A common polymorphism that protects from cardiovascular disease increases fibronectin processing and secretion

Sébastien Soubeyrand, PhD[a]*, Paulina Lau MSc[a], Majid Nikpay, PhD[a], Anh-Thu Dang MSc[a], and Ruth McPherson, MD, PhD [a,b]*.

[a] Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada; [b] Department of Medicine, Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Canada.

Author Contributions: S.S. and P.L designed, performed and analyzed experiments; A.-T. D performed experiments; M.N. completed bioinformatic analyses; R.M. and S.S. wrote the manuscript.

* To whom correspondence may be addressed:

ssoubeyrand@ottawaheart.ca

rmcpherson@ottawaheart.ca

Competing Interest Statement: The authors declare no competing interests.

Keywords: Fibronectin 1, coronary artery disease, Single nucleotide polymorphism, signal peptide, glycosylation, inflammation.

Abstract

Recent large scale bioinformatic analyses have identified common genetic variants within the fibronectin (*FN1*) gene that predispose to cardiovascular disease, through mechanisms that remain to be investigated. This work explores the underlying mechanisms and identifies a novel process controlling fibronectin secretion. First, we demonstrate that higher levels of FN1 protein in plasma associate with a reduced risk of cardiovascular disease.. Next, cellular models were leveraged to demonstrate that the CAD associated region encompasses a L15Q polymorphism within the FN1 signal peptide that impacts secretion of FN1 both qualitatively and quantitatively. Thus, by reducing FN1 secretion, a variant within the signal peptide contributes to lower circulating FN1 and increased CAD risk. In addition to providing novel functional evidence implicating FN1 in cardiovascular disease, these findings demonstrate that a common variant within a secretion signal peptide regulates protein function.

Introduction

Genome-wide Association Studies (GWAS) have identified hundreds of common single nucleotide polymorphisms (SNPs) that are significant for cardiovascular disease (CAD) risk [1–3]. Although GWAS signals are enriched for eQTLs, the identification of causal genes is challenging since 1) the vast majority of common trait related SNPs do not overlap protein coding genes and 2) are in a eQTLs for multiple genes [4,5]. Validation of statistical associations by experimental approaches is an essential first step in the development of novel therapeutic interventions. As the majority of GWAS identified variants are unlikely to be causal for several reasons, the very identification of causal SNPs among the list of GWAS identified variants is itself a complex process [6]. Indeed, predictions place at least 80% of GWAS identified SNPs within a substantially wide 34 kbp window of causal variants in Europeans [7]. Clearly, mechanistic insights are limited at that level of resolution, especially since *trans* (long-distance) acting variants are prevalent and may account for significant heritability [8]. In order to pinpoint causal SNPs ("finemapping") and identify functionally important gene targets, various approaches have been used that leverage expression data, epigenetic information, etc. [9]. This approach has yielded surprising findings including variants located within and outside genes that regulate distal genes, as well as evidence of pervasive transcription independent mechanisms [10–12].

The Fibronectin 1 gene (*FN1*) encodes a group of protein isoforms that differ in sequence and localization: plasma (pFN) and cellular (cFN) [13]. Both forms are synthesized as precursors that are processed during ER/Golgi trafficking and either enrich the local matrix environment (cFN) or secreted into the circulation (pFN) [14]. The cellular forms exist as multiple variants that act as key structural components and regulators of the extracellular matrix

(ECM), where they are deposited as insoluble fibers involved in cell adhesion. The second major form of FN1, pFN, is secreted by the liver into the circulation where it is abundant. Mice deficient in pFN display largely normal hemostasis and wound-healing, consistent with a predominant or exclusive role of cFN in these processes [15]. Interestingly, pFN deficient mice display increased neuronal apoptosis and larger infarct areas following focal brain ischemia, suggesting that pFN plays a protective role, possibly by activating anti-apoptotic mechanisms via integrin signaling [15]. While pFN is not essential to vascular integrity, pFN has been shown to penetrate the vessel wall and to constitute a significant portion of arterial FN where it may participate in tissue remodeling [16,17].

Here, we explore and clarify the mechanisms linking CAD to common GWAS identified variants that map to the *FN1*gene. Using bioinformatic and molecular approaches we provide evidence that differential post-transcriptional regulation underlies the *FN1*-CAD association. More specifically a polymorphism within the propeptide of FN1 was found to regulate the ability of FN1 to be secreted. These findings provide a unique portrait of a common coding variant linked to CAD that has functional consequences at the protein level without affecting its mature amino acid sequence.

Results

rs1250259 links FN1 protein expression to coronary artery disease

The CAD-linked haplotype harbors several tightly linked SNPs that correlate with the disease (including top SNP rs1250229) that are causal candidates (**Figure 1, Table S1, Figure S1**).

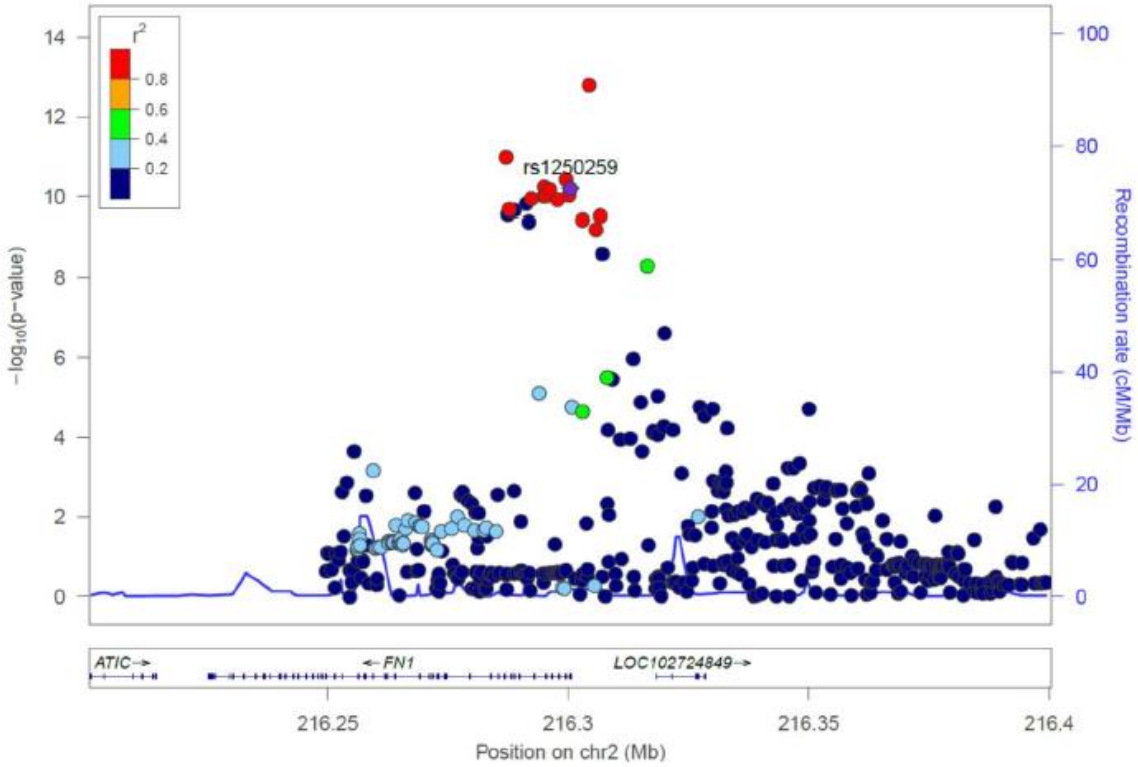

Figure 1. Local Manhattan plot of CAD association. CAD association data centred on rs1250259 (± 0.2 Mb) from Van der Harst (https://doi.org/10.1161/CIRCRESAHA.117.312086) plotted using LocusZoom, showing a signal enrichment around the upstream region of FN1.

Strikingly, the region contains a single coding SNP (rs1250259), central to the CAD associated region, which was prioritized for follow-up. Interrogation of genome-wide association studies using PhenoScanner and Open Targets points to an association between the most common allele (rs1250259-A) and lower pulse pressure, reduced CAD risk, as well as to changes in blood FN1 levels (**Table S2, Table S3**) [2,18–21]. While FN1 levels as a function of the rs1250259 genotype are not available, the proximal CAD protective T allele (rs1250258-T), closely linked (R^2 =0.99) to rs1250259-A is associated with increased circulating FN1 and fragments thereof, suggesting that it may play a cardioprotective role [22].

We next performed Mendelian randomization to test a causal role for FN1 per se. In this analysis, all SNPs associated with changes in FN1 protein expression are pooled and tested for association of each of CAD risk and FN1 levels. Consistent with individual SNP contributions, FN1 and CAD were inversely correlated, with higher circulating FN1 linked to lower CAD prevalence (**Table 1**).

Table 1. Mendelian randomization reveals an of impact FN1 on CAD. Probes (protein concentration) and corresponding CAD values are from Suhre et al [22]. Bxy, regression coefficient of x and y; se, Standard Error; p, pvalue of the beta; nsnp: number of snps used in model; Z, Z-score of the correlation. All values are rounded to 2 significant figures.

Probe	Outcome	bxy	se	p	nsnp	Z
3434-34_1	CAD	-0.059	0.0094	3.70E-10	5	-6.3
4131-72_2	CAD	-0.061	0.0094	1.00E-10	4	-6.5

Identification of a missense mutation within FN1 linked to CAD that is predicted to affect secretion

Although the above analysis focused on FN1 protein expression, the contribution to *FN1* mRNA expression remained to be tested. Genotype-Tissue Expression (GTEx) data indicate that the CAD linked haplotype region (using the top CAD SNP rs1250229 and rs1250259) was not associated with statistically significant changes in *FN1* expression in any of the available tissues (data not shown). This suggests that the haplotype may affect (harder to detect) distal genes, tissues not part of the GTEx panel or a combination thereof. Alternatively, the region may affect FN1 post-transcriptionally. Translation of rs1250259 is predicted to yield protein variants harboring either a Gln (rs1250259-T) or Leu (rs1250259-A) at position 15. Of note, the SNP haplotype is defined on the positive strand while the gene is transcribed in the negative orientation (**Fig S2**). FN1 is synthesized as a precursor that undergoes removal of a ~30 amino acid region containing a hydrophobic signal peptide (which includes residue 15) and a hydrophilic short pro-sequence [14].

The rs1250259 affects secretion of a FN1 fusion construct in transformed and primary cell models

To examine the impact of this substitution on FN1 secretion, a model fusion protein consisting of amino acids 1-182 of FN1 fused to a GFP-HA tag moiety was generated (**Figure 2A**).

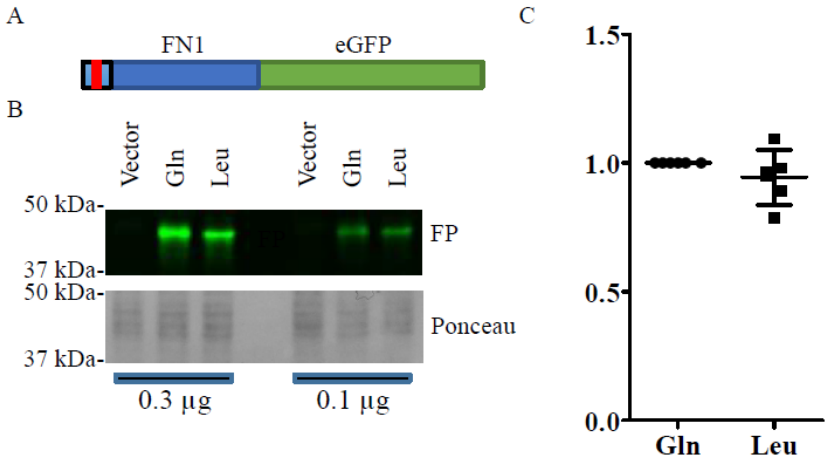

Figure 2. Both FN1 variants express similarly in HEK293T cells. A, Schema of the FN1-FP construct used. Drawing is approximately to scale. FN1 region corresponds to AA1-182, which encompasses, the signal peptide as well as 3 complete Fibronectin type-I domains corresponding to a previously reported crystal structure (2CG7, PDB entry). Signal peptide is in lighter blue. Red bar indicates position of the L15Q polymorphism. B, SDS-PAGE of HEK293T cell lysates expressing the common coding variants of FN1. HEK293T cells were transfected for 48 h with constructs encoding FN1-GFP fusion proteins. The Gln variant exhibited a slight retardation relative to the Leu variant. C, quantification of the cellular FP intensity after correction for transfection efficiency (Renilla).

This moiety is conserved in all FN1 forms and addresses technical limitations linked to the large size of FN1. The FN1 region chosen corresponds to the N-terminal heparin binding domain, which forms a well-defined region by crystallography and NMR and is shared by both secreted and cellular forms. Expression was first tested in HEK293T, a readily transfectable and widely available cell line. Following SDS-PAGE of cell lysates, a shift was observed: the Q15 form migrates slightly slower than the L15 variant (**Figure 2B**). Both fusion variants were present at comparable levels in HEK293T lysates after correcting for transfection efficiency (**Figure 2 C**).

Presence of the secreted protein in the media was tested next (**Figure 3**).

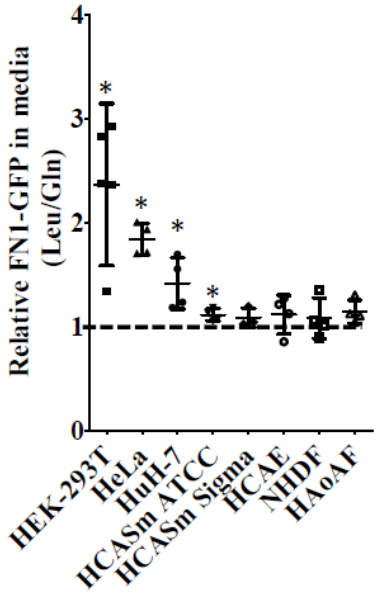

Figure 3. Presence of the CAD protective allele results in increased secretion of a FN1 model construct. Media and lysates from cells transduced for 48 hours with FN1-GFP plasmids encoding either Leu15 or Gln15 were analyzed by fluorometry. Ratios of media to cellular fluorescence were first assessed for each variant (L/Q) and the values for the Leu allele were divided by the corresponding Gln values. Results represent the mean from 3-5 biologicals ± 95% C.I. HCASm: human coronary artery smooth muscle cells from either ATCC or Sigma; HCAE: human coronary artery endothelial cells; NHDF: Normal Human Dermal fibroblast; HAoAF: Human aorta Adventitial fibroblast.

In HEK293T and HeLa cells, transfection resulted in the secretion of FN1-GFP in the media, with the L15 exhibiting greater propensity to be secreted, defined as the signal in the culture media relative to the cellular signal. To examine the impact of this polymorphism on secretion by the liver, which is the major physiological source of pFN1, HuH-7 hepatoma cells, a widely model of hepatocyte function, were transfected next. Although the difference was smaller than observed in the epithelial models above, L15 FN1-GFP was also more readily secreted by HuH-7 cells. Finally, FN1-GFP was transduced into several primary cell models with relevance to CAD, i.e., adventitial fibroblasts, endothelial cells, and coronary smooth muscle cells. In all models, the Leu form seemed on average better secreted than the Gln form, although the difference reached statistical significance only in a lot of coronary smooth muscle cells.

159 **Secreted forms of Q15 qualitatively differ in some primary cells**

160 Examination of the variants by SDS-PAGE revealed some unexpected findings. Delivery of
161 FN1-GFP demonstrated isoform-specific differences in the secreted forms, in a cell-type
162 different manner (**Figure 4**).

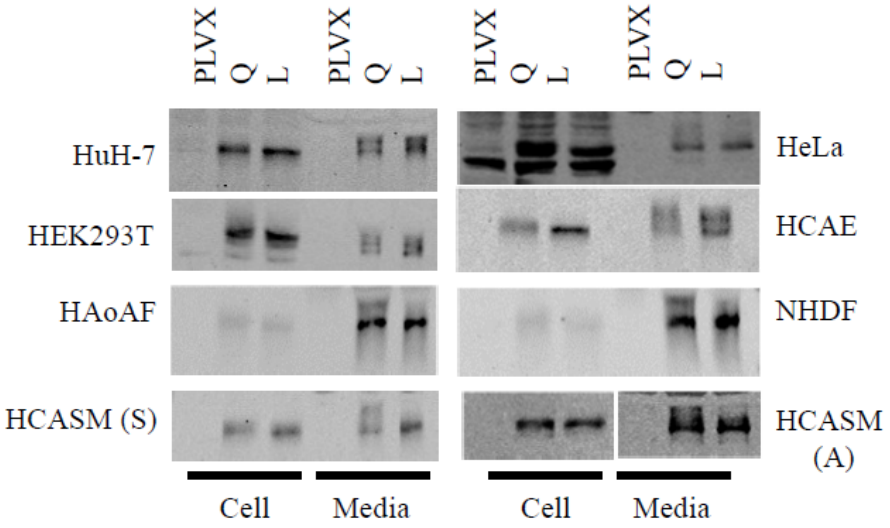

Figure 4. Multiple FN1-GFP species are secreted in cell cultures. Media (2% total well) and lysates (10% total well) from cells transduced for 72 hours with FN1-GFP plasmids encoding either L15 (T allele) or Q15 (A allele) were analyzed by Western blot. HCAE: human coronary artery endothelial cells; NHDF: Normal Human Dermal Fibroblast; HAoAF: Human aorta Adventitial fibroblast); HCASM: human coronary artery smooth muscle cells from either ATCC (A) or Sigma (S). Data is representative of at least 3 distinct biological replicates.
163

164 In some cells (fibroblasts, muscle models as well as HeLa cells), transduction of the Q15 form
165 led to enrichment relative to the L15 form of a slower migrating band on SDS-PAGE. By
166 contrast, FN1-GFP from endothelial cells and HEK293T resembled HuH-7 cells in that both
167 secreted forms exhibited qualitatively more similar profiles. Thus, in some cell types, the L15Q
168 polymorphism appears to dictate both quality and quantity of FN1-GFP secreted.

Differences in O-glycosylation account for the difference in migration

We hypothesized that this 3-5 kDa difference was due to variable levels of post-translational modifications, possibly glycosylation and/or retention of pro-peptides of different lengths. As full-length pFN1 is modified post-transcriptionaly by O and N-linked glycosylation, events commonly associated with secretion, glycosylation was examined first. Both variants secreted from dermal fibroblasts were subjected to deglycosylation reactions *in vitro* using a cocktail of enzymes targeting a wide range of glycosylation chains. The incubation resulted in the disappearance of the slower migrating form (**Figure 5**).

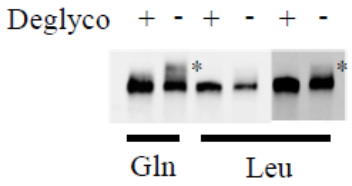

Figure 5. Glycosylation patterns of Q15 and L15 FN1-GFP from dermal fibroblast differ. Media from NHDF transduced for 72 hours with FN1-GFP plasmids encoding either L15 (T allele) or Q15 (A allele) were recovered by immunoprecipitation with anti-HA beads, denatured and treated with (+) or without (-) deglycosylation enzymes prior to Western blotting using an anti-GFP antibody. A higher exposure of the Leu samples is included to facilitate L-Q comparison. * indicates the position of a larger, glycosylated form.

Interestingly, a longer exposure of the L15 form also shows the presence of a slower migrating band that is also sensitive to glycosylation treatment. Thus, both forms are glycosylated, albeit to different extent.

The type of glycosylation involved was examined by treating cells with tunicamycin, which blocks N-glycosylation thereby interfering with protein transit through the Golgi

apparatus and secretion. Inclusion of tunicamycin severely reduced the amount, and altered the migration, of full-length FN1 recovered from the media but its impact on FN1-GFP was minor (**Figure S3**). These findings point to differential O-glycosylation between the two short FN1-GFP constructs.

The L15Q polymorphism results in similar N-terminal sequences

Although the slower form reflects distinct glycosylation, the underlying cause(s) remained to be clarified. We hypothesized that distinct glycosylation profiles might result from a shift in cleavage position of the signal peptide, as suggested by SignalP (**Figure S4**). Mass spectrometry of FN1-GFP fusions isolated from the culture media of NHDF however revealed that all forms consisted of either Gln or pyroGlu at their N-termini, consistent with previous studies on full-length pFN1 [23] (**Figure S5**). Thus, qualitative differences in N-terminal processing are unlikely to singly account for the different glycosylation patterns. Moreover, analysis of the gel region from the L15 sample, corresponding to a putative slower form, identified the unequivocal presence of FN1-GFP of lower abundance (~ 20% of the lower form), further suggesting that glycosylation occurs on both forms albeit to different extent, with the Q15 form showing increased glycosylation.

Quantitative differences in the secretion of the full-length FN1 variants

The impact of the L15Q polymorphism on full-length FN1 was tested next. Due to its large size, expression of a recombinant FN1 is challenging since 1) primary cells are difficult to transfect and 2) its coding sequence is too large for lentiviral delivery. For these reasons, analyses were performed on HEK293T, which are readily transfectable and wherein the polymorphism had a sizeable impact on the secretion of the short FN1-GFP form. Moreover, analysis was focused on

the pFN1 given its established link as a pQTL to the L15Q variant. The pFN1 construct was modified to express a C-terminal HA tag to simplify analysis. Western blot analysis revealed an unexpected difference in expressing cells, as the introduction of Q15 variant resulted in 2 distinct bands in cell lysates, in contrast to the L15 which showed only one (**Figure 6A**).

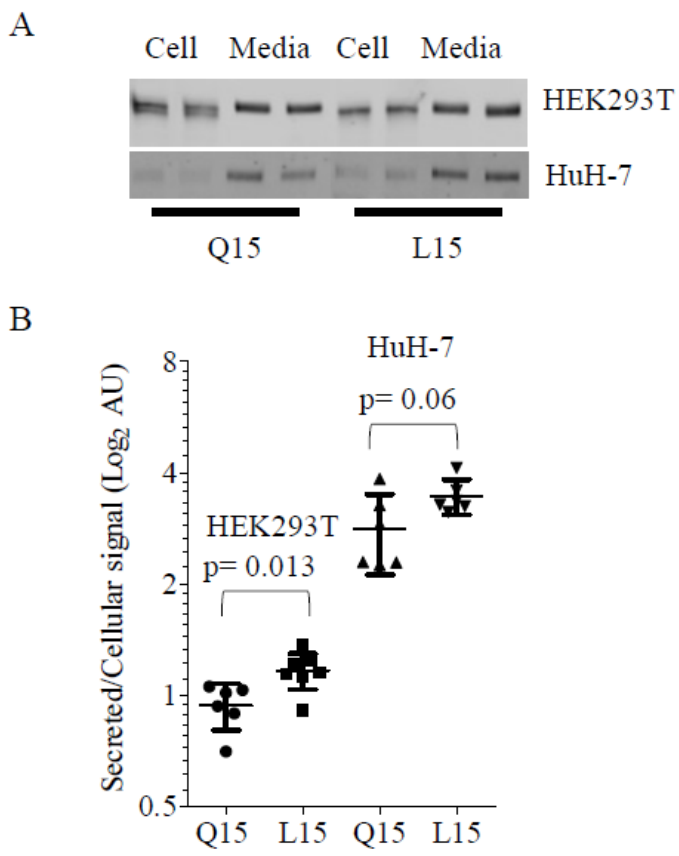

Figure 6. The full-length FN1 Q15 variant is less efficiently processed and secreted. Constructs encoding either variant of pFN1, tagged with a C-terminal Hemagglutinin tag, were transfected for 22 h in HEK293T or HuH-7 cells, as indicated A, Media and lysates were then analyzed by Western blot using an HA-specific antibody. Two biologics are shown for each construct. B, Quantification of FN1-HA Western blots. Signals from media and cytosol were quantified for each transfection and data is expressed as the secreted to cellular signals for Q15 and L15 (± 95% C.I.). Each point represents a distinct biologic.

By contrast the forms recovered from the media were similar. Interestingly, the additional Q15 band migrated faster than the L15 band, suggesting a lower mass, that may represent an incompletely glycosylated protein. We reasoned that incomplete glycosylation might reflect a slower or impaired processing which would result in decreased secretion of mature FN1. Quantification of plasma and cellular signals indicated trends (not statistically significant) in

facilitated L15 secretion (increased media signal and reduced cell signal) **(Figure S6)**. After internal correction to cellular signal however, a clear pattern emerged whereby the protective L15 variant showed statistically significant greater secretion (**Figure 6B)**

Macrophage polarization is associated with increased cellular and secreted FN1

The above experiments with the FN1-GFP fusion suggested that the cardioprotective form (L15) is secreted with greater efficiency and may be glycosylated to different extent. How increased FN1 might translate into reduced CAD risk or which form (cFN and/or pFN) may be most affected is unclear. Previous findings demonstrated that cFN expressing smooth muscle cells were associated with macrophage infiltration in plaque lesions [24]. In addition, deposited transcription data of human macrophages derived from *in vitro* differentiated blood monocytes, show increased FN1 expression in alternatively activated (anti-inflammatory M2) vs classically activated (inflammatory M1) or unpolarized macrophages [25] (**Figure S7**). That analysis however provided only total FN1 levels and did not examine secreted and cellular forms separately, which may be regulated differently. Using exon bridging strategies targeting either form in blood derived macrophages, we observed that pFN1 and cFN1 were similarly affected by polarization, although levels of pFN1 were more sensitive than cFN1 to the polarization status (M2/M1 pFN1vs cFN1 = 1.3; $p = 0.008$) (**Figure 7**).

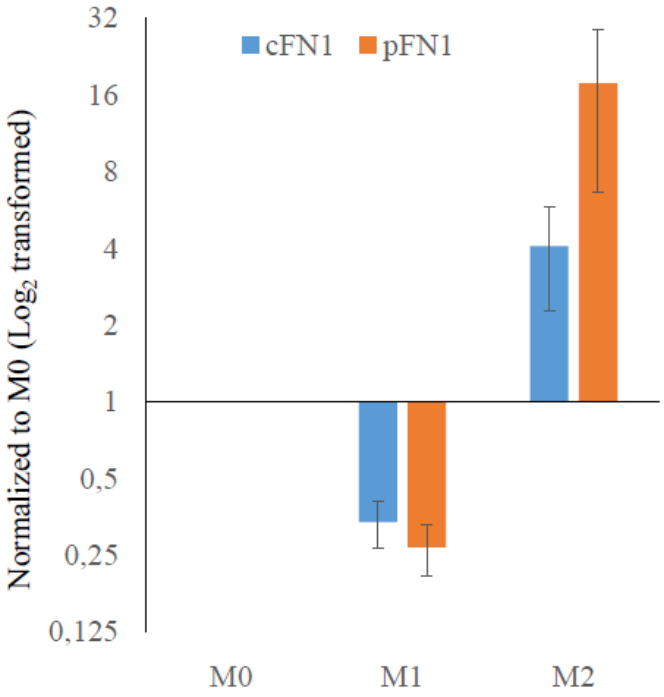

Fig 7. FN1 isoforms are decreased in M1- and increased in M2-activated macrophages. Levels of cFN1, pFN1 and SPR14 in human blood derived macrophages were measured by qRT-PCR. FN1 values were first normalized to the matching SRP14 values and are graphed relative to the M0 values (set to 1).

232

233

234

Discussion

Here, we provide experimental evidence that the CAD protective allele of a GWAS-identified SNP increases FN1 secretion. This work provides new insights into the more global issue surrounding the role of FN1 in the pathogenesis of CAD, which is explained by genetic and environmental factors in approximately equal proportion [26]. Moreover, it points to the role of a common variant that contributes to the heritable component of the disease. The pervasiveness of both alleles (Global Mean Allele Frequency: 0.23/0.77) in diverse ethnic groups, albeit with an uneven geographic distribution, suggests that the CAD risk variant encoding the Q15 form may confer some evolutionary benefit. Perhaps lower FN1 expression, at the expense of slightly greater risk of CAD, may enable a more potent immune response that may be advantageous to combat infections or protect from cancer. Examination of the UK biobank data via the PheWeb interface (http://pheweb.sph.umich.edu/SAIGE-UKB/gene/FN1) demonstrates an association of *FN1* with neoplasm (P=8.9e-7). While demonstrating that *FN1* may be linked to cancer, the link is with rs139452116 which results in a rare P2016L substitution and no significant association is evident for rs1250259, a SNP in strong LD with the signal peptide SNP rs1250258.

Signal peptides are critical for the proper maturation and secretion of extracellular proteins. Thus, mutations within secretion signal peptides can have profound repercussions if they affect the ability of the secretory apparatus to process them. A very rare R14W mutation within the signal sequence of carbonic anhydrase IV (CA4) is linked to retinitis pigmentosa and attributed to the accumulation of the immature protein within the ER, triggering the unfolded protein response and apoptosis [27]. Unlike this extreme and rare example, the common L15Q substitution has modest quantitative and qualitative impacts on FN1. Impacts on glycosylation observed on both $FN1_{1\text{-}182}$ GFP and full-length fibronectin did not yield a coherent pattern: the

Q15 form, while consistently less well secreted, exhibited increased glycosylation in $FN1_{1\text{-}182}$ GFP, at least in some primary models, but showed reduced glycosylation of full-length FN1 in our transformed cell models. Perhaps this reflects a cell-specific role of glycosylation in the control of FN1 secretion or, alternatively, a subsidiary role in defining the impact of the Q15L polymorphism. Additional investigations will be needed to resolve this question.

We demonstrate that the cardioprotective allele is linked to increased FN1 secretion indicating that circulating FN1 protects against CAD. One limitation to this interpretation is that it is derived from an integrative analysis of distinct cohorts: a UKBiobank/- and CARDIoGRAMplusC4D meta-analyis focusing on the genetics of CAD and correlative GWAS/pQTL derived from a healthy cohort [18,22]. One advantage of this approach is that by examining the impact of the SNPs predisposing to CAD on pFN1 levels in a largely healthy population, one avoids confounders frequently associated with CAD (additional underlying conditions, medications, etc.). This comes with an important limitation, as the impact of pFN1 levels on CAD is an extrapolation, albeit an informed one.

One mechanism underlying the role of FN1 in CAD involves the inflammatory compartment. As shown for alveolar macrophages, maturation from monocytes *in vitro* is linked to increased fibronectin production and secretion, which is in turn reduced upon inflammatory stimuli [28]. We observed that FN1 expression followed a >50-fold expression gradient, ranging from its lowest in M1 to intermediate in M0 and maximal in M2. Higher FN1 concentration may help maintain lesional macrophages in a more differentiated and anti-inflammatory state, thereby inhibiting local macrophage proliferation and associated atherosclerotic inflammation [29–31]. Alternatively, or in addition, increased production of FN1 may directly contribute to the suppression of the M1-like phenotype. Changes in FN1 levels conferred by this common genetic

variant are unlikely sufficient to induce polarization since induction of macrophage markers involves concerted changes including upregulation of both FN1 and integrins [32,33]. Rather, by subtly modulating the amount of available FN1, the variant may contribute to the formation of distinct inflammatory signatures within each macrophage subgroup.

Atherosclerosis has a complex, heterogeneous etiology, involving extensive tissue remodelling characterized by smooth muscle cell proliferation which is exacerbated by hypertension as well as invasion by circulating immune cells [34–36]. It was hypothesized that FN accumulation in the aortic media may play a role in the remodelling of the aortic wall in response to increased shear stress [37]. This is consistent with the observation that *FN1* SNPs are also linked to blood pressure traits, suggesting that FN1 might contribute to CAD in part through the regulation of the vascular tone. Thus, the cardioprotective property of FN1 might ultimately stem from its ability to regulate vascular wall ECM assembly, by jointly affecting vascular elasticity and inflammation.

Materials and Methods

Tissue culture

HuH-7 were obtained from and grown in low glucose DMEM supplemented with 1 g/L glucose and penicillin (0.1 mg/ml) and streptomycin (0.1mg/ml). HEK-293T and HeLa were from the ATCC and grown in DMEM with 4.5 g/L glucose supplemented with penicillin (0.1 mg/ml) and streptomycin (0.1mg/ml). Coronary smooth muscle cells were obtained from Sigma and the ATCC and maintained in the recommended media. Human coronary adventitial fibroblasts. Normal human dermal fibroblast. human aorta adventitial fibroblasts were purchased from Lonza. All primary cells were maintained in their recommended media. For fluorescence measurements, cells were shifted to phenol-free media for 48-72 h.

DNA constructs

The short GFP-HA fusion proteins (L15 and Q15) were obtained by chemical synthesis of two dsDNA block variants (BioBasic) encoding amino acid 1-182 of FN1 and inserted via restriction cloning in pLVX-puro digested with EcoRI/BamHI. The full-length pFN1 construct was obtained from Addgene (Fibronectin-human-plasma in pMAX; Plasmid #120401 [38]). A Q15L substitution was achieved by Q5 mutagenesis (New England Biolab) on a N-terminal Hind III/AvrII fragment transferred in pCMV5 digested similarly. Following validation by Sanger sequencing, the fragment was returned to the pMAX construct. A Hemagglutinin A epitope tag was then inserted via high fidelity assembly (NEBuilder HiFi DNA Assembly; New England Biolab) by swapping a synthetic fragment containing a C-terminal HA containing sequence within the RsrII digested pMAX pFN1 construct. The final assembly and sequences of these constructs are included in Supplemental Materials.

Transfection and transduction

Cells were transfected with lipofectamine 3000 (ThermoFisher) using a ratio of 3:2:1 (lipofectamine 3000 (µl): P3000 reagent (µl): DNA (µg)). For infection, viral particles were first generated in HEK-293FT cells using PVLX-puro (Clontech) alongside psPAX2 and pMD2.G obtained from Addgene. Virus containing supernatants were filtered through 400 nm filters and frozen at -80 C as is. Infections were performed in the presence of polybrene (2 µg/ml).

Immunoprecipitation and Western blotting

Cells were lysed in IP buffer (50 mM Tris-HCl, pH 7.4, 0.15 M NaCl, 0.1% Nonidet P40 (IGEPAL), 5 mM $MgCl_2$) for 2 min at 4 ºC. Lysates (1 mg protein equivalent) were then cleared by centrifugation (17,000 Xg) for 5 min and 20 µl of prewashed Anti-HA magnetic beads (Pierce) were added. For isolation from the media, 20 µl of beads were added to 3 ml of 400 nm filtered media harvested 72 h post-infection. Western blot was performed using 8 or 10% mini gels followed by wet transfer (1 h, 100 V) to Western grade nitrocellulose (Bio-Rad). Blots were incubated in Intercept blocking buffer (LI-COR) for 1 h and incubated for 16 h at 4 ºC in the presence of cognate primary antibodies diluted 1:2000 in TBS/T (50 mM Tris-HCl, pH 7.4, 0.15 M NaCl, 0.1% Tween-20). Secondary antibodies (donkey anti-mouse (680) or -rabbit (700); LI-COR) were diluted 1:20,000. Four 1 min washes in PBS were performed after each antibody incubation.

RNA isolation and qRT-PCR

RNA was isolated using the High Pure Isolation Kit (Roche). The Transcriptor First Strand cDNA Synthesis Kit (Roche) was used to generate cDNA using a 1:1 mixture of random hexamer and oligodT. PCR amplification and quantification were performed on a Roche

LightCycler 480 using the SYBR Green I Master reaction mix (Roche). For each experiment, relative amounts of target cDNAs were first expressed relative to SRP14. Results shown represent the means of 3 biological replicates. Oligonucleotides used are described in Supplemental Materials.

Mendelian Randomization

To investigate the possibility of an association between plasma protein level of FN1 and CAD, we did multi-SNP summary-based Mendelian randomization (MR) analysis which is also known as 2-sample Mendelian randomization [39]. For this purpose, we obtained summary association statistics (Beta and Standard error) for SNPs (pQTLs) that are independently (r^2<0.2) associated ($P<5e^{-8}$) with FN1 protein level and used these as an instrument to investigate a causal effect. This means, for SNPs in our instrument (MR N_{SNP}), we also obtained their summary association statistics (Beta and Standard error) with CAD and contrasted the effect sizes of the SNPs on FN1 (exposure) with the effect sizes of the SNPs on the CAD (outcome), to estimate the causal effect of FN1 on CAD. In this context, a significant negative association indicates individuals genetically susceptible to have higher levels of FN1 are at lower risk of CAD. MR analysis was done using the GSMR (*Generalised Summary-data-based Mendelian Randomisation*) algorithm implemented in GCTA software (version 1.92)[39]. As compared to other methods for 2-sample MR analysis, GSMR automatically detects and removes SNPs that have pleiotropic effect on both exposure and outcome using the HEIDI test; in addition, GSMR accounts for the sampling variance in β (beta) estimates and the linkage disequilibrium (LD) among SNPs, as such it is statistically more powerful than other 2-sample MR approaches. GSMR also incorporates a variety of quality assurance and helpful functions, notably aligning both GWAS summary

363 datasets to the same reference allele at each SNP. Excluding SNPs that difference between their
364 allele frequency in GWAS summary datasets and the LD reference sample is greater than 0.2, a
365 clumping function to only keep non-correlated (r^2<0.2) SNPs (with association P-value < $5e^{-8}$) in
366 the instrument and a function to generate the scatter plot of SNP effects. Previously we used this
367 approach to investigate the role of circulating miRNAs with regard to cardiometabolic
368 phenotypes [40]. We obtained GWAS summary statistics for CAD from the most recent meta-
369 analysis of CARDIoGRAMplusC4D and UK Biobank [18] and GWAS summary statistics for
370 SNPs that influence FN1 protein level from Suhre et al [22].

371 **Immunoprecipitation and deglycosylation reactions**

372 Culture media from Q15 and L15 NHDF infected for 96 h with lentiviral constructs expressing
373 FN1-GFP-HA were recovered, supplemented with 1 mM PMSF and centrifuged (1000 X g, 2
374 min) to remove cellular debris, and further cleared at high speed for 5 min (13,000 X g).
375 Recombinant FN1-GFP-HA was isolated from 10 ml of media (corresponding to a 10 cm culture
376 dish) using 25 µl anti-HA Pure Proteome magnetic beads (Pierce). Beads were washed 4 X 0.5
377 ml of PBS/1 % Triton X-100 and resuspended in 250 µl of the same buffer. Aliquots (10%) of
378 the isolates were used per deglycosylation reaction. Deglycosylation was performed using the
379 Protein Deglycosylation Mix II according to the supplier's protocol (New England Biolab).
380 Briefly, the immunoisolated material was denatured for 10 min at 75 °C and subjected to a
381 deglycosylation reaction for 30 min at 20 °C and 180 min at 37 C, using enzyme mix (2.5 µl) or
382 a mock reaction (no enzyme mix) in 25 µl of bead suspension. Samples were then denatured in
383 SDS-PAGE sample buffer and analyzed by Western blotting.

384 **Protein Analysis by LC-MS/MS**

For mass spectrometry, Q15 and L15 FN1-GFP samples were immunoprecipitated from the media of transduced NHDF as described above, resolved by SDS-PAGE and stained by colloidal Coomassie blue (Simply blue); NHDF were chosen for their greater proliferative ability over coronary models while exhibiting similar shifts on SDS-PAGE. Gel pieces were than excised and destained; a gel area matching a putative, lower abundance glycosylated L form was also included, for a total of 4 samples. Two distinct biologics per sample were analyzed. Proteomics analysis was performed at the Ottawa Hospital Research Institute Proteomics Core Facility (Ottawa, Canada). Proteins were digested in-gel using trypsin (Promega) according to the method of Shevchenko [41]. Peptide extracts were concentrated by Vacufuge (Eppendorf). LC-MS/MS was performed using a Dionex Ultimate 3000 RLSC nano HPLC (Thermo Scientific) and Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). MASCOT software version 2.6.2 (Matrix Science, UK) was used to infer peptide and protein identities from the mass spectra. The observed spectra were matched against custom sequences and against an in-house database of common contaminants. The results were exported to Scaffold (Proteome Software, USA) for further validation and viewing.

Acknowledgements and Funding: This work was funded by a Canadian Institutes for Health Research Foundation grant (FRN:154308; RM).

406 **References**

407 **1**. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an
408 Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ Res.
409 Wolters Kluwer Health; 2018;122: 433–443.

410 **2**. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide
411 association analysis identifies novel blood pressure loci and offers biological insights into
412 cardiovascular risk. Nat Genet. Nature Publishing Group; 2017;49: 403–415.

413 **3**. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association
414 analyses based on false discovery rate implicate new loci for coronary artery disease. Nat
415 Genet. 2017;

416 **4**. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-Associated SNPs
417 Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. Gibson G,
418 editor. PLoS Genet. 2010;6: e1000888.

419 **5**. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al.
420 Candidate causal regulatory effects by integration of expression QTLs with complex trait
421 genetic associations. PLoS Genet. 2010;6.

422 **6**. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal
423 variants by statistical fine-mapping. Nat Rev Genet. 2018;19: 491–504.

424 **7**. Wu Y, Zheng Z, Visscher PM, Yang J. Quantifying the mapping precision of genome-
425 wide association studies using whole-genome sequencing data. Genome Biol. 2017;18:
426 86.

427 **8**. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic
428 Inheritance. Cell. 2019;177: 1022-1034.e6.

429 **9**. Cannon ME, Mohlke KL. Deciphering the Emerging Complexities of Molecular
430 Mechanisms at GWAS Loci. Am J Hum Genet. 2018;103: 637–653.

431 **10**. Gupta RM, Hadaya J, Trehan A, Zekavat SM, Roselli C, Klarin D, et al. A Genetic
432 Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1
433 Gene Expression. Cell. 2017;170: 522-533.e15.

434 **11**. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-
435 associated variants within FTO form long-range functional connections with IRX3.
436 Nature. Nature Publishing Group; 2014;507: 371–375.

437 **12**. Wang Y, He B, Zhao Y, Reiter JL, Chen SX, Simpson E, et al. Comprehensive Cis-
438 Regulation Analysis of Genetic Variants in Human Lymphoblastoid Cell Lines. Front
439 Genet. Frontiers Media S.A.; 2019;10.

440 **13**. To WS, Midwood KS. Plasma and cellular fibronectin: Distinct and independent functions
441 during tissue repair. Fibrogenesis and Tissue Repair. 2011.

442 **14**. Gutman A, Yamada KM, Kornblihtt A. Human fibronectin is synthesized as a pre-
443 propolypeptide. FEBS Lett. 1986;207: 145–8.

444 **15**. Sakai T, Johnson KJ, Murozono M, Sakai K, Magnuson MA, Wieloch T, et al. Plasma
445 fibronectin supports neuronal survival and reduces brain injury following transient focal
446 cerebral ischemia but is not essential for skin-wound healing and hemostasis. Nat Med.
447 2001;7: 324–330.

448 **16**. Moretti FA, Chauhan AK, Iaconcig A, Porro F, Baralle FE, Muro AF. A major fraction of
449 fibronectin present in the extracellular matrix of tissues is plasma-derived. J Biol Chem.
450 2007;282: 28057–28062.

451 **17**. Kumra H, Sabatier L, Hassan A, Sakai T, Mosher DF, Brinckmann J, et al. Roles of
452 fibronectin isoforms in neonatal vascular development and matrix integrity. PLoS Biol.
453 Public Library of Science; 2018;16.

454 **18**. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an
455 Expanded View on the Genetic Architecture of Coronary Artery DiseaseNovelty and
456 Significance. Circ Res. 2018;122: 433–443.

457 **19**. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: A
458 database of human genotype-phenotype associations. Bioinformatics. Oxford University
459 Press; 2016;32: 3207–3209.

460 **20**. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al.
461 PhenoScanner V2: an expanded tool for searching human genotype–phenotype
462 associations. Bioinformatics. Oxford University Press (OUP); 2019;

463 **21**. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory
464 networks of human serum proteins link genetics to disease. Science (80-). American
465 Association for the Advancement of Science; 2018;361.

466 **22**. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting
467 genetic risk to disease end points through the human blood plasma proteome. Nat
468 Commun. 2017;8: 14357.

469 **23**. Garcia-Pardo A, Pearlstein E, Frangione B. Primary structure of human plasma
470 fibronectin. J Biol Chem. 1985;260: 10320–10325.

471 **24**. Dietrich T, Perlitz C, Licha K, Stawowy P, Atrott K, Tachezy M, et al. ED-B fibronectin
472 (ED-B) can be targeted using a novel single chain antibody conjugate and is associated
473 with macrophage accumulation in atherosclerotic lesions. Basic Res Cardiol. 2007;102:
474 298–307.

475 **25**. Martinez FO, Gordon S, Locati M, Mantovani A. Transcriptional Profiling of the Human
476 Monocyte-to-Macrophage Differentiation and Polarization: New Molecules and Patterns
477 of Gene Expression. J Immunol. The American Association of Immunologists; 2006;177:
478 7303–7311.

479 **26**. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. Circ Res.
480 2016;118: 564–578.

481 **27**. Rebello G, Ramesar R, Vorster A, Roberts L, Ehrenreich L, Oppon E, et al. Apoptosis-
482 inducing signal sequence mutation in carbonic anhydrase IV identified in patients with the
483 RP17 form of retinitis pigmentosa. Proc Natl Acad Sci U S A. Proc Natl Acad Sci U S A;
484 2004;101: 6617–6622.

485 **28**. Yamauchi K, Martinet Y, Crystal RG. Modulation of fibronectin gene expression in
486 human mononuclear phagocytes. J Clin Invest. 1987;80: 1720–1727.

487 **29**. Tang J, Lobatto ME, Hassing L, van der Staay S, van Rijs SM, Calcagno C, et al.
488 Inhibiting macrophage proliferation suppresses atherosclerotic plaque inflammation. Sci
489 Adv. 2015;1: e1400223–e1400223.

490 **30**. Robbins CS, Hilgendorf I, Weber GF, Theurl I, Iwamoto Y, Figueiredo J-L, et al. Local
491 proliferation dominates lesional macrophage accumulation in atherosclerosis. Nat Med.
492 2013;19: 1166–72.

493 **31**. Jenkins SJ, Ruckerl D, Cook PC, Jones LH, Finkelman FD, van Rooijen N, et al. Local
494 macrophage proliferation, rather than recruitment from the blood, is a signature of TH2
495 inflammation. Science (80-). 2011;332: 1284–8.

496 **32**. Xie B, Laouar A, Huberman E. Fibronectin-mediated cell adhesion is required for
497 induction of 92-kDa type IV collagenase/gelatinase (MMP-9) gene expression during
498 macrophage differentiation: The signaling role of protein kinase C-β. J Biol Chem.
499 1998;273: 11576–11582.

500 **33**. Ferreira OC, Valinsky JE, Sheridan K, Wayner EA, Bianco C, Garcia-Pardo A. Phorbol
501 ester-induced differentiation of U937 cells enhances attachment to fibronectin and
502 distinctly modulates the α5β1 and α4β1 fibronectin receptors. Exp Cell Res. Exp Cell Res;
503 1991;193: 20–26.

504 **34**. Laurent S, Boutouyrie P. The Structural Factor of Hypertension: Large and Small Artery
505 Alterations. Circulation Research. Lippincott Williams and Wilkins; 2015. pp. 1007–1021.

506 **35**. Brown IAM, Diederich L, Good ME, DeLalio LJ, Murphy SA, Cortese-Krott MM, et al.
507 Vascular smooth muscle remodeling in conductive and resistance arteries in hypertension.
508 Arterioscler Thromb Vasc Biol. Lippincott Williams and Wilkins; 2018;38: 1969–1985.

509 **36**. Intengan HD, Schiffrin EL. Vascular remodeling in hypertension: roles of apoptosis,
510 inflammation, and fibrosis. Hypertension. 2001. pp. 581–587.

511 **37**. Bézie Y, Lamazière JMD, Laurent S, Challande P, Cunha RS, Bonnet J, et al. Fibronectin
512 expression and aortic wall elastic modulus in spontaneously hypertensive rats. Arterioscler
513 Thromb Vasc Biol. Lippincott Williams and Wilkins; 1998;18: 1027–1034.

514 **38**. Rossnagl S, Altrock E, Sens C, Kraft S, Rau K, Milsom MD, et al. EDA-Fibronectin
515 Originating from Osteoblasts Inhibits the Immune Response against Cancer. PLoS Biol.
516 Public Library of Science; 2016;14.

517 **39**. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations
518 between risk factors and common diseases inferred from GWAS summary data. Nat
519 Commun. Nature Publishing Group; 2018;9: 224.

520 **40**. Nikpay M, Beehler K, Valsesia A, Hager J, Harper M-E, Dent R, et al. Genome-wide
521 identification of circulating-miRNA expression quantitative trait loci reveals the role of
522 several miRNAs in the regulation of cardiometabolic phenotypes. Cardiovasc Res.
523 2019;115: 1629–1645.

524 **41**. Shevchenko A, Tomas H, Havliš J, Olsen J V., Mann M. In-gel digestion for mass
525 spectrometric characterization of proteins and proteomes. Nat Protoc. Nat Protoc; 2007;1:
526 2856–2860.

527

528

Table S1. LD structure of the CAD associated variants proximal and overlapping FN1. Haploreg visualization of the top CAD associated SNP (according to Van der Harst et al; green highlight) and its relationship to other variants (r^2>0.8), overlapping GENCODE genes and dbSNP functional annotation. The missense variant (rs1250259) is highlighted in yellow

Pos (Chr 2; hg38)	LD (r^2)	Variant	Ref	Alt	Ref frequency (EUR)	Gene	dbSNP funct annot.
215422370	0.83	rs1250248	A	G	0.79	FN1	intronic
215423073	0.83	rs13423742	C	G	0.21	FN1	intronic
215427608	0.83	rs1837121	G	A	0.78	FN1	intronic
215430234	0.84	rs1250239	C	G	0.78	FN1	intronic
215430291	0.84	rs1250240	A	G	0.78	FN1	intronic
215430589	0.84	rs1250241	T	A	0.78	FN1	intronic
215431534	0.84	rs1250242	G	C	0.78	FN1	intronic
215433073	0.86	rs1250244	G	C	0.78	FN1	intronic
215434906	0.85	rs1250247	C	G	0.78	FN1	intronic
215435462	0.93	rs1250258	C	T	0.78	FN1	intronic
215435759	0.94	rs1250259	T	A	0.79	FN1	missense
215438330	0.8	rs3910516	A	G	0.76	1.3kb 3' of AC012462.1	
215439661	1	rs1250229	T	C	0.79	2.7kb 3' of AC012462.1	
215441111	0.9	rs1250231	G	A	0.78	4.1kb 3' of AC012462.1	
215441912	1	rs1250232	C	T	0.79	4.9kb 3' of AC012462.1	

Table S2. Association between rs1250259 and metabolic traits and diseases as identified by PhenoScanner (http://www.phenoscanner.medschl.cam.ac.uk/). Note that the effect allele displayed by PhenoScanner for rs1250259 corresponds to the common allele for rs1250259 and the alternate allele for rs1250258. The phased haplotype for the most common alleles are shown in green. Note that the beta for Fibronectin is in unit **decrease. Thus the less common/effect allele (a1) reduces fibronectin levels and is linked to increased CAD.**

snp	rsid	hg19_coordinates	hg38_coordinates	a1 (effect allele)	a2	trait	efo	study	pmid	ancestry	year	beta	se	p	direction	n	n_cases	n_controls	n_studies	unit	dataset
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Nonsyndromic striae distensae stretch n	-	Tung	23633020	European	2013	NA	NA	1.51E-06	NA	33930	-	-	-	-	GRASP
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Blood protein levels [Fibronectin Fragme	EFO_0008140	Suhre K	28240269	European	2017	0.3092	0.05061	1.00E-09	+	-	-	-	-	unit decrease	NHGRI-EBI_GWAS_Catalog
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Blood protein levels [Fibronectin Fragme	EFO_0008140	Suhre K	28240269	European	2017	0.6675	0.0489	2.00E-42	+	-	-	-	-	unit decrease	NHGRI-EBI_GWAS_Catalog
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Blood protein levels [Fibronectin]	EFO_0008140	Suhre K	28240269	European	2017	0.7113	0.04798	1.00E-49	+	-	-	-	-	unit decrease	NHGRI-EBI_GWAS_Catalog
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Deep ovarian andor rectovaginal disease	EFO_0001065	Uimari O	28333195	European	2017	NA	NA	3.00E-08	NA	-	-	-	-	-	NHGRI-EBI_GWAS_Catalog
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Comparative height size at age 10	-	Neale B	UKBB	European	2017	0.008653	0.001884	4.40E-06	+	332021	0	332021	1	-	Neale-B_UKBB_EUR_2017
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Height	EFO_0004339	Neale B	UKBB	European	2017	0.009749	0.001962	6.74E-07	+	336474	0	336474	1	IVNT	Neale-B_UKBB_EUR_2017
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Impedance of leg right	-	Neale B	UKBB	European	2017	0.01191	0.002499	1.86E-06	+	331301	0	331301	1	IVNT	Neale-B_UKBB_EUR_2017
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Systolic blood pressure	EFO_0006335	Neale B	UKBB	European	2017	0.01882	0.002796	1.71E-11	+	317754	0	317754	1	IVNT	Neale-B_UKBB_EUR_2017
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Coronary artery disease	EFO_0000378;EFO_(	van der Ha	29212778	Mixed	2018	0.047	0.007955	3.50E-09	+	296525	34541	261984	1	log OR	van-der-Harst-P_CAD-UKBB_Mixed_2018
rs1250258	rs1250258	chr2:216300185	chr2:215435462	C	T	Coronary artery disease	EFO_0000378;EFO_(	van der Ha	29212778	Mixed	2018	0.0419	0.0065	9.12E-11	+	547261	122733	424528	2	log OR	van-der-Harst-P_CAD_Mixed_2018

snp	rsid	hg19_coordinates	hg38_coordinates	a1 (effect allele)	a2	trait	efo	study	pmid	ancestry	year	beta	se	p	direction	n	n_cases	n_controls	n_studies	unit	dataset
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Low density lipoprotein	EFO_0004611	GLGC	24097068	European	2013	0.0298	0.006	1.46E-06	+	89888	0	89888	60	IVNT	GLGC_LDL_EUR_2013
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Total cholesterol	EFO_0004574	GLGC	24097068	European	2013	0.0257	0.0059	7.56E-06	+	94595	0	94595	60	IVNT	GLGC_TC_EUR_2013
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Nonsyndromic striae distensae stretch n	-	Tung	23633020	European	2013	NA	NA	9.01E-07	NA	33930	-	-	-	-	GRASP
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Coronary artery disease	EFO_0000378	van der Ha	29212778	Mixed	2018	0.048	0.007305	5.00E-11	+	-	-	-	-	unit decrease	NHGRI-EBI_GWAS_Catalog
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Pulse pressure	EFO_0005763	Warren HF	28135244	European	2017	0.314	0.03549	9.00E-19	+	-	-	-	-	unit decrease	NHGRI-EBI_GWAS_Catalog
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Comparative height size at age 10	-	Neale B	UKBB	European	2017	-0.00878	0.001884	3.16E-06	-	332021	0	332021	1	-	Neale-B_UKBB_EUR_2017
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Height	EFO_0004339	Neale B	UKBB	European	2017	-0.00984	0.001961	5.30E-07	-	336474	0	336474	1	IVNT	Neale-B_UKBB_EUR_2017
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Impedance of leg right	-	Neale B	UKBB	European	2017	-0.01167	0.002498	2.97E-06	-	331301	0	331301	1	IVNT	Neale-B_UKBB_EUR_2017
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Systolic blood pressure	EFO_0006335	Neale B	UKBB	European	2017	-0.01885	0.002795	1.57E-11	-	317754	0	317754	1	IVNT	Neale-B_UKBB_EUR_2017
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Coronary artery disease	EFO_0000378;EFO_(	van der Ha	29212778	Mixed	2018	-0.04722	0.007952	2.90E-09	-	296525	34541	261984	1	log OR	van-der-Harst-P_CAD-UKBB_Mixed_2018
rs1250259	rs1250259	chr2:216300482	chr2:215435759	A	T	Coronary artery disease	EFO_0000378;EFO_(	van der Ha	29212778	Mixed	2018	-0.0424	0.0065	6.16E-11	-	547261	122733	424528	2	log OR	van-der-Harst-P_CAD_Mixed_2018

Table S3. Phenotypic studies associated with variations in FN1 expression. Open Targets Genetics was interrogated for studies associated with FN1. The tool identifies SNPs associated with the FN1 term that were linked to various phenotypical traits. All variants are in tight LD (R^2> 0.8). A subset of the genome-wide significant (< 5E-8) associations are shown.

Lead Varian	Study ID	Trait	Lead Varian	Beta	Odds R	PMID	Effect	Common allele
rs1250229	GCST005196	Coronary artery disease	3.00E-19	-0.0644		PMID:29212778	C	T
rs1250229	GCST005194	Coronary artery disease	1.58E-13		0.957	PMID:29212778	C	T
rs1250229	GCST004787	Coronary artery disease (myocardial infarction, percutaneo	3.00E-13		0.934	PMID:28714975	C	T
rs1250229	GCST004233_2	LDL cholesterol levels [Trans-ethnic initial]	2.00E-09	0.0243		PMID:28334899	C	T
rs1250229	GCST003302	Cholesterol, total	1.00E-08	0.023		PMID:26780889	C	T
rs1250229	GCST002222	LDL cholesterol	3.13E-08	0.0243		PMID:24097068	C	T
rs1250231	NEALE2_4080_	Systolic blood pressure, automated reading	8.70E-13	-0.3427			A	G
rs1250247	GCST007096	Pulse pressure	1.00E-21			PMID:27841878	G	C
rs1250247	GCST007099	Systolic blood pressure	1.00E-12			PMID:27841878	G	C
rs1250247	GCST007097	Pulse pressure	7.00E-11			PMID:27841878	G	C
rs1250247	GCST007097_3	Pulse pressure [EA]	3.00E-09			PMID:27841878	G	C
rs1250247	NEALE2_30090	Platelet crit	2.06E-08	-0.0007			G	C
rs1250248	GCST007081	Lung function (FVC)	1.00E-08			PMID:30595370	G	C
rs1250248	GCST004235_2	Total cholesterol levels [Trans-ethnic initial]	5.00E-08	0.0204		PMID:28334899	G	C
rs1250258	GCST004365_2	Blood protein levels [Fibronectin]	1.00E-49	0.7113		PMID:28240269	T	C
rs1250258	GCST004365_2	Blood protein levels [Fibronectin Fragment 3]	2.00E-42	0.6675		PMID:28240269	T	C
rs1250258	GCST007067	Waist-hip ratio	7.00E-12			PMID:30595370	T	C
rs1250258	GCST004365_2	Blood protein levels [Fibronectin Fragment 4]	1.00E-09	0.3092		PMID:28240269	T	C
rs1250258	GCST004370	Deep ovarian and/or rectovaginal disease with dense adhes	3.00E-08			PMID:28333195	T	C
rs1250259	GCST006585_6	Blood protein levels [FN1]	5E-89			PMID:30072576	A	T
rs1250259	GCST006585_1	Blood protein levels [NPNT]	1.00E-59			PMID:30072576	A	T
rs1250259	GCST007087	Systolic blood pressure	2.00E-22			PMID:30595370	A	T
rs1250259	GCST004278	Pulse pressure	9.00E-19			PMID:28135244	A	T
rs1250259	GCST007268	Diastolic blood pressure	1.00E-16			PMID:30578418	A	T
rs1250259	GCST007269	Pulse pressure	3.00E-16			PMID:30578418	A	T
rs1250259	GCST006585_1	Blood protein levels [SSR1]	2.00E-13			PMID:30072576	A	T
rs1250259	GCST006585_1	Blood protein levels [NDUFS4]	5.00E-13			PMID:30072576	A	T
rs1250259	GCST006585_1	Blood protein levels [P4HB]	2.00E-11			PMID:30072576	A	T
rs1250259	GCST007267	Systolic blood pressure	1.00E-09			PMID:30578418	A	T
rs1250259	GCST005195	Coronary artery disease	2.90E-09		0.954	PMID:29212778	A	T
rs1250259	GCST006585_1	Blood protein levels [MUSK]	5.00E-09			PMID:30072576	A	T
rs1250259	GCST006585_1	Blood protein levels [SET]	9.00E-09			PMID:30072576	A	T

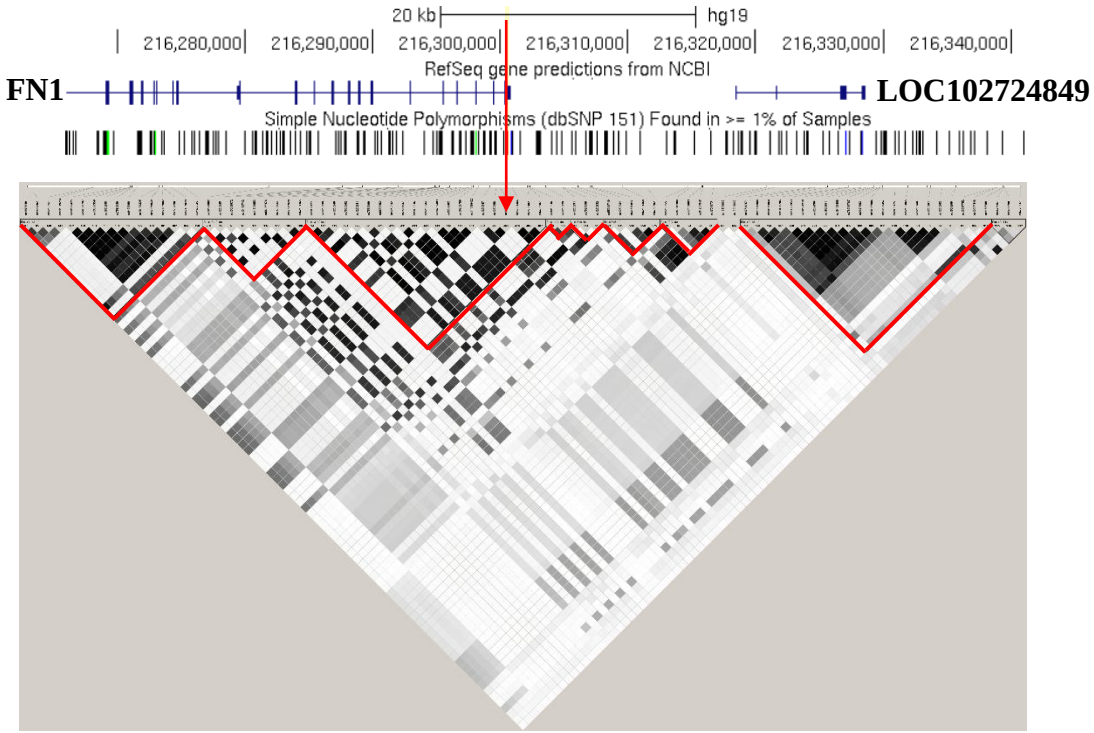

Fig S1. Haploview map with overlapping genes. Region spanning rs1975319 to rs6726337 is shown. LD intensity is proportional to linkage (R^2) values. Blocks were defined using the LD spine method. LD Only common SNPs (frequency > 0.05) are shown. Red arrow points to rs1250259.

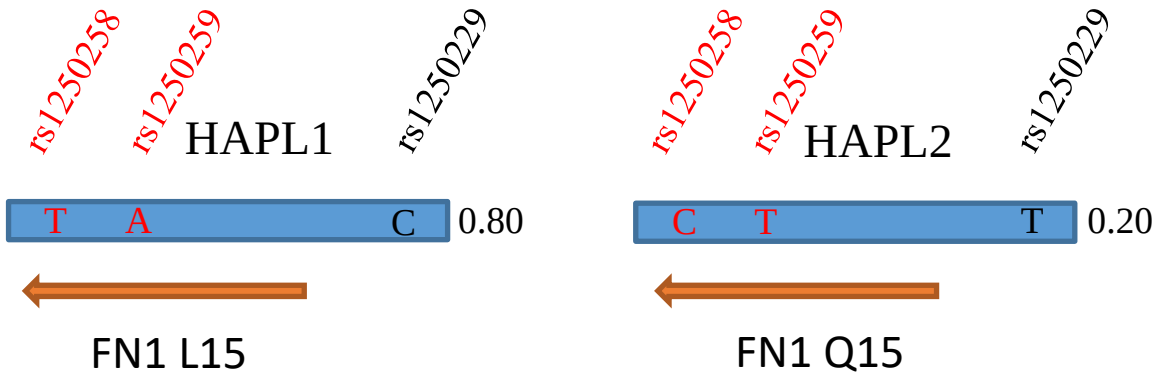

Fig S2. **Haplotype structure around rs1250229, linked to both CAD and blood pressure**. The T allele at rs1250229 (MAF 0.8 in EU), correlates with the presence of rs1250259-A (r^2=0.94), resulting in T on the coding strand of FN1 which is expressed from the negative strand. The corresponding codon encodes a Leucine at position 15 of FN1 while the alternate allele codes for Glutamine. Numbers on the right are the fraction of the corresponding phased haplotype over the total number of observed haplotypes. Values are from the 1000 Genomes Project, using rs1250259 values (all populations). Genotype information for rs125058 and 59 were verified and found to be consistent with the Ottawa cohort genotyping.

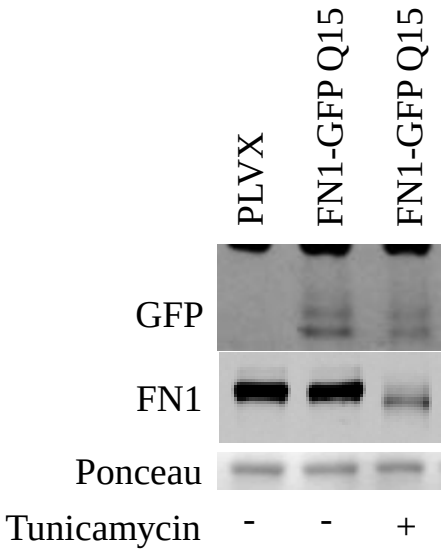

Fig S3. **Secreted $FN1_{1-182}$-GFP is insensitive to Tunicamycin.** HuH-7 cells stably transduced with FN1 $FN1_{1-182}$-GFP were treated for 24 h with 10 µg/ml tunicamycin. Media were harvested and analyzed by Western blotting for FN1-GFP and full-length FN1, as a control for Tunicamycin efficacy. Ponceau stain of a ~ 150 kDa section matching the samples is included as a loading control.

A

	Signal peptide prediction	Cleavage site	Probability of cleavage
L	0.99	TGA-SK (POS 26-27)	0.27
Q	0.97	GTA-VP (POS 20-21)	0.26

B

	Cleavage site	Probability of cleavage
L	STG-AS (POS 25-26)	0.32
Q	STG-AS (POS 25-26)	0.27

Fig S4. Bioinformatic predictions of L15Q variations on signal peptide cleavage. The N-terminal domain of FN1 was analyzed via SignalP -5.0 (http://www.cbs.dtu.dk/services/SignalP/) (**A**) or TargetP-2.0 (http://www.cbs.dtu.dk/services/TargetP/) (**B**) to predict the impact of the Q15L natural variant on processing. Both approaches predict modestly reduced cleavage for the Q15 variant but only SignalP predicts a shift in the cleavage position.

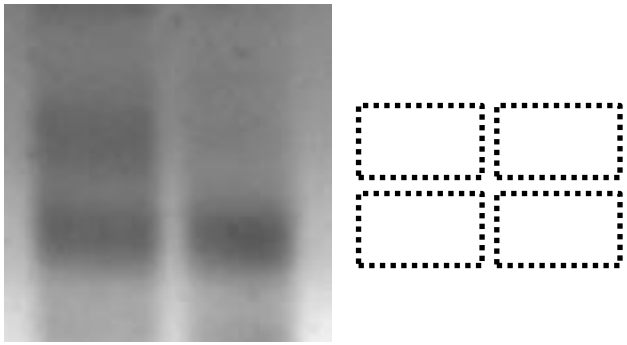

Fig S5. **Representative Coomassie stain of a SDS-PAGE gel of immunoprecipitated FN1-GFP-HA.** Q15 and L15 fusion proteins were isolated from 2.5 ml of media and analyzed by Western blot. Gel pieces derived from the slower (glycosylated) and faster forms were analyzed separately by LC-MS/MS. The boxes indicate the corresponding regions isolated from the gel (for clarity shown on the side of the gel).

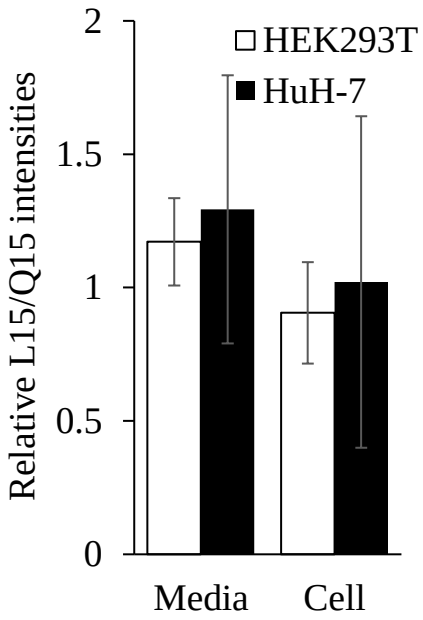

Fig S6. Quantification of L15 and Q15 variants in the media and cell. Data is expressed as the ratio of the L to Q Western blot signals in each compartment. Differences were not statistically significant.

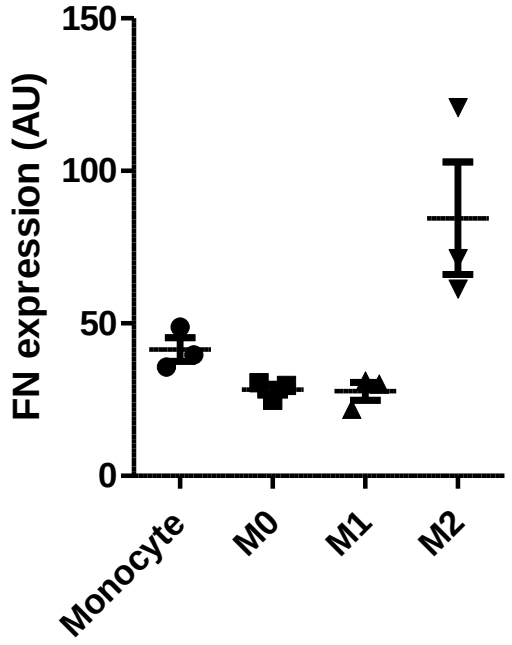

Fig S7. **FN1 expression as a function of polarization status**. Transcription array data from GDS2430 (PMID: 17082649) showing levels of FN1 (Probe ID: 214702_at).

Ab:

GFP (Rabbit; Invitrogen: A11122)

GFP (Rabbit; Sigma: G1544); this Ab showed less non-specific signal than A11122

FN1 (Rabbit; GeneTex : GTX112794)

HA (Mouse; Covance: MMS-101R)

Oligonucleotides :

Mutagenic primers (Q15L)

Forward: CTGGCCGTCCtGTGCCTGGGG

Reverse: CAGCAGCAGCCCGGGCCC

qPCR primers

sFN1

Forward: CCATCATCCCAGCTGTTCCT

Reverse: GTTCGTAGACACTGGAGACAC

pFN1

Forward: CTGCAGTAACCAACATTGATCG

Reverse: TGAGGCCTTGCAGCTCTG

SRP14

Forward: ACTTCCGGCTCTCACTGCTA

Reverse: TCAAAGCCCTCCACAGTACC

pLVXFN1_1-182GFPHA

Short FN1 GFP1 expression construct

Q15L substitution highlighted in red

GFP moiety in green

HA tag in red

TGGAAGGGCTAATTCACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCC
TGATTAGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACC
AGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATG
GGATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGA
GAGCTGCATCCGGAGTACTTCAAGAACTGCTGATATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAG
GGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCCTGT
ACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTC
AATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAG
ACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGA
GGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGT
ACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGA
GAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTA
TGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATA
CTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCT
ATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAA
AGTAAGACCACCGCACAGCAAGCGGCCGGCCGCTGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGG
AGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGA
GTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCAC
TATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAA
TTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAG
AATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTG
CACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATG
GAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAA
AAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTG
TGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGT
GAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCC
CGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCTCGACGG
TATCGCCTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACA
GACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAG
ATCCAGTTTATCGATAAGCTTGGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT
GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGA
CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTG
ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT
TCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG
CAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGA

CTCTACTAGAGGATCGCTAGCGCTACCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTCAACATGCTTAGGGGT
CCGGGGCCCGGGCTGCTGCTGCTGGCCGTCCAGTGCCTGGGGACAGCGGTGCCCTCCACGGGAGCCTCGAAGAG
CAAGAGGCAGGCTCAGCAAATGGTTCAGCCCCAGTCCCCGGTGGCTGTCAGTCAAAGCAAGCCCGGTTGTTATGA
CAATGGAAAACACTATCAGATAAATCAACAGTGGGAGCGGACCTACCTAGGCAATGCGTTGGTTTGTACTTGTTAT
GGAGGAAGCCGAGGTTTTAACTGCGAGAGTAAACCTGAAGCTGAAGAGACTTGCTTTGACAAGTACACTGGGAA
CACTTACCGAGTGGGTGACACTTATGAGCGTCCTAAAGACTCCATGATCTGGGACTGTACCTGCATCGGGGCTGG
GCGAGGGAGAATAAGCTGTACCATCGCAAACCGCTGCCATGAAGGGGGTCAGTCCTACAAGATTGGTGACACCT
GGAGGAGACCACATGAGACTGGTGGTTACATGTTAGAGTGTGTGTCTTGGTAATGGAAAAGGAGAATGGACC
TGCAAGCCCATAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGG
CGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGA
AGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTG
CTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG
AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTA
CAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT
GCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGAGGTGCCGCCGGATACC
CTTATGATGTGCCAGATTATGCCTGAGGATCCCGCGACTCTAGATAATTCTACCGGGTAGGGGAGGCGCTTTTCCC
AAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCAC
ACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCCT
AGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAG
TCTCGTGCAGATGGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTT
GCTCCTTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGG
CGGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCGCGCTGTT
CTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCAGCCCAAGCTTACCATGACCGAGTACAAGCCCACGGTGCGCCT
CGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCA
CACCGTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGCTCGA
CATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCG
GGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGAT
GGAAGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCGCCCGACCA
CCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCC
TTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGG
TGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGACCGCGTCTGGAACAATCAACCTCTG
GATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTA
ATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTT
ATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTT
GGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATC
GCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAG
CTGACGTCCTTTCCATGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTC
GGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCC
CTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGGAATTAATTCTGCAGTCGAGACCTAGAAAAACAT
GGAGCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGATTGTGCCTGGCTAGAAGCACAAGAGGAGGAGGA
GGTGGGTTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTT

TAAAAGAAAAGAGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACC
ACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATG
GTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTAC
ACCCTGTGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGATATCGAGCTTGCTACAAGGGACTTT
CCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAA
GCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGG
AACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGG
TAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTAGTAGTTCATGTCATCTTATTATTC
AGTATTTATAACTTGCAAAGAAATGAATATCAGAGAGTGAGAGGCCTTGACATTGCTAGCGTTTTACCGTCGACCT
CTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC
ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGC
TCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGC
GGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGG
TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA
AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACG
AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCC
CTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTG
TGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTT
CTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC
TTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA
ACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAA
TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCAC
CTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAG
GGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAA
ACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGT
TGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCAT
GGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAA
CCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
CACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCT
GTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGG
GTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGA
AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCAAC
TTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTG
CATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCAACTGGATAACTCAAGCTAACCA
AAATCATCCCAAACTTCCCACCCCATACCCTATTACCACTGCCAATTACCTGTGGTTTCATTTACTCTAAACCTGTGA
TTCCTCTGAATTATTTTCATTTTAAAGAAATTGTATTTGTTAAATATGTACTACAAACTTAGTAGTTTTTAAAGAAATT
GTATTTGTTAAATATGTACTACAAACTTAGTAGT

pMAX pFN1HA

Full-length pFN1 construct with C-terminal HA tag

HA tag is red

N and C-terminal pFN1 amino acids are in green

Hind III site for screening purposes is highlighted

Q15L substitution is highlighted

TCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTT
GTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTGATTATTGACTAG
TTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAA
ATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCC
AATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCT
TACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC
ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT
TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGC
GTGTACGGTGGGAGGTCTATATAAGCAGAGGTCGTTTAGTGAACCGTCAGATCACTAGTAGCTTTATTGCGGTAG
TTTATCACAGTTAAATTGCTAACGCAGTCAGTGCTCGACTGATCACAGGTAAGTATCAAGGTTACAAGACAGGTTT
AAGGAGGCCAATAGAAACTGGGCTTGTCGAGACAGAGAAGATTCTTGCGTTTCTGATAGGCACCTATTGGTCTTA
CTGACATCCACTTTGCCTTTCTCTCCACAGGGGTACCGCCATCATGAAGTTTAAACAAGCTTGAATTCTCTAGAGAT
ATCCTGCAGAGATCTGGATCCCTCGAGGCTAGCTGTCAACATGCTTAGGGGTCCGGGGCCCGGGCTGCTGCTGCT
GGCCGTCC(A/T)GTGCCTGGGGACAGCGGTGCCCTCCACGGGAGCCTCGAAGAGCAAGAGGCAGGCTCAGCAAA
TGGTTCAGCCCCAGTCCCCGGTGGCTGTCAGTCAAAGCAAGCCCGGTTGTTATGACAATGGAAAACACTATCAGAT
AAATCAACAGTGGGAGCGGACCTACCTAGGCAATGCGTTGGTTTGTACTTGTTATGGAGGAAGCCGAGGTTTTAA
CTGCGAGAGTAAACCTGAAGCTGAAGAGACTTGCTTTGACAAGTACACTGGGAACACTTACCGAGTGGGTGACAC
TTATGAGCGTCCTAAAGACTCCATGATCTGGGACTGTACCTGCATCGGGGCTGGGCGAGGGAGAATAAGCTGTAC
CATCGCAAACCGCTGCCATGAAGGGGGTCAGTCCTACAAGATTGGTGACACCTGGAGGAGACCACATGAGACTG
GTGGTTACATGTTAGAGTGTGTGTGTCTTGGTAATGGAAAAGGAGAATGGACCTGCAAGCCCATAGCTGAGAAGT
GTTTTGATCATGCTGCTGGGACTTCCTATGTGGTCGGAGAAACGTGGGAGAAGCCCTACCAAGGCTGGATGATGG
TAGATTGTACTTGCCTGGGAGAAGGCAGCGGACGCATCACTTGCACTTCTAGAAATAGATGCAACGATCAGGACA
CAAGGACATCCTATAGAATTGGAGACACCTGGAGCAAGAAGGATAATCGAGGAAACCTGCTCCAGTGCATCTGCA
CAGGCAACGGCCGAGGAGAGTGGAAGTGTGAGAGGCACACCTCTGTGCAGACCACATCGAGCGGATCTGGCCCC
TTCACCGATGTTCGTGCAGCTGTTTACCAACCGCAGCCTCACCCCCAGCCTCCTCCCTATGGCCACTGTGTCACAGA
CAGTGGTGTGGTCTACTCTGTGGGGATGCAGTGGCTGAAGACACAAGGAAATAAGCAAATGCTTTGCACGTGCCT
GGGCAACGGAGTCAGCTGCCAAGAGACAGCTGTAACCCAGACTTACGGTGGCAACTCAAATGGAGAGCCATGTG
TCTTACCATTCACCTACAATGGCAGGACGTTCTACTCCTGCACCACAGAAGGGCGACAGGACGGACATCTTTGGTG
CAGCACAACTTCGAATTATGAGCAGGACCAGAAATACTCTTTCTGCACAGACCACACTGTTTTGGTTCAGACTCGA
GGAGGAAATTCCAATGGTGCCTTGTGCCACTTCCCCTTCCTATACAACAACCACAATTACACTGATTGCACTTCTGA
GGGCAGAAGAGACAACATGAAGTGGTGTGGGACCACACAGAACTATGATGCCGACCAGAAGTTTGGGTTCTGCC

CCATGGCTGCCCACGAGGAAATCTGCACAACCAATGAAGGGGTCATGTACCGCATTGGAGATCAGTGGGATAAGC
AGCATGACATGGGTCACATGATGAGGTGCACGTGTGTTGGGAATGGTCGTGGGGAATGGACATGCATTGCCTACT
CGCAGCTTCGAGATCAGTGCATTGTTGATGACATCACTTACAATGTGAACGACACATTCCACAAGCGTCATGAAGA
GGGGCACATGCTGAACTGTACATGCTTCGGTCAGGGTCGGGGCAGGTGGAAGTGTGATCCCGTCGACCAATGCC
AGGATTCAGAGACTGGGACGTTTTATCAAATTGGAGATTCATGGGAGAAGTATGTGCATGGTGTCAGATACCAGT
GCTACTGCTATGGCCGTGGCATTGGGGAGTGGCATTGCCAACCTTTACAGACCTATCCAAGCTCAAGTGGTCCTGT
CGAAGTATTTATCACTGAGACTCCGAGTCAGCCCAACTCCCACCCCATCCAGTGGAATGCACCACAGCCATCTCAC
ATTTCCAAGTACATTCTCAGGTGGAGACCTAAAAATTCTGTAGGCCGTTGGAAGGAAGCTACCATACCAGGCCACT
TAAACTCCTACACCATCAAAGGCCTGAAGCCTGGTGTGGTATACGAGGGCCAGCTCATCAGCATCCAGCAGTACG
GCCACCAAGAAGTGACTCGCTTTGACTTCACCACCACCAGCACCAGCACACCTGTGACCAGCAACACCGTGACAGG
AGAGACGACTCCCTTTTCTCCTCTTGTGGCCACTTCTGAATCTGTGACCGAAATCACAGCCAGTAGCTTTGTGGTCT
CCTGGGTCTCAGCTTCCGACACCGTGTCGGGATTCCGGGTGGAATATGAGCTGAGTGAGGAGGGAGATGAGCCA
CAGTACCTGGATCTTCCAAGCACAGCCACTTCTGTGAACATCCCTGACCTGCTTCCTGGCCGAAAATACATTGTAAA
TGTCTATCAGATATCTGAGGATGGGGAGCAGAGTTTGATCCTGTCTACTTCACAAACAACAGCGCCTGATGCCCCA
CCTGACCCGACTGTGGACCAAGTTGATGACACCTCAATTGTTGTTCGCTGGAGCAGACCCCAGGCTCCCATCACAG
GGTACAGAATAGTCTATTCGCCATCAGTAGAAGGTAGCAGCACAGAACTCAACCTTCCTGAAACTGCAAACTCCGT
CACCCTCAGTGACTTGCAACCTGGTGTTCAGTATAACATCACTATCTATGCTGTGGAAGAAAATCAAGAAAGTACA
CCTGTTGTCATTCAACAAGAAACCACTGGCACCCCACGCTCAGATACAGTGCCCTCTCCCAGGGACCTGCAGTTTG
TGGAAGTGACAGACGTGAAGGTCACCATCATGTGGACACCGCCTGAGAGTGCAGTGACCGGCTACCGTGTGGAT
GTGATCCCCGTCAACCTGCCTGGCGAGCACGGGCAGAGGCTGCCCATCAGCAGGAACACCTTTGCAGAAGTCACC
GGGCTGTCCCCTGGGGTCACCTATTACTTCAAAGTCTTTGCAGTGAGCCATGGGAGGGAGAGCAAGCCTCTGACT
GCTCAACAGACAACCAAACTGGATGCTCCCACTAACCTCCAGTTTGTCAATGAAACTGATTCTACTGTCCTGGTGA
GATGGACTCCACCTCGGGCCCAGATAACAGGATACCGACTGACCGTGGGCCTTACCCGAAGAGGACAGCCCAGG
CAGTACAATGTGGGTCCCTCTGTCTCCAAGTACCCACTGAGGAATCTGCAGCCTGCATCTGAGTACACCGTATCCCT
CGTGGCCATAAAGGGCAACCAAGAGAGCCCCAAAGCCACTGGAGTCTTTACCACACTGCAGCCTGGGAGCTCTAT
TCCACCTTACAACACCGAGGTGACTGAGACCACCATTGTGATCACATGGACGCCTGCTCCAAGAATTGGTTTTAAG
CTGGGTGTACGACCAAGCCAGGGAGGAGAGGCACCACGAGAAGTGACTTCAGACTCAGGAAGCATCGTTGTGTC
CGGCTTGACTCCAGGAGTAGAATACGTCTACACCATCCAAGTCCTGAGAGATGGACAGGAAAGAGATGCGCCAAT
TGTAAACAAAGTGGTGACACCATTGTCTCCACCAACAAACTTGCATCTGGAGGCAAACCCTGACACTGGAGTGCTC
ACAGTCTCCTGGGAGAGGAGCACCACCCCAGACATTACTGGTTATAGAATTACCACAACCCCTACAAACGGCCAGC
AGGGAAATTCTTTGGAAGAAGTGGTCCATGCTGATCAGAGCTCCTGCACTTTTGATAACCTGAGTCCCGGCCTGGA
GTACAATGTCAGTGTTTACACTGTCAAGGATGACAAGGAAAGTGTCCCTATCTCTGATACCATCATCCCAGCTGTTC
CTCCTCCCACTGACCTGCGATTCACCAACATTGGTCCAGACACCATGCGTGTCACCTGGGCTCCACCCCCATCCATT
GATTTAACCAACTTCCTGGTGCGTTACTCACCTGTGAAAAATGAGGAAGATGTTGCAGAGTTGTCAATTTCTCCTTC
AGACAATGCAGTGGTCTTAACAAATCTCCTGCCTGGTACAGAATATGTAGTGAGTGTCTCCAGTGTCTACGAACAA
CATGAGAGCACACCTCTTAGAGGAAGACAGAAAACAGGTCTTGATTCCCCAACTGGCATTGACTTTTCTGATATTA
CTGCCAACTCTTTTACTGTGCACTGGATTGCTCCTCGAGCCACCATCACTGGCTACAGGATCCGCCATCATCCCGAG
CACTTCAGTGGGAGACCTCGAGAAGATCGGGTGCCCCACTCTCGGAATTCCATCACCCTCACCAACCTCACTCCAG
GCACAGAGTATGTGGTCAGCATCGTTGCTCTTAATGGCAGAGAGGAAAGTCCCTTATTGATTGGCCAACAATCAAC
AGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTGGGATGCTCCT
GCTGTCACAGTGAGATATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGGAGTTCACTGTG
CCTGGGAGCAAGTCTACAGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACTGTGTATGCTGTCA
CTGGCCGTGGAGACAGCCCCGCAAGCAGCAAGCCAATTTCCATTAATTACCGAACAGAAATTGACAAACCATCCCA
GATGCAAGTGACCGATGTTCAGGGCAACAGCATTAGTGTCAAGTGGCTGCCTTCAAGTTCCCCTGTTACTGGTTAC

AGAGTAACCACCACTCCCAAAAATGGACCAGGACCAACAAAAACTAAAACTGCAGGTCCAGATCAAACAGAAATG
ACTATTGAAGGCTTGCAGCCCACAGTGGAGTATGTGGTTAGTGTCTATGCTCAGAATCCAAGCGGAGAGAGTCAG
CCTCTGGTTCAGACTGCAGTAACCACTATTCCTGCACCAACTGACCTGAAGTTCACTCAGGTCACACCCACAAGCCT
GAGCGCCCAGTGGACACCACCCAATGTTCAGCTCACTGGATATCGAGTGCGGGTGACCCCCAAGGAGAAGACCG
GACCAATGAAAGAAATCAACCTTGCTCCTGACAGCTCATCCGTGGTTGTATCAGGACTTATGGTGGCCACCAAATA
TGAAGTGAGTGTCTATGCTCTTAAGGACACTTTGACAAGCAGACCAGCTCAGGGTGTTGTCACCACTCTGGAGAAT
GTCAGCCCACCAAGAAGGGCTCGTGTGACAGATGCTACTGAGACCACCATCACCATTAGCTGGAGAACCAAGACT
GAGACGATCACTGGCTTCCAAGTTGATGCCGTTCCAGCCAATGGCCAGACTCCAATCCAGAGAACCATCAAGCCA
GATGTCAGAAGCTACACCATCACAGGTTTACAACCAGGCACTGACTACAAGATCTACCTGTACACCTTGAATGACA
ATGCTCGGAGCTCCCCTGTGGTCATCGACGCCTCCACTGCCATTGATGCACCATCCAACCTGCGTTTCCTGGCCACC
ACACCCAATTCCTTGCTGGTATCATGGCAGCCGCCACGTGCCAGGATTACCGGCTACATCATCAAGTATGAGAAGC
CTGGGTCTCCTCCCAGAGAAGTGGTCCCTCGGCCCCGCCCTGGTGTCACAGAGGCTACTATTACTGGCCTGGAACC
GGGAACCGAATATACAATTTATGTCATTGCCCTGAAGAATAATCAGAAGAGCGAGCCCCTGATTGGAAGGAAAAA
GACAGACGAGCTTCCCCAACTGGTAACCCTTCCACACCCCAATCTTCATGGACCAGAGATCTTGGATGTTCCTTCCA
CAGTTCAAAAGACCCCTTTCGTCACCCACCCTGGGTATGACACTGGAAATGGTATTCAGCTTCCTGGCACTTCTGGT
CAGCAACCCAGTGTTGGGCAACAAATGATCTTTGAGGAACATGGTTTTAGGCGGACCACACCGCCCACAACGGCC
ACCCCCATAAGGCATAGGCCAAGACCATACCCGCCGAATGTAGGTGAGGAAATCCAAATTGGTCACATCCCCAGG
GAAGATGTAGACTATCACCTGTACCCACACGGTCCGGGACTCAATCCAAATGCCTCTACAGGACAAGAAGCTCTCT
CTCAGACAACCATCTCATGGGCCCCATTCCAGGACACTTCTGAGTACATCATTTCATGTCATCCTGTTGGCACTGAT
GAAGAACCCTTACAGTTCAGGGTTCCTGGAACTTCTACCAGTGCCACTCTGACAGGCCTCACCAGAGGTGCCACCT
ACAACATCATAGTGGAGGCACTGAAAGACCAGCAGAGGCATAAGGTTCGGGAAGAGGTTGTTACCGTGGGCAAC
TCTGTCAACGAAGGCTTGAACCAACCTACGGATGACTCGTGCTTTGACCCCTACACAGTTTCCCATTATGCCGTTGG
AGATGAGTGGGAACGAATGTCTGAATCAGGCTTTAAACTGTTGTGCCAGTGCTTAGGCTTTGGAAGTGGTCATTTC
AGATGTGATTCATCTAGATGGTGCCATGACAATGGTGTGAACTACAAGATTGGAGAGAAGTGGGACCGTCAGGG
AGAAAATGGCCAGATGATGAGCTGCACATGTCTTGGGAACGGAAAAGGAGAATTCAAGTGTGACCCTCATGAGG
CAACGTGTTATGATGATGGGAAGACATACCACGTAGGAGAACAGTGGCAGAAGGAATATCTCGGTGCCATTTGCT
CCTGCACATGCTTTGGAGGCCAGCGGGGCTGGCGCTGTGACAACTGCCGCAGACCTGGGGGTGAACCCACTCCCG
AAGGCACTACTGGCCAGTCCTACAACCAGTATTCTCAGAGATACCATCAGAGAACAAACACTAATGTTAATTGCCC
AATTGAGTGCTTCATGCCTTTAGATGTACAGGCTGACAGAGAAGATTCCCGAGAG**GGATACCCTTATGATGTGCC**
AGATTATGCCTGAAGCTT*GTGCCAGATTATGCCTGATTTAAACAGAGCTCGATGAGTTTGGACAAACCACAACTAG*
AATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATA
AACAAGTTAACAACAACAATTGCATTCATTTTATTAAGGCCTCACGTGACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGA
CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACGGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATA
GCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAG
*CAGCCACTGGTAACAGGATT*AGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACT
ACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTA
GCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGG
GATTTTGGTCATGCCGTCTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGG
CGATACCGTAAAGCACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACG
CTATGTCCTGATAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCAT

GATATTCGGCAAGCAGGCATCGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCTCGCCTTGAGCCTGGC
GAACAGTTCGGCTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGA
GTACGTGCTCTCTCGATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGC
CGCATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACT
TCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGTACAGCTGCGCAAGGAACGCCCGTCGTG
GCCAGCCACGATAGCCGCGCTGCCTCGTCTTGCAGTTCATTCAGGGCACCGGACAGGTCGGTCTTGACAAAAAGA
ACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCATAG
CCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCATAATATTATTGAAGC
ATTTATCAGGGTTCGTCTCGTCCCGGTCTCCTCCCATGCATG