and Adult Rat Cerebellum and Participates in the Formation of Transcription Complexes Binding DNA at the sis-Inducible Element. Journal of Neurochemistry. 1997; 68(4):1345–51. PMID: 9084404

52. Qin J-Z, Kamarashev J, Zhang C-L, Dummer R, Burg G, Dobbeling U. Constitutive and Interleukin-7- and Interleukin-15-Stimulated DNA Binding of STAT and Novel Factors in Cutaneous T Cell Lymphoma Cells. 2001; 117(3):583–9.

53. Müller H-P, Sogo J, Schaffner W. An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. Cell. 1989; 58(4):767–77. PMID: 2548735

54. Ptashne M. Gene regulation by proteins acting nearby and at a distance. Nature. 1985; 322(6081):697–701.

55. Ptashne M. How eukaryotic transcriptional activators work. Nature. 1988; 335(6192):683–9. doi: 10.1038/335683a0 PMID: 3050531

56. Ptashne M, Gann A. Transcriptional activation by recruitment. Nature. 1997; 386(6625):569–77. doi: 10.1038/386569a0 PMID: 9121580

57. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002; 295 (5558):1306–11. doi: 10.1126/science.1067799 PMID: 11847345

58. Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nature protocols. 2007; 2(7):1722–33. doi: 10.1038/nprot.2007.243 PMID: 17641637

59. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. Genome research. 2006; 16(10):1299–309. doi: 10.1101/gr.5571506 PMID: 16954542