

ARTICLE OPEN



Novel start codon variant in the 5'UTR of *LDLR* associated with familial hypercholesterolaemia

Martin Bird¹✉, Chris Jyun-Peng Tung², Alan M. Pittman¹, Elijah R. Behr¹, Axel Nohturfft² and Marta Futema^{1,3}✉

© The Author(s) 2025

Familial hypercholesterolaemia (FH) is a genetic disorder due to pathogenic variants in *LDLR*, *APOB*, and *PCSK9* genes, characterised by elevated low-density lipoprotein cholesterol (LDL-C) concentration and a significantly increased risk of premature coronary heart disease. Annotating whole genome sequencing data of 536 FH patients using the VEP plugin UTRannotator, we identified a novel variant c.-35C > G in the 5' untranslated region (5'UTR) of *LDLR*, predicted to introduce an upstream translation initiation codon and upstream open reading frame (uORF) that is out of frame with the *LDLR* coding sequence. Using promoter and epitope reporter assays, we demonstrate that the c.-35C > G variant leads to the preferential utilisation of the upstream AUG codon over the wild-type *LDLR* translation start site. We additionally conducted reporter assays for a previously reported variant that introduces a novel AUG codon through a deletion at position -22 of the 5'UTR (c.-22del) and obtained similar results. These findings confirm a novel type of FH-causing *LDLR* variants, leading to a premature start of translation and a truncation, underscoring the need for expanded genetic screening beyond coding regions. Future studies should focus on further characterising 5'UTR variants to better understand their role in FH.

European Journal of Human Genetics; <https://doi.org/10.1038/s41431-025-01893-y>

INTRODUCTION

Familial hypercholesterolaemia (FH) is a genetic disorder marked by elevated levels of low-density lipoprotein cholesterol (LDL-C), which increases the risk of developing premature coronary heart disease (CHD) [1]. Clinical diagnosis of possible FH is widely based on the well-established Simon Broome and Dutch Lipid Clinic Network criteria [2]. Where probands meeting the clinical FH criteria were followed up by sequencing, most of the identifiable pathogenic variants have been found in the *LDLR*, *APOB*, and *PCSK9* genes [1, 3–5]. Defects in the proteins encoded by these genes result in impaired clearance of circulating LDL particles, leading to a lifelong burden of increased cholesterol in affected individuals [6]. More recently, a specific variant in the *APOE* gene (p.Leu167del) has also been linked to FH [7, 8]. Additionally, a recessive form of hypercholesterolaemia can arise from homozygous or compound heterozygous pathogenic variants in *LDLRAP1* [9].

Variants in the protein-coding region of the *LDLR* gene make up the great majority (> 80%) of genetic changes that are known to cause FH [10, 11]. Moreover, FH-associated variants have been identified in the promoter region of *LDLR* that are predicted to inhibit the binding of important transcription factors such as sterol regulatory element-binding proteins (SREBPs), specificity protein 1 (Sp1), cAMP responsive element-binding protein (CREB), and TATA box binding protein. In functional studies, these variants showed 3–72% wild-type promoter activity [12].

Currently, there is limited functional evidence supporting the pathogenicity of 5'UTR variants. To date, only 2 *LDLR* 5'UTR (region

c.-86 to c.-1) variants have been functionally studied (c.-13A > G, c.-36T > G), both showing no effect on promoter activity [11–13]. These variants are likely benign because they do not create new upstream start sites (uAUG) or upstream open reading frames (uORF). To our knowledge, only one instance of a uORF creating variant in *LDLR* has been described in a Turkish male proband with FH who was found to be compound heterozygous for a c.-22del variant resulting in an uAUG [14]. However, no functional analyses of this variant have been reported.

The prevalence of FH-causing variants in known FH genes is estimated to be around 1 in 250 individuals (95% CI 1:345–1:192) across diverse populations, including those of South Asian and African ancestry [15, 16]. Early diagnosis and proactive lipid-lowering treatments are crucial for minimising the likelihood of cardiovascular events in this group [17]. Identifying the pathogenic variant in individuals with FH is crucial not only for guiding targeted treatment but also for effective cascade testing in family members, thereby enabling early detection of those at risk. However, in mutational analyses at least half of probands with clinical FH remained without a genetic diagnosis [1, 3–5], which highlights the importance of exploring non-coding regions and other genetic contributors. Here, using whole genome sequencing (WGS) data we apply in silico annotation tools to look specifically for 5'UTR variants, followed by functional testing. In our analyses of promoter activity, we included the previously published variant c.-22del which creates a novel uAUG, to provide novel insights into the regulatory mechanisms that contribute to FH and the role of non-coding variants in the disease.

¹Cardiovascular and Genomics Research Institute, School of Health & Medical Sciences, City St George's, University of London, London, UK. ²Neuroscience and Cell Biology Research Institute, School of Health & Medical Sciences, City St George's, University of London, London, UK. ³Institute of Cardiovascular Science, Faculty of Population Health, University College London, London, UK. ✉email: mbird@sgul.ac.uk; mfutema@sgul.ac.uk

Received: 5 February 2025 Revised: 17 April 2025 Accepted: 2 June 2025

Published online: 24 July 2025

METHODS

100,000 Genomes project familial hypercholesterolaemia cohort

The current study (project ID: RR123) has been approved by the Genomics England Clinical Interpretation Partnership (GeCIP) cardiovascular domain committee [18]. The study cohort comprised 467 probands and 69 affected relatives, classified as having Possible or Definite FH based on the Simon Broome diagnostic criteria, as per National Institute for Health and Care Excellence (NICE) recommendations [2]. Inclusion criteria were LDL-C levels exceeding 4.9 mmol/L and a family history of myocardial infarction or severe hypercholesterolaemia. Participants were recruited through the Rare Diseases: Cardiovascular domain of the 100,000 Genomes Project (100KGP). This study follows our previous bioinformatic analyses of the FH 100KGP whole genomes to provide new knowledge on a variant of uncertain significance (VUS) [19].

In silico 5'UTR variant annotation

UTR variant annotation utilised ensembl Variant Effect Predictor (VEP) [20] v99 in conjunction with UTRannotator [21]. Variants which were not present or had a minor allele frequency (MAF) ≤ 0.0001 in GnomAD v4.1.0 were retained. The variants that were selected, were further classified to one of the following types of high impact 5'UTR variants: uAUG gain, uAUG loss, upstream stop site (uSTOP) loss, and upstream frameshift (uFrame-shift). Variants identified after this filtering process were then searched for in the rest of the 100KGP WGS participants to verify if they were unique to the 100KGP FH cohort.

Plasmids

Reagents are listed in the Supplementary Methods S1. Plasmids pRL-CMV (expressing *Renilla reniformis* luciferase), pGL4.32 and pGL4.36 were purchased from Promega and plasmid pCMV- β Gal from Stratagene. Plasmid pANG32 expresses the Luc2P reporter gene (*Photinus pyralis* luciferase fused to an hPEST protein destabilisation sequence) from a promoter containing five copies of an NF- κ B response element and was generated by cutting pGL4.32 with XbaI, followed by religation to remove the hygromycin expression cassette. To generate plasmid pANG183, the HindIII-Adel fragment of pANG32 was replaced with that of pGL4.36, which removed the TATA box. Plasmid pLDLR-Luc2P was derived from pANG183 by replacing the KpnI-HindIII fragment with a sequence that includes nucleotides -428 to -91 of the human *LDLR* gene (all coordinates are relative to translation start site) [22]. In plasmid pLDLR-UTR-Luc2P and its derivatives the reporter is driven from nucleotides -428 to -1 of the *LDLR* gene with the consensus start site of transcription at -86 (MANE transcript ENST00000558518.6); the constructs were generated by replacing the BmgBI-NcoI fragment of pLDLR-Luc2P with annealed oligonucleotides corresponding to nucleotides -97 to -2 (see Supplemental Methods S2). Plasmid pLDLR-UTR-HiBiT-Luc2P and related constructs differ from pLDLR-UTR-Luc2P by the fusion of a sequence expressing an 11-amino acid HiBiT peptide tag to the N-terminus of Luc2P [23]. These plasmids were constructed by recombining BmgBI and NcoI-digested pLDLR-Luc2P with double-stranded DNA fragments using NEBuilder HiFi DNA Assembly reagent (New England Biolabs, Hitchin, UK). All plasmids were verified by sequencing.

Cell culture

Medium A refers to DMEM/F12 supplemented with antibiotics (100 units per ml of penicillin and 100 μ g per ml of streptomycin sulphate) and 10% FBS (Fetal bovine serum). Medium B refers to DMEM/F12 supplemented with antibiotics, 5% lipoprotein-deficient serum (LPDS), 0.5 μ M lovastatin and 50 μ M mevalonate. Human alveolar adenocarcinoma A549 cells (ATCC CCL-185) were grown at 37 °C in an atmosphere of 5% CO₂ in medium A.

Plasmid transfection and enzyme assays

A549 cells were transfected using TurboFect™ reagent according to the manufacturer's instructions (Thermo Fisher Scientific, MA, USA) at a ratio of 3 μ l Turbofect per μ g DNA. To measure the activities of *P. pyralis* and *R. reniformis* luciferases, cells in 96-well plates were lysed and analysed with Dual-Glo Luciferase Assay reagents (Promega) as described [24]. Expression of the HiBiT tag was detected based on its binding to the LgBiT complementation partner, which reconstitutes a NanoLuc luciferase activity that was measured using the Nano-Glo® HiBiT Lytic Detection System (Promega). To measure the activity of bacterial β -galactosidase expressed from the pCMV- β Gal plasmid, A549 cells in 96-well plates were

lysed with 50 μ l lysis buffer and incubated on ice for 30 min. For β -Gal assays, 20 μ l of lysate were mixed with 30 μ l of 5 mM ONPG in a black-walled clear-bottom microplate and incubated at 37 °C for 1–3 h. Luminescence and 405-nm absorbance readings were performed using a Promega GloMax® Discovery microplate reader. Linearity of the activity signals was verified by serial dilutions.

Statistics

Normal distribution of relative luciferase expression data was assessed in R using the Shapiro test. To assess differences between data groups, an unpaired Student's t test was performed where each dataset was normally distributed; otherwise, an unpaired Wilcoxon test was performed. Where the average difference between two sample groups is cited, the error associated with the difference was calculated by adding the coefficients of variation from each data group, and the resulting sum was then multiplied by the difference.

Variant nomenclature and accession

The variants described in this paper are done so in accordance with HGVS recommendations. Nucleotide numbering is based on reference GenBank sequence NM_000527.4. All variants described are available on ClinVar.

RESULTS

LDLR 5'UTR variants

The 5'UTR analysis pipeline, utilising UTRannotator, was applied to 536 participants of the FH cohort in 100KGP. One 5'UTR variant, NM_000527.5(*LDLR*):c.-35C>G ((Accession: VCV003893685.1) (<https://www.ncbi.nlm.nih.gov/clinvar/variation/3893685/>)) was identified as potentially affecting the open reading frame of *LDLR* (Table 1) (Fig. 1). This variant is a single nucleotide variant (SNV) located 35 base pairs upstream of the 1st base of the protein-coding sequence (CDS) of *LDLR* and consists of a cytosine to guanine transversion. UTRannotator classified this variant as an upstream start (uAUG) gain variant with a moderate Kozak consensus, which is predicted to introduce a novel overlapping open reading frame (oORF) 63 amino acids in length. This variant was found in an FH proband with no Tier 1 or Tier 2 variants (<https://re-docs.genomicsengland.co.uk/tiering/>). In gnomAD 4.1.0 one non-Finnish European out of 798,530 sequenced individuals was found to carry the variant, and no other carriers were identified in the 77,277 100KGP non-FH participants.

A literature search for other *LDLR* 5'UTR variants associated with FH found a single-nucleotide deletion at position -22, NM_000527.5(*LDLR*):c.-22del ((Accession: VCV000250962.1) (<https://www.ncbi.nlm.nih.gov/clinvar/variation/250962/>)), reported by Sözen et al. that is predicted to introduce a new ATG start codon and 58 amino acid long open reading frame out of frame with the wild-type CDS [14]. The variant was identified in a compound heterozygous patient presenting with cholesterol levels indicative of a homozygous familial hypercholesterolaemia (FH) phenotype (Total cholesterol (mmol/l): 20.12, LDL (mmol/l): 18.41, with xanthomas and CHD at age 8 years). Genetic analysis showed that the father carried only the c.-22del variant, while the mother had both *LDLR* c.762 G>T p.(Gln254His) and c.763 T>G p.(Cys255Gly) alterations on the same allele. The father, from whom the proband inherited the variant, had a phenotype consistent with heterozygous FH and total cholesterol of 8.83 mmol/L (HDL: 0.95 mmol/L, LDL: 6.73 mmol/L) [14]. The c.-22del variant was not found in gnomAD 4.1.0 or in any participant in the 100KGP.

FH-associated 5'UTR variants reduce expression of the wild-type CDS

If the novel AUG codons introduced in the 5'UTR were used preferentially for the initiation of translation, the variants would be expected to reduce the expression of the wild-type *LDLR* open reading frame. To test this hypothesis, we constructed plasmids in which a reporter gene encoding firefly luciferase (FLuc2P) was

Table 1. Information of c.–35C > G variant with data derived from UTRannotator, also included is the c.–22del variant.

Chrom:Pos	Ref	Alt	DNA change	Variant type	Consequence	Distance to CDS	Distance to next in-frame STOP	Kozak Context	GnomAD (v4) AF
19:11089514	C	G	c.–35C > G	5' UTR, uAUG gain	Creation of oORF	37	186	GTGATGC	0.0000007
19:11089527	C		c.–22del	5' UTR, uAUG gain	Creation of oORF	24	173	GACATGC	0

Chrom Chromosome, Pos Position, Ref Reference allele, Alt Alternate allele, CDS Coding sequence, AF Allele frequency.

preceded by the promoter (bases –428 to –93) plus 5'UTR (bases –92 to +4) of the human *LDLR* gene, such that the entire Kozak sequence (AGCATGG) at the luciferase translation initiation site corresponded to that of the *LDLR*. Following transfection, cells were grown in cholesterol-depleting medium containing lipoprotein-deficient serum (LPDS) and lovastatin to promote maximal expression through the promoter's SREBP response element [25]. A low concentration of mevalonate was supplemented to ensure survival of the cells in the presence of lovastatin. As a control, some cells received SREBP-suppressing 25-hydroxycholesterol to confirm normal regulation of the reporters. As shown in Fig. 2A (black bars), expression of firefly luciferase was reduced by $59.4\% \pm 4.5\%$ (Mean \pm SEM, $p < 0.0001$) following the introduction of the c.–35C > G variant. For the plasmid with the c.–22del variant, expression was reduced by $46.7\% \pm 3.4\%$ ($p < 0.002$). In Fig. 2B similar results can be seen for cells grown purely with in standard growth medium containing FBS or for cells grown with LPDS (without lovastatin and mevalonate). Compared to wild-type, expression of the c.–35C > G variant expression was reduced by $42.3\% \pm 4.9\%$ (FBS, $p = 0.01$) and $53.3\% \pm 10.7\%$ (LPDS, $p < 0.0001$), respectively. In cells transfected with the c.–22del variant, expression was reduced by $43.1\% \pm 3.8\%$ (FBS, $p = 0.01$) and $47.5\% \pm 13.5\%$ (LPDS, $p = 0.004$).

To test whether the novel ATG sequence at position –35 of the UTR is in fact used for the initiation of translation, we constructed another set of plasmids in which the *LDLR* promoter and 5'UTR preceded a firefly luciferase gene with an N-terminal HiBiT tag. The HiBiT peptide can be detected with a luminescence assay through complementation of a split NanoLuc luciferase system [23]. Confirming the results in Fig. 2, HiBiT expression in cholesterol-depleting medium was high when the HiBiT-FLuc2P fusion protein was cloned in frame with the *LDLR* translation initiation codon (Fig. 3, “Wild-type” construct) but reduced by about 79% following introduction of a novel ATG at position –35 of the 5'UTR (construct “c.–35C > G”). Expression of the reporter was restored, however, when the HiBiT-FLuc2P CDS was brought in frame with the novel ATG (construct “c.–35C > G; insGG”), measurements of FLuc2P activity in the same extracts mirrored those obtained for the HiBiT tag (data not shown).

We were initially surprised by the high expression of the reporter when cloned in frame with the novel AUG because the surrounding sequence is a very poor match for the Kozak consensus sequence [26]. To further explore this observation, we generated another reporter plasmid in which the HiBiT-FLuc2P CDS was again in frame with the AUG at –35 but the nucleotides surrounding the novel initiation codon were changed to match the core Kozak sequence (construct “c.–40_–34 delinsACCATGG; insGG”). The expression of this reporter was comparable to that of the reference construct. Taken together, the results in Figs. 2 and 3 show that the c.–35C > G variant leads to preferential initiation of translation from the novel AUG, leading to a significant reduction of protein synthesis from the usual *LDLR* translation start site.

Variant classification with The American College of Medical Genetics and Genomics (ACMG) guidelines

To assess the pathogenicity of the c.–35C > G and c.–22del *LDLR* variants, combined criteria from the ACMG guidelines adapted by Chora et al., specific to FH, and Ellingford et al., specific to non-coding variants, were used [27, 28]. The c.–22del variant was not present in GnomAD v4.1.0, and c.–35C > G had a PopMax MAF ≤ 0.0002 (0.02%), therefore both meet the PM2 moderate pathogenic criteria as they are sufficiently rare in the population (PM2, Chora et al.). The variants were found in patients who have been diagnosed with FH, therefore PP4 supporting pathogenic criteria for a variant present in a patient with FH (Chora et al.) was applied. *LDLR* haploinsufficiency is a cause of FH and the variants have the same predicted impact as previously identified pathogenic variants, which cause *LDLR* haploinsufficiency [29, 30]. Translation from each variant's resultant uAUG would create an overlapping open reading frame (oORF) that overlaps



Fig. 1 Schematic of 5'UTR and WT translated *LDLR* exon 1. The variants c.-35C>G and c.-22del are shown imposed on the 5'UTR **A**. Hypothetical translated products starting from the uATGs resulting from c.-35C>G and c.-22del are shown below as **B**, **C** respectively. c.-35C>G, c.-22del, and the wild type ATG have their respective Kozak consensus sequences shown in flanking brackets.

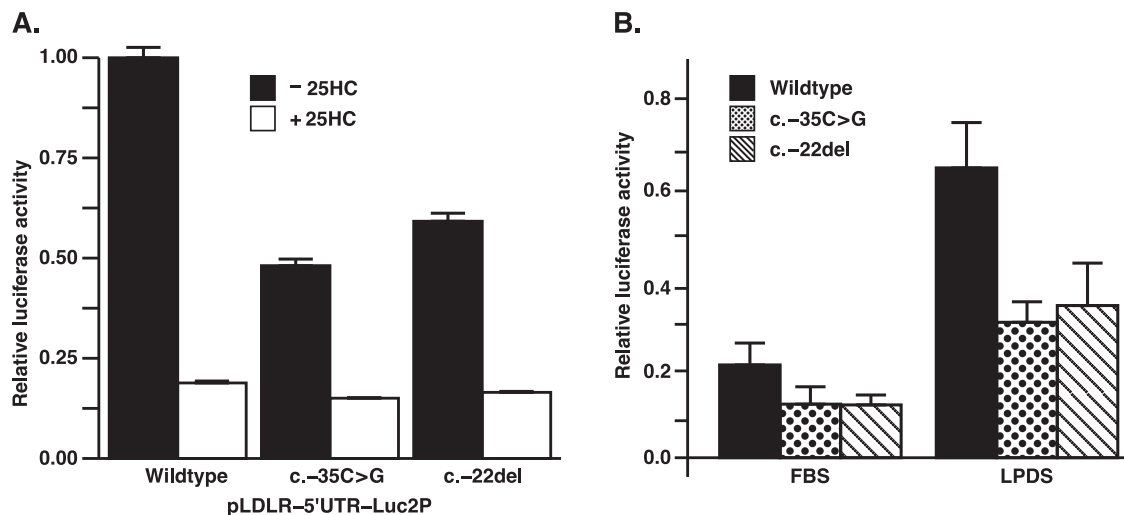


Fig. 2 Measurement of luciferase reporter activity. On day 1 of growth, A549 cells in 96-well plates were transfected with 1 ng/well of pRL-CMV (expressing *Renilla* luciferase) plus 125 ng/well of pLDLR-UTR-Luc2P reporter plasmids containing either the wild-type *LDLR* 5'UTR or 5'UTR sequences with the indicated mutations. On day 2, cells were switched to cholesterol-depleting medium B in the absence or presence of 1 µg/ml 25-hydroxycholesterol (25HC). Cells were harvested and lysates analysed on day 3. Firefly luciferase activity was divided by *Renilla* luciferase activity. The results from three independent experiments were each normalised and then averaged. Error bars indicate SEM. **A** Relative firefly luciferase activity in DMEM/F12 supplemented with antibiotics (100 units per ml of penicillin and 100 µg per ml of streptomycin sulphate), 5% lipoprotein-deficient serum (LPDS), 0.5 µM lovastatin and 50 µM mevalonate (LPDS/Lov) and LPDS/Lov plus 25-hydroxycholesterol (LPDS/Lov + 25HC). **B** Relative firefly luciferase activity in all tested media; DMEM/F12 supplemented with antibiotics and 10% FBS (FBS), DMEM/F12 supplemented with antibiotics, 5% lipoprotein-deficient serum (LPDS), LPDS/Lov, and LPDS/Lov + 25HC.

the coding sequence out-of-frame with the canonical start site (PM5, Ellingford et al.). Luciferase functional assays demonstrated a $59.4\% \pm 4.5\%$ (Mean \pm SEM) reduction in activity compared with wild type, which is sufficient to apply the PS3_Supporting (Chora et al.) criterion for c.-35C>G. For the c.-22del variant a < 50% reduction in activity was observed ($46.7\% \pm 3.4\%$ (Mean \pm SEM)) therefore not sufficient for PS3_Supporting criteria (Chora et al.). Following this, for the c.-22del variant, we are unable to apply PP1 as currently the segregation data is only available for one family (as per Chora et al. guidelines) [14]. Additionally, we are unable to apply PM3 as the other two variants ((NM_000527.5(*LDLR*):c.762 G>T, NM_000527.5(*LDLR*):c.763 T>G)) the proband inherited in *trans* both have conflicting classifications of pathogenicity and therefore are classified as VUSs at the time of writing (Chora et al.) [14]. Altogether there is sufficient evidence for c.-35C>G to be classified as Likely Pathogenic, while c.-22del met criteria of a VUS.

DISCUSSION

Here we demonstrate a novel molecular mechanism for non-coding variation in *LDLR* leading to FH. Luciferase assays comparing wild-type *LDLR* 5'UTR to the c.-35C>G and

c.-22del variants confirm a significant reduction in *LDLR* promoter activity, of 59.4% and 46.7% activity, respectively. These variants are thus likely to lead to an impaired expression of LDL receptors, which would explain the insufficient clearance of LDL-C from the blood of the affected probands.

Both sequence changes lead to the introduction of novel uAUG codons that are out of frame with the wild-type translation start codon. In humans, translation is initiated when the small ribosomal subunit encounters the first AUG codon, which let us to hypothesise that the c.-35C>G and c.-22del variants reduce expression due to the competition of the novel and normal start sites for the ribosome. This hypothesis was borne out by experiments where a reporter CDS was cloned in frame with the variant uAUGs (Fig. 3). Although the sequence surrounding the uAUG introduced at position -35 is a poor match to the consensus initiation site [26], the inclusion of an ideal Kozak sequence around this uAUG (construct "c.-40_34 delinsACCATGG; insGG") did not enhance translational efficiency of the in-frame reporter, suggesting that the presence of a uAUG alone is sufficient to disrupt *LDLR* protein production.

The short peptides that are predicted to be produced from the two uAUG variant have an unknown function at present, and it is unclear whether they have any biological effect.

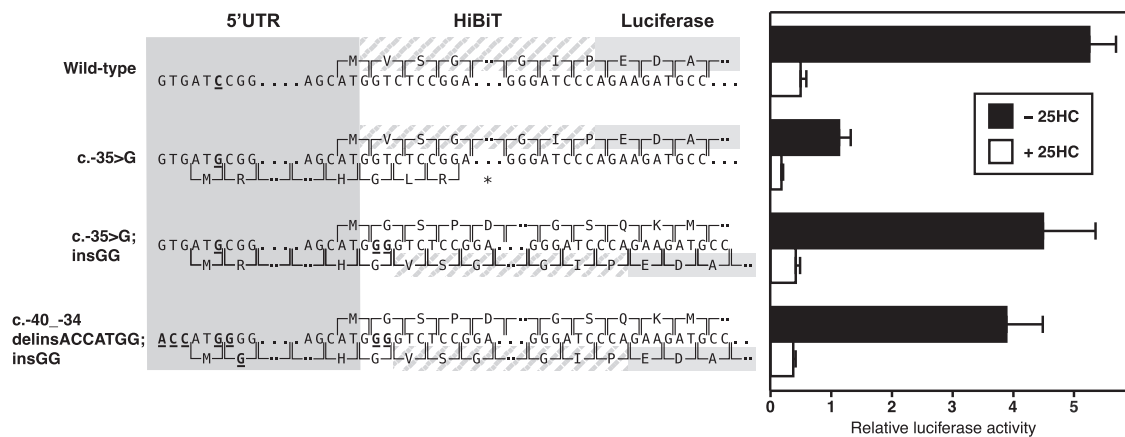


Fig. 3 Measurement of luciferase and HiBiT reporter activity. A549 cells were grown as in Fig. 2 and transfected with 10 ng/well of pCATlac (expressing β -galactosidase) plus 125 ng/well of pLDLR-UTR-HiBiT-Luc2P reporter plasmids. On day 2, cells were switched to cholesterol-depleting medium B in the absence or presence of 1 μ g/ml 25-hydroxycholesterol (25HC). Enzyme activities were analysed on day 3. NanoLuc (indicating HiBiT expression) was corrected for β -galactosidase. Error bars indicate SD ($n = 3$). Sequences that differ between the four reporters are highlighted in red. Plasmid (b) differs from wild-type plasmid (a) by a c.-35C > G mutation in the *LDLR* 5'UTR; the mutation introduces an upstream AUG that is out of frame with the HiBiT-Luc2P CDS. Plasmid (c) differs from plasmid (b) by the insertion of two G nucleotides following the *LDLR* Kozak sequence, which brings HiBiT-Luc2P in frame with the upstream AUG. Plasmid (d) differs from plasmid (c) in that sequences surrounding the upstream AUG were modified to match the core Kozak consensus sequence.

Our findings underscore the importance of considering non-coding variants in genetic diagnostics for FH, as they can significantly impact gene expression even when the coding sequence appears intact. However, compared to other pathogenic variants in the *LDLR* promoter, these uAUG creating variants have a somewhat stronger effect. SP1 binding site variants were shown to alter promoter activity (% of wildtype) by 6–26%, SREBP2 binding site variants 11–40%, and SREBP1 binding site variants 10% [31–34]. Therefore, identification of a likely pathogenic uAUG creating variant can provide the patient with a genetic diagnosis, which can then be used for cascade testing and to personalise the treatment.

The use of UTRannotator in a large sequencing dataset was particularly valuable in identifying the c.-35C > G variant as a uAUG gain variant and predicting its impact on translation [21]. In addition to potential implementation in current FH genetic testing of *LDLR*, this tool can be further employed to investigate other important FH-related genes. For example, the *PCSK9* mRNA encodes a uORF with the uAUG at Chr1-55039566-55039568. Loss-of-uORF variants are thought to be associated with enhanced translation [35–37], which in the case of *PCSK9* would be expected to cause increased turnover of the *LDLR* protein and might thus impair the clearance of LDL. Though our study did not find any such uAUG loss variants in the 100KGP cohort, 5 SNVs are listed in GnomAD v4.1.0 which would cause a uAUG loss (all described SNVs have MAF of <0.0001). Whether these SNVs are associated with hypercholesterolaemia has not yet been reported.

One limitation of our study is the inability to access lipid profile data for the 100KGP FH cohort due to Genomics England confidentiality constraints. This restricts our capacity to fully correlate the functional impacts of identified variants with phenotypic data across the entire cohort but all patients within the 100KGP FH cohort are confirmed to have been diagnosed according to the Simon Broome criteria. Family co-segregation of the c.-35 variant and collection of patients cells to perform western blotting to confirm the expression defect were not possible due to the loss of contact with the study participant.

The implications of this research are twofold. First, it highlights the need for broader screening of non-coding regions in FH diagnostics, as variants like c.-35C > G and c.-22del may contribute to the disease in patients lacking known pathogenic variants in coding regions. Second, it provides a potential

molecular mechanism by which non-coding variants can contribute to FH, specifically through the creation of uORFs that interfere with normal translation. Future studies should focus on expanding the catalogue of 5'UTR variants in FH patients and investigating their functional impact using in vitro and in vivo models.

DATA AVAILABILITY

Genetic and phenotypic data for the 100KGP study participants are available through the Genomics England Research Environment via the application at <https://www.genomicsengland.co.uk/research/academic/join-research-network>.

REFERENCES

- Nordestgaard BG, Chapman MJ, Humphries SE, Ginsberg HN, Masana L, Descamps OS, et al. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: Consensus Statement of the European Atherosclerosis Society. *Eur Heart J*. 2013;34:3478–3490.
- Marks D, Thorogood M, Neil HAW, Humphries SE. A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. *Atherosclerosis*. 2003;168:1–14.
- Futema M, Whittall RA, Kiley A, Steel LK, Cooper JA, Badmus E, et al. Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic. *Atherosclerosis*. 2013;229:161–168.
- Medeiros AM, Alves AC, Bourbon M. Mutational analysis of a cohort with clinical diagnosis of familial hypercholesterolemia: considerations for genetic diagnosis improvement. *Genet Med*. 2016;18:316–324.
- Reeskamp LF, Tromp TR, Defesche JC, Grefhorst A, Stroes ESG, Hovingh GK, et al. Next-generation sequencing to confirm clinical familial hypercholesterolemia. *Eur J Prev Cardiol*. 2021;28:875–883.
- Sharifi M, Rakhit RD, Humphries SE, Nair D. Cardiovascular risk stratification in familial hypercholesterolaemia. 102. *Heart*. BMJ Publishing Group; 2016. p. 1003–1008.
- Awan Z, Choi HY, Stitzel N, Ruel I, Bamimore MA, Husa R, et al. APOE p.Leu167del mutation in familial hypercholesterolemia. *Atherosclerosis*. 2013;231:218–222.
- Marduel M, Ouguerram K, Serre V, Bonnefont-Rousselot D, Marques-Pinheiro A, Erik Berge K, et al. Description of a large family with autosomal dominant hypercholesterolemia associated with the APOE p.Leu167del mutation. *Hum Mutat*. 2013;34:83–87.
- Garcia CK, Wilund K, Arca M, Zuliani G, Fellin R, Maioli M, et al. Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science*. 2001;292:1394–1398.

10. Leigh SEA, Foster AH, Whittall RA, Hubbart CS, Humphries SE. Update and Analysis of the University College London Low Density Lipoprotein Receptor Familial Hypercholesterolemia Database. *Ann Hum Genet.* 2008;72:485–498.
11. 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. *Ann Hum Genet.* 2012;76:387–401.
12. Khamis A, Palmén 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. *Eur J Hum Genet.* 2015;23:790–795.
13. Medeiros AM, Alves AC, Francisco V, Bourbon M. Update of the portuguese familial hypercholesterolaemia study. *Atherosclerosis.* 2010;212:553–558.
14. Sözen MM, Whittall R, Öner C, Tokatli A, Kalkanoğlu HS, Dursun A, et al. The molecular basis of familial hypercholesterolaemia in Turkish patients. *Atherosclerosis.* 2005;180:63–71.
15. Akiyamen LE, Genest J, Shan SD, Reel RL, Albaum JM, Chu A, et al. Estimating the prevalence of heterozygous familial hypercholesterolaemia: a systematic review and meta-analysis. *BMJ Open.* 2017;7:e016461.
16. Gratton J, Humphries SE, Futema M. Prevalence of FH-Causing Variants and Impact on LDL-C Concentration in European, South Asian, and African Ancestry Groups of the UK Biobank-Brief Report. *Arterioscler Thromb Vasc Biol [Internet].* 2023;43:1737–1742.
17. Versmissen J, Oosterveer DM, Yazdanpanah M, Defesche JC, Basart DCG, Liem AH, et al. Efficacy of statins in familial hypercholesterolaemia: a long term cohort study. *BMJ.* 2008;337:223–226.
18. Mark C, Jim D, Martin D, Leila E, Tom F, Sue H, et al. figshare. 2017. The National Genomics Research Library v5.1, Genomics England.
19. Bird M, Rimbart A, Pittman AM, Humphries SE, Futema M. Variants in *LPA* are associated with Familial Hypercholesterolaemia: whole genome sequencing analysis in the 100,000 Genomes Project. *Eur J Prev Cardiol.* 2024.
20. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17:1–14.
21. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5'untranslated region variants with the UTRannotator. *Bioinformatics.* 2021;37:1171–1173.
22. Castoreno AB, Wang Y, Stockinger W, Jarzylo LA, Du H, Pagnon JC, et al. Transcriptional regulation of phagocytosis-induced membrane biogenesis by sterol regulatory element binding proteins. *Proc Natl Acad Sci USA.* 2005;102:13129–13134.
23. Schwinn MK, Machleidt T, Zimmerman K, Eggers CT, Dixon AS, Hurst R, et al. CRISPR-Mediated tagging of endogenous proteins with a luminescent peptide. *ACS Chem Biol.* 2018;13:467–474.
24. Zhang S, Nohturfft A. Studying membrane biogenesis with a luciferase-based reporter gene assay. *J Vis Exp.* 2008;920.
25. Briggs MR, Yokoyama C, Wang X, Brown MS, Goldstein JL. Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter. I. Identification of the protein and delineation of its target nucleotide sequence. *J Biol Chem.* 1993;268:14490–14496.
26. Kozak M. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 1984;12:857–872.
27. Chora JR, Iacocca MA, Tichý L, Wand H, Kurtz CL, Zimmermann H, et al. The Clinical Genome Resource (ClinGen) Familial Hypercholesterolemia Variant Curation Expert Panel consensus guidelines for LDLR variant classification. *Genet Med.* 2022;24:293–306.
28. Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, Campbell C, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med.* 2022;14:73.
29. Austin MA, Hutter CM, Zimmern RL, Humphries SE. Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. *Am J Epidemiol.* 2004;160:407–420.
30. Marduel M, Carrié A, Sassolas A, Devillers M, Carreau V, Di Filippo M, et al. Molecular spectrum of autosomal dominant hypercholesterolemia in france. *Hum Mutat.* 2010;31:E1811–E1824.
31. De Castro-Orós I, Pampín S, Bolado-Carrancio A, De Cubas A, Palacios L, Plana N, et al. Functional analysis of LDLR promoter and 5' UTR mutations in subjects with clinical diagnosis of familial hypercholesterolemia. *Hum Mutat.* 2011;32:868–872.
32. Sun XM, Neuwirth C, Wade DP, Knight BL, Soutar AK. A mutation (T→45C) in the promoter region of the low-density-lipoprotein (LDL)-receptor gene is associated with a mild clinical phenotype in a patient with heterozygous familial hypercholesterolaemia (FH). *Hum Mol Genet.* 1995;4:2125–2129.
33. Dedoussis GVZ, Pitsavos C, Kelberman D, Skoumas J, Prassa ME, Choumerianou DM, et al. FH-Pyrgos: a novel mutation in the promoter (–45delT) of the low-density lipoprotein receptor gene associated with familial hypercholesterolemia. *Clin Genet.* 2003;64:414–419.
34. 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;15:11186–11189.
35. Schulz J, Mah N, Neuenschwander M, Kischka T, Ratei R, Schlag PM, et al. Loss-of-function uORF mutations in human malignancies. *Sci Rep.* 2018;8:2395.
36. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA.* 2009;106:7507–7512.
37. Plassmeyer SP, Florian CP, Kasper MJ, Chase R, Mueller S, Liu Y, et al. A Massively Parallel Screen of 5'UTR mutations identifies variants impacting translation and protein production in neurodevelopmental disorder genes. *medRxiv.* 2023.

ACKNOWLEDGEMENTS

This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

AUTHOR CONTRIBUTIONS

MB and AN designed and conducted experiments, interpreted data, and wrote the manuscript, AN also provided project supervision; CJPT contributed to experimental data collection and helped write the manuscript; AP, EB, and MF revised the manuscript drafts and provided project supervision.

FUNDING

MB was funded by St George's University of London PhD studentship (internal funding 12729-58).

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL

The 100KGP, launched in the UK in 2013, generated WGS data to provide molecular diagnosis to participants with rare diseases and cancers. The 100KGP was approved by East of England – Cambridge Central Research Ethics Committee ref:20/EE/0035. Only participants who provided written informed consent for their data to be used for research were included in the analyses. The current project (RR123) has been approved by the Genomics England Clinical Interpretation Partnership (GeCIP) cardiovascular domain committee.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-025-01893-y>.

Correspondence and requests for materials should be addressed to Martin Bird or Marta Futema.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025