THESIS TITLE

EFFECTS OF RNA POLYMORPHISMS ON *RPS26* (*RIBOSOMAL PROTEIN S26)* TRANSLATIONAL EFFICIENCY

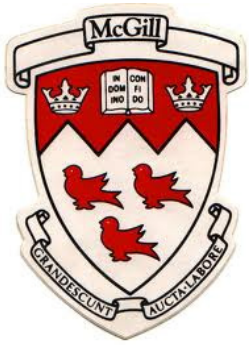
Angeliki Makri

Faculty of Medicine

Department of Human Genetics

McGill University, Montreal, Quebec, Canada

April 2012



A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

© ANGELIKI MAKRI 2012

Table of contents

1.1 Abstract

Background: The effect of common polymorphisms on transcription levels of a large number of genes has been well documented and defined genome-wide in numerous studies. However, the effects of common mRNA polymorphisms on translation have received little attention. *RPS26* exhibited one of the strongest associations ($p = 10^{-10}$) in our group's genome-wide search to identify transcripts whose relative distribution between heavy polysomes (active translation) and soluble RNA (inactive) was skewed by single-nucleotide polymorphisms. The *RPS26* SNPs most strongly associated with ribosomal distribution, rs17118262 (C/G) and rs1131017(C/G), are located 9 bases apart, are in strong linkage disequilibrium (D'=1) but have different allele frequencies (r^2=0.54). The goal of the present study was to confirm and quantify the effect of those *RPS26* SNPs on translational efficiency at the protein level. *RPS26* was selected for this purpose because rs1131017 is also associated with type 1 diabetes (T1D), the second strongest novel association in our genome-wide association study (GWAS).

Methods: To independently confirm the translational imbalance found by ribosomal distribution, we studied allelic ribosomal distribution in single heterozygous samples (allelic translation) with Single Nucleotide Primer Extension (SNuPE) and Pyrosequencing. To test the hypothesis that mRNA with the allele most associated with ribosomes also produces more *RPS26* protein, we applied both an *ex vivo* direct measurement and an *in vitro* translation approach.

Results: Pyrosequencing and SNuPE confirmed the translational imbalance favouring the G allele of the rs17118262. Both *ex vivo* and *in vitro* translation showed that the same allele

that was found at a higher proportion in the polyribosomal fraction than in the soluble RNA produces more protein product under the same translational environment.

Conclusion: Both SNPs (the common rs1131017 and the rare rs17118262 variant) exert a significant effect on the amount of *RPS26* protein that is produced. Adding to the well-studied polymorphic transcriptional effects, the effects of exonic single-nucleotide polymorphisms (SNPs) on translational efficiency may be equally important in determining protein expression levels and therefore contributing significantly to the disease risk. *RPS26* in type 1 diabetes is an excellent example of such an association.

1.2 Resume

L'effet des polymorphismes fréquents (SNPs) sur les niveaux de transcription d'un grand nombre de gènes a été bien documenté et défini à l'échelle du génome dans de nombreuses études. Toutefois, les effets des polymorphismes communs sur la traduction de l'ARN messager (ARNm) ont reçus peu d'attention. Dans notre étude à l'échelle du génome qui étudiait des polymorphismes donnant une distribution différentielle entre les polysomes lourds (traduction active) et de l'ARN soluble (inactif), le gène *RPS26* a montré une des plus fortes associations ($p = 10^{-10}$). Les SNPs dans *RPS26* les plus fortement associés à une différence de distribution des ribosomes, rs17118262 (C / G) et rs1131017 (C / G), sont séparés par 9 bases et sont en fort déséquilibre de liaison (D '= 1) mais ont des fréquences alléliques différentes (r^2=0.54). L'objectif de la présente étude était de confirmer et de quantifier l'effet de ces SNP dans *RPS26* sur son efficacité de traduction au niveau protéique.

Méthodes: Pour confirmer de manière indépendante le déséquilibre constaté au niveau de la distribution des ribosomes, nous avons étudié la distribution des allèles dans des échantillons hétérozygotes pour chaque un des 2 SNPs avec à l'aide des techniques d'extension nucléotidique simple d'amorces (SNuPE) et de pyroséquençage. Pour tester l'hypothèse que l'ARNm avec un allèle particulier produit plus de protéines *RPS26*, nous avons appliqué à la fois une approche de quantification *ex vivo* et traduction *in vitro*.

Résultats: Le SNuPE et le pyroséquençage ont confirmé le déséquilibre au niveau de la traduction montrant une supériorité de l'allèle G du SNP rs17118262. Les approches *ex vivo* et *in vitro* ont montré que la traduction de l'allèle trouvé dans une proportion plus élevée dans la fraction de polysomes lourds produit plus de protéines par rapport à l'autre allèle dans les même conditions.

Conclusion: Les deux SNPs (le rs1131017 qui est fréquent et le rs17118262, de moindre fréquence) exercent un effet significatif sur la quantité de protéine *RPS26* produite. Combiné à l'effet déjà connu des SNPs sur les niveaux de transcription, leur effet sur la variation du niveau de traduction est peut-être aussi important dans la détermination du niveau total de protéines et par le fait même contribuer au risque de maladie de manière significative. *RPS26* au niveau du diabète de type I en est un parfait exemple.

2. Abbreviations

AI: allelic imbalance

AREs: AU-rich elements

BCA: Bicinchoninic Acid

BSA: Bovine Serum Albumin

cDNA: complementary DNA

CEPH: Centre de l'Étude du Polymorphisme Humain

CEU: HapMap samples-Utah residents with ancestry from northern and western Europe

CHOP: Children's Hospital of Philadelphia

CPE: Cytoplasmic Polyadenylation Element

CPEB: Cytoplasmic Polyadenylation Element Binding

DBA: Diamond Blackfan Anemia

EDEN: Embryo Deadenylation Element

eIFs: eukaryotic Initiation Factors

eQTL: expression Quantitative Trait Locus

FBS: Fetal Bovine Serum

GWAS: Genome-Wide association Study

HapMap: Haplotype Map

HLC: Human Liver Cohort

IRE: Iron Responsive Element

IRES: Internal Ribosome Entry Site

IRP: Iron Regulatory Protein

LD: Linkage Disequilibrium

LCL: Lymphoblastoid Cell Lines

miRNA: microRNA

MLPA: Multiplex Ligation-dependent Probe Amplification

mRNA: messenger RNA

M.Sc: Master of Science

m^7G cap: 7-methylguanosine cap structure

OD: Optical density

PABP: poly(A) binding protein

PBS: Phosphate Buffered Saline

PPAR: Peroxisome Proliferators-Activated Receptors

RNA: RiboNucleic Acid

RPS26: Ribosomal Protein S26

SDS: Sodium Dodecyl Sulfate

siRNAs: small interfering RNAs

SNP: Single Nucleotide Polymorphism

SNuPE: Single Nucleotide Primer Extension

T1D: Type 1 Diabetes

TI: Translational Imbalance

tRNA: transfer RNA

uAUGs: upstream AUGs

uORFs: upstream Open-Reading Frames

WTCCC: Wellcome Trust Case Control Consortium

5'UTR: 5' Untranslated Region

3. Introduction and literature review

3.1 Ribosomal Association as a proxy for translational efficiency

The effect of common polymorphisms on transcription levels of a large number of genes has been well documented and defined genome-wide in numerous studies [1]. However, the effects of common mRNA polymorphisms on translation have received little attention. We hypothesised that a non-negligible part of human variation in expression at the protein level is due to such effects. Effects of exonic single-nucleotide polymorphisms (SNPs) on translational efficiency may be equally important in determining protein expression levels but, to date, they have never been systematically examined because possible methodologies are much more complex.

A well-known method to study translational efficiency is the use of ribosomal fractions. Polyribosome or polysome profile analysis uses sucrose density gradient to separate polysomes (mRNAs bound to multiple ribosomes) from monosomes (mRNAs bound to a single ribosome). Monosomes are translationally much less active than polyribosome fractions, which represent the active part of the translation. Ribosomal populations can then be isolated and quantified [2]. After separation of extracts, RNA from the heavy (polyribosomal) and light (monosomal) ribosomal fractions is recovered [3,4] enabling downstream applications to investigate the role of translational control in regulating gene expression [5].

Using ribosomal association as a proxy for translational efficiency, my supervisor's laboratory screened 17,495 autosomal genes by transcriptome profiling of ultracentrifugal ribosomal fractions of cell lines from 40 individuals from the CEU set of HapMap, to identify transcripts whose relative distribution between heavy polysomes (active

translation) and soluble RNA (inactive) was skewed by single-nucleotide polymorphisms. Our analysis revealed 72 genes whose ribosomal distribution was strongly associated with at least one SNP, at a significance level that met the Bonferroni threshold of 4×10^{-9}, for the number of gene-SNP pairs examined. A total of 1,007 genes passed the less stringent FDR threshold of q=0.05 ($p<2.4 \times 10^{-5}$). One of the strongest associations ($p = 10^{-10}$) was with a 5'UTR SNP of the Ribosomal Protein S26 *(RPS26).* The exact same SNP also had the most significant association with T1D at the 12q13 locus, the second strongest novel association in our genome-wide association study (GWAS) [6], also found independently in the GWAS by the Wellcome Trust Case-Control Consortium [7]. Locus 12q13 is known to be one of the most strongly associated loci with type 1 diabetes, thus making *RPS26* an interesting candidate for further downstream investigations.

Homo sapiens ribosomal protein S26 *(RPS26)* is a 2,322 bp gene located on chromosome 12q13 that encodes a protein component of the ribosomal 40S subunit *(gene ID 6231, NCBI Reference Sequence: NG_023201.1).* The 699 bp mRNA transcript includes 4 exons *(NCBI Reference Sequence: NM_001029.3).* There are totally 6 SNPs in the 5'UTR region, 32 intronic SNPs, and 14 exonic SNPs (5 missense, 6 synonymous, 2 nonsense). The SNPs most strongly associated with ribosomal distribution, rs17118262 (C/G) and rs1131017(C/G), are located 9 bases apart, are in strong linkage disequilibrium (D'=1) but have different allele frequencies (r^2=0.54). The rs17118262 is less common than rs1131017, which is found more commonly in the general population. The goal of my M.Sc. project was to confirm and quantify the effect of those SNPs on translational efficiency at the protein level.

3.2 Translational control

Translation is the process of converting a DNA sequence into a polypeptide sequence. It occurs in three consecutive steps: initiation, elongation and termination. Numerous studies have investigated the translational machinery of the eukaryotic cells [8-12]. Briefly, the small ribosomal subunit (40S) and the initiator tRNA assemble to form the 80S initiation complex. Critical in this stage is the recognition of the 7-methylguanosine cap structure (m^7G cap) located in the 5′ end of the mRNA by the eukaryotic initiation factor 4F (eIF4F) complex. The elongation stage includes three functional binding sites on the 80S ribosome that allow the peptide formation: the A (aminoacyl) site where the aminoacyl-tRNA binds, the P (peptidyl) site where a peptide bond forms between the new amino acid and the peptidyl-tRNA and finally the E (exit) site where the tRNA is released. At the termination stage, one of the three stop codons (UAA, UGA, or UAG), binds to the A site and the completed protein is released from the ribosome.

Regulation of the translational process is complicated and occurs during all the three steps; however, the well-studied mechanisms of translational control are found in the initiation stage [12,13]. In the following sections, we will analyze how three known structural elements (5'untranslated region, Kozak sequence and 3'untranslated region) affect translational efficiency, providing examples of genetic variants with known clinical effects in each case.

3.2 A. The 5'UTR in translational control

Several elements in the 5'UTR affect translation regulation: upstream open-reading frames (uORF), stem-loop structures, length of 5'UTR, number of start codons and finally the internal ribosome entry site.

As mentioned before, translation starts with the recognition of the m^7G cap by the eIF4F complex. The eIF4F complex includes the cap-binding protein eIF4E, the RNA helicase eIF4A and the scaffolding protein eIF4G [14,15]. eIF4G binds to the poly(A) binding protein (PABP) which coats the 3′ poly(A) tail of an mRNA [16,17] causing circularization of the mRNA in a closed loop. This closed-loop promotes recruitment of the 43S pre-initiation complex, a complex of the 40S small ribosomal subunit and specific associating factors. The RNA helicase eIF4A unfolds secondary structures in the mRNA's 5′UTR [14] which further enhances binding of the 43S pre-initiation complex to the mRNA. The mRNA is then scanned by the pre-initiation complex moving along the RNA in the 5'-3' direction until an appropriate start codon (AUG) is found [10]. As soon as AUG is recognized in the scanning sequence, 60S ribosomal subunit is assembled leading to the initiation of protein synthesis [9].

Upstream Open-Reading Frames (uORF) are well-recognized *cis*-acting elements that negatively affect translation. Presence of starting codons upstream of the physiological start site (uAUG) defines short upstream ORFs. These uAUGs do not allow the ribosome to reach the physiological AUG resulting in premature initiation and translation inhibition [8,9]. Translation of many known genes is partially regulated by the presence of uORFs in the 5'UTR, including several proto-oncogenes and cytokines; interleukin 7, interleukin 15, tgf-b3 and bcl-2 [18-21].

A well-studied example of this association is the translation of the thrombopoietin gene (*TPO*). Presence of several uAUG codons in the 5'UTR strongly inhibits translation while directed mutagenesis of all uAUGs has been shown to restore full translational efficiency [22]. Normal *TPO* serum concentrations are very low preventing potent activation of platelet production. However, a gain-of-function mutation in the *TPO* 5'UTR abolishes an uORF that disrupts the normal translational repression of *TPO*, thus causing dramatic increase in platelet production leading to the disease hereditary thrombocythemia [22,23].

Stem-loop structures are the second major mechanism of translation inhibition. They are formed by self-complementary sequences within the 5'UTR. These structures can be very stable impeding the unwinding activity of helicase eIF4A during the normal ribosomal scanning [9]. Human platelet-derived growth factor 2 (*PDGF2*), transforming growth factor b1 (*TGFb1*) and ornithine decarboxylase are some examples of mRNA transcripts where stable stem-loop structures interfere with the initiation of the translation [24-26]. The helicase eIF4A is capable of unwinding less stable stem loops, (free energies up to 2126 kJ/mol). However those loops are often reinforced by RNA-binding proteins, resulting in translation inhibition [9].

A well-defined example of 5'UTR stem loops stabilized by mRNA-binding proteins is the iron homeostasis. Specially-designed mRNA-binding proteins, known as iron regulatory proteins (IRPs) [27], sense the endogenous iron load and interact with mRNA stem-loop structures known as iron-responsive elements (IREs) [28]. Ferritin and erythroid-specific d-aminolevulinate synthase (*ALAS2*) have IREs that bind to IRPs with high affinity when intracellular iron is low, thus inhibiting ferritin translation. Mutations

causing the hereditary hyperferritinemia-cataract syndrome change the stem loop structure of IRE in ferritin mRNA, disrupting the fine control of ferritin translation [29].

Many studies have correlated the 5'UTR length and the number of start codons to translational efficiency of the mRNA transcript [30-33]. Longer 5′UTRs with numerous start codons are more likely to produce false starts or short ORFs leading to non-functional translation products [8]. McClelland et al [32] studied the role of 5'UTR in regulating translational efficiency in seven distinct transcripts encoded by the PPAR-γ gene. Peroxisome proliferators-activated receptors (PPAR) are a family of nuclear receptors implicated in the regulation of carbohydrate and lipid metabolism [34,35]. Alternative splicing results in multiple PPAR-*γ* transcripts that differ only in the 5'UTR. Studying translation *in vitro* and *in vivo*, the investigators showed that variants with short 5'UTR length, low numbers of start codons and the least stable mRNA secondary structure, are the variants that translate more efficiently.

Jin et al [33] studied a naturally occurring 32-base pair deletion mutation in the human chemokine receptor 5 *(CCR5)* gene that serves as coreceptor of the human immunodeficiency virus type 1 (HIV-1) [36]. Individuals homozygotes for the mutation are highly protected against HIV-1 infection [37]. Several *CCR5* transcripts with different 5'UTRs exist as the result of complex splicing and multiple transcriptional starts [38]. Translational efficiency of the different 5'UTRs was studied *in vitro* revealing that the shortest 5'UTR was the most efficiently translated.

Finally, the internal ribosome entry site (IRES) is a 5'UTR sequence that mediates translation independent of the m^7G cap, particularly found in picornaviruses [39]. Interestingly, researchers have found that the 5' UTR of the proto-oncogene *c-myc*

contains an IRES, showing that the translational regulation of *c-myc* is mediated through cap-independent mechanisms [40]. A mutation in the *c-myc* 5'UTR IRES has been identified in multiple myeloma patients. RNA transcripts bearing the mutation show enhanced binding of cellular proteins [41]. Further research has revealed IRES in many other eukaryotic genes [42].

3.2 B. Kozak sequence in translational control

In 1986, Kozak [43] studied point mutations around the ATG initiator codon in a cloned preproinsulin gene and identified the sequence ACCATGG in the translation start site as the most favourable context promoting initiation of translation. This experimentally-derived optimal sequence was later replicated in 699 vertebrate mRNAs; the gccRCCAUGG emerged as the consensus sequence for translation [44]. The A nucleotide of the AUG start codon is defined as position 1, and the preceding C nucleotide is referred as position -1. R at the position -3 (R^{-3}) represents A or G. The most conserved positions, likely representing the most functionally significant positions, are the G^{+4} and R^{-3}; 46% of vertebrate mRNAs have G^{+4} and 97% have R^{-3} [43,45]. Replacement of the nucleotides on those two positions greatly affects translation efficiency [43,46].

Genetic variations in the Kozak sequence have significant clinical consequences. The *CDKN2A* (cyclin-dependent kinase inhibitor 2A) gene encodes the tumor suppressor protein p16, a CDK4 kinase inhibitor [47] that causes cell arrest at G_1 phase, thus inhibits cell growth [48]. A G→T mutation at the 5'UTR of the *CDKN2A* gene, produces a strong matched Kozak consensus sequence 34nt upstream of the normal translation start site,

therefore the production of the tumor suppressor p16 is impaired, increasing susceptibility to melanoma [49].

De Angioletti et al [50] report a G→C mutation at the -6 position of the Kozak sequence of the β-hemoglobin gene that decreases the translational efficiency of the β-globin chain by about 30%. Although the mutation alone is not disease-causing, it significantly worsens the clinical phenotype of thalassaemia intermedia when it coexists with a thalassaemia mutation.

3.2 C. The 3'UTR in translational control

Translation begins at the 5' end of the mRNA. However the 3'UTR has a significant regulatory effect mainly mediated by two mechanisms: 1) the ribosomal subunit binding through the interaction of eIF4G and the poly(A) binding protein (PABP) [51,52] that results in mRNA circularization bringing the 3' UTR in close proximity to the 5' end of the mRNA [17,53] and 2) the cytoplasmic polyadenylation/deadenylation process that regulates the length of poly(A) tail [54,55].

Almost all eukaryotic pre-mRNAs undergo processing at their 3′ end by cleavage at the polyadenylation site and addition of a non-templated polyadenylated tail, consisting of 200–300 adenosine residues [54,56]. This first round of nuclear polyadenylation, following the pre-mRNA cleavage, seems to be a default mechanism [54]. Subsequent polyadenylation steps occur in the cytoplasm, a process mainly driven by CPEB (cytoplasmic polyadenylation element binding), a highly conserved sequence-specific RNA-binding protein that binds to the cytoplasmic polyadenylation element (CPE) [57].

However, the final length of the poly(A) tail will be defined by concomitant deadenylation and polyadenylation in the nucleus and the cytoplasm, depending on the cellular conditions. mRNA deadenylation results in the shortening of the poly(A) tail and is a process triggered by specific *cis*-acting sequence elements in the 3'UTR; APA (alternative polyadenylation) signals, AU-rich elements (AREs) including the embryo deadenylation element (EDEN), and microRNA target sites [54,58].

MicroRNAs (miRNAs) are highly conserved, non-coding small RNA molecules consisting of around 20 nucleotides. They are able to regulate gene expression by binding to the complementary mRNAs. More than 900 miRNAs have been identified in humans with almost 60% of human protein-coding genes harboring miRNA target sites in their 3' untranslated region [59]. Numerous studies implicate miRNAs in multiple biological processes, including the pathogenesis of cancer and apoptosis [60], the cardiovascular biology [61], the development and function of the female reproductive tract [62], neurodevelopment and behavioural disorders (fragile X syndromes, schizophrenia, depression and drug addiction) [63]. The list of reported miRNA functions continues to grow rapidly.

TNF-α (tumor necrosis factor-a) and COX-2 (cyclooxygenase-2) are two examples of ARE-containing mRNAs. The ARE sequence impairs the uncontrolled production of TNF-α, regulating the inflammatory response. Transgenic mice that lack ARE in their TNF-α mRNA show high levels of circulating TNF-α, and subsequently develop joint- and gut-associated immunopathologies [64,65]. Interestingly, genetic variations in the polyadenylation signal sequence have been detected in β^{+}-thalassemia patients. In 1985, Orkin et al [66] described a mutation in the cleavage-polyadenylation

signal of the human beta-globin gene and Rund et al [67] detected two mutations in the beta-globin poly(A) signal in Israeli patients with β^{+}-thalassemia.

3.3 Ribosomal Protein S26 *(RPS26)* and its association to complex human diseases

3.3 A. Ribosomal Protein S26

Ribosomes are the organelles responsible for protein synthesis. They consist of two subunits, the small 40S and the large 60S subunit, which contain 4 RNA species and approximately 80 structurally distinct proteins. The ribosomal proteins have a similar structural organization, sharing common features in their 5'ends, where transcription- and translation- control sequences are found [68].

The ribosomal protein S26 *(RPS26)* is one of the 80 distinct ribosomal proteins, it is located in the mRNA-binding center of the ribosome [69,70] and is significantly conserved in eukaryotes [71] including dog, cow, mouse, rat, zebrafish, fruit fly, mosquito, *C.elegans, S.cerevisiae, A.thaliana*, rice, and *P.falciparum*. In prokaryotes there is no ribosomal protein homologue of *RPS26* [72], and in plants *RPS26* is differentially expressed during environmental stress [73].

Ribosomal protein S26 was firstly cloned and sequenced by Vincent et al. in 1993 [74]. One year later, Filipenko et al [75] isolated and amplified the corresponding sequence of human cDNA from total placenta cDNA, based on the known sequence of rat S26 ribosomal protein. The fragment of S26 cDNA has 87.7% homology between coding regions of *RPS26* mRNAs in rat liver and human placenta. The same group a few

years later [68] was able to map the intron-containing *RPS26* gene to chromosome 12 by using hybrid cell lines with a different set of human chromosomes.

Ivanov et al [72] trying to explain the large variations in the expression of the human ribosomal protein genes that take place in different normal adult human tissues [76] showed that human *RPS26* suppresses the splicing of its pre-mRNA, thus contributing to gene regulation. *RPS26* interacts with both the intron I of its pre-mRNA and with another mRNA fragment that contains the 5'UTR and the coding part. In vitro splicing of an RNA fragment that contained exon I, intron I and part of exon II of the *RPS26* pre-mRNA produced alternative mRNAs, whose formation was inhibited by adding recombinant *RPS26*.

In 2010, Doherty et al [77] assessed the role of *RPS26* in pre-rRNA processing. *RPS26* protein expression was knocked down with siRNAs in HeLa cells. siRNAs are short double-stranded RNA molecules that interfere with the RNA transcript and silence the gene expression by causing degradation of the corresponding endogenous mRNA. Forty-eight hours post-transfection, RNA was extracted and high levels of the 18S precursors (43S, 26S, and 18S-E RNA) were detected, indicating that *RPS26* is necessary for the production of the small subunit. The same pre-RNA pattern was also detected *ex vivo* in RNAs extracted from lymphoblastoid cells derived from patients with *RPS26* mutations, who were diagnosed with Diamond-Blackfan anemia, a disease causatively linked to *RPS26*, as it is explained in the next sections. The authors conclude that mutations in *RPS26* are capable of changing the function of ribosomal proteins in ribosome biogenesis.

Finally an interesting property of the *RPS26* gene was found in yeasts. The *RPS26* homologue of *Saccharomyces cerevisiae* is involved in the normal cell growth of the species, as disruption of its expression results in the formation of micro-colonies [71]. Strittmatter et al [78] also showed that *RPS26* is critical in the dimorphic switch of the baker's yeast from a single cell yeast to an adhesive filamentous growth form.

3.3 B. The association between RPS26 and type 1 diabetes

RPS26 is located on 12q13 chromosome within a locus that was initially identified in the McGill–CHOP large-scale genome wide association study (GWAS) as a novel susceptibility locus for type 1 diabetes [6]. This finding was independently reported in a number of cohorts, including the GWAS of the Wellcome Trust Case-Control Consortium [7,79]. The first stage of the McGill–CHOP GWAS [80] had revealed a significant association of T1D with variation within a 233-kb linkage disequilibrium block on chromosome 16p13. Twenty-three loci that were next in rank of statistical significance were further investigated in two independent cohorts of type 1 diabetic patients; the Type 1 Diabetes Genetics Consortium replication cohort including 549 families with at least one child diagnosed with diabetes (946 total affected) and the Canadian replication cohort consisting of 364 nuclear family trios with one type 1 diabetes-affected offspring and two parents (1,092 individuals). The highest statistical significance (p = 9.13 x 10^{-10}, OR 1.25 [95% CI 1.12-1.40]) was detected with a novel locus at 12q13 chromosome that confers susceptibility for type 1 diabetes comparable with that of the 16p13 locus [6].

Furthermore, *RPS26* was identified as the most likely susceptibility gene for the T1D 12q13 locus in a human liver gene expression microarray study [81]. Schadt et al studied the genetic architecture in the human liver by applying an integrative genomics approach; they profiled 427 human liver samples targeting more than 39,000 transcripts and they genotyped 782,476 unique SNPs to uncover genetic associations in the highly-metabolic liver tissue. Since the T1D SNPs genotyped in the GWAS of the Wellcome Trust Case Control Consortium (WTCCC) [7] were also genotyped in the human liver cohort (HLC) study [81], Schadt et al. examined the association between those SNPs and the HLC expression levels of the corresponding candidate susceptibility gene for that SNP. Comparing data from the two studies [7,81], *RPS26* and not *ERBB3* (the gene nearest the SNP) was found to have HLC expression levels significantly associated ($p = 4.03 \times 10^{-22}$) with the novel T1D-associated SNP (rs2292239) at the 12q13 region.

Apart from the liver tissue where the expression of *RPS26* is strongly affected by the T1D-associated SNPs, the same association was also revealed by expression quantitative trait locus (eQTL) mapping in lymphoblastoid cell lines [82]. Expression QTL (eQTL) is defined as a locus at which genetic allelic variation is associated with variation in gene expression [83]. Dixon et al [82] genotyped 408,273 SNPs and identified eQTL from 54,675 transcripts in 20,599 genes in Epstein-Barr virus–transformed lymphoblastoid cell lines; *RPS26* was one of the most significant eQTL detected.

Furthermore, a recent study by Rotival et al [84] identified association between the T1D 12q13 locus (lead SNP rs11171739) and two *cis* eQTLs, *RPS26* and *SUOX* ($p<10^{-300}$ and 3.1×10^{-18} respectively). The authors analyzed 12,808 genes detected in at

least 5% of circulating monocyte samples from 1,490 European unrelated subjects to investigate *trans* associations of co-expression modules. Three genomic regions were significantly associated *in trans* with modules of co-expressed genes. 12q13 locus was one of them, with the *trans* effects being mediated by two above mentioned *cis* eQTLs. The *cis* regulation of *RPS26* detected in monocytes confirms findings in other cell types; the human liver [81] and lymphoblastoid cell lines [82].

An interesting connection between T1D and *RPS26* is suggested by Schadt et al [81]. *RPS26* belongs to the ribosomal complex on the endoplasmic reticulum (ER). ER stress results in protein unfolding, a process also implicated in inflammation and oxidative stress [85]. Given the autoimmune nature of T1D on one hand and the increased susceptibility of the highly-secreting pancreatic beta cells to ER stress on the other hand, a potential connection of the ribosomal complex and T1D can be foreseen. It has been shown that T1D patients show higher levels of ER stress [86].

A paper by Plagnol et al in 2009 [87] directly questioned whether *RPS26* is the causative gene in the T1D. The authors combined data from two different studies: the genome-wide eQTL expression study by Dixon et al [82] and the British Juvenile Diabetes Research Foundation/Wellcome Trust (JDRF/WT) T1D case–control study with 4,000 cases and 4,000 controls. An initial set of 4 SNPs (rs705704, rs705699, rs1131017, and rs877636) was selected using a significance threshold of $p = 10^{-10}$ for the T1D association and $p = 10^{-50}$ for *RPS26* expression. Using a regression approach, they found a single SNP that could explain the T1D association (rs705704, $p = 10^{-50}$) and a single SNP (rs1131017, $p = 10^{-13}$) that could explain the *RPS26* eQTL association. The pairwise r^2 (correlation coefficient) between the two SNPs is 0.7. The probability that T1D

susceptibility in the 12q13 locus is mediated through the *RPS26* expression, was calculated as p=0.001. Based on the assumption of a sole causal variant, their correlation model does not support *RPS26* as the responsible trait for T1D susceptibility at the 12q13 region.

However, given the findings of three different studies in monocytes [84], LCLs [82] and human liver tissue [81], and taking into account the McGill–CHOP GWAS, where the rs1131017 *RPS26* SNP conferred the highest T1D risk, it seems reasonable to re-examine potential correlation between *RPS26* expression and T1D susceptibility using all available data.

3.3 C. RPS26 in Diamond-Blackfan anemia

Most of the current literature on *RPS26* focuses on its role in the pathogenesis of Diamond-Blackfan Anemia (DBA). DBA or inherited erythroblastopenia is a congenital bone marrow failure syndrome, mainly characterized by defects in red cell maturation that usually presents early in infancy. Haploinsufficiency of either a small or large subunit-associated ribosomal protein causes more than 50% of DBA cases, thus classifying DBA as a ribosomopathy [88,89]. The malfunctioning bone marrow of DBA patients predisposes them to other serious complications, including increased risk of myelodysplastic syndrome, acute myeloid leukemia (AML) and osteosarcoma. From 30 to 40% of DBA patients present with short stature and congenital anomalies; craniofacial malformations, thumb or upper limb abnormalities, cardiac defects, urogenital malformations, and cleft palate are the most commonly encountered [77,88].

Ribosomal Protein S19 *(RPS19)* was the first gene reported to bear a mutation in a DBA patient [90] in 1999. Since then, mutations in nine distinct ribosomal protein genes, including *RPS26*, have been described as disease-causing variants [91]. Doherty et al [77], were the first to report pathogenic *RPS26* mutations in 12 DBA patients. Their study cohort included 103 probands who were negative for mutations in the seven, known back then, DBA genes: RPS19, RPS24, RPL35A, RPS17, RPS7, RPL5, and RPL11. In six probands and in two family members a missense mutation in the first codon of *RPS26* was detected; methionine was changed to leucine, valine, or arginine. These sequence changes eliminate the translation-initiation codon and the protein is predicted to not be translated. Missense mutations were detected in a total of 8 probands with 4 of them bearing a *de novo* change. 3 probands had splice-site mutations in introns 1 and 2 and finally 1 proband had a *de novo* insertion of G in exon 2 causing frameshift at codon 11 and stop at codon 25. Interestingly, genotype/phenotype correlation in 11 *RPS26*-mutated patients revealed that only 3 patients had physical malformations.

In 2011, a second group discovered four *de novo* heterozygous mutations in the *RPS26* gene in a cohort of 14 DBA probands from the Czech DBA Registry, who were negative for mutations in the known DBA-associated genes [91]. Two *RPS26* mutations were detected in the start codon, supporting the data by Doherty et al [77] that the *RPS26* tranlsation-initiation codon may be a mutational hot spot. The third *RPS26* mutation was located in the acceptor splice site of intron 1, likely interfering with proper splicing that could result in degradation of the forming mRNA. The fourth *RPS26* mutation was detected in exon 3; cysteine 77, which is conserved in yeasts, plants and humans, was

changed to tryptophan. All the reported mutations likely result in *RPS26* protein haploinsufficiency.

The authors also make some interesting *RPS26* genotype/phenotype correlations; in contrast to patients bearing *RPL5* and *RPL11* mutations, individuals with an *RPS26* mutation do not have any thumb abnormalities. Also, a female *RPS26*-mutated patient was diagnosed with Klippel–Feil syndrome (KFS), a rare condition characterized by failed segmentation of the cervical vertebrae [91,92].

3.3 D. RPS26 in psoriasis

A recent meta-analysis paper [93] of two previously published GWAS in psoriasis [94,95] implicates *RPS26* in the disease susceptibility loci. The rs12580100 SNP that resides 1.2 kb upstream of *RPS26*, even though it did not attain genome-wide significance (combined $p = 1 \times 10^{-6}$, replication stage OR = 1.17) it still captured the authors' attention. It is located 300 kb from a confirmed association signal at rs2066808 near *IL23A*. However, the two SNPs have low correlation coefficient (pairwise LD $r^2 = 0.23$), as they reside in different LD blocks. A logistic regression model including both SNPs generated significant partial association for each ($p = 0.016$ for rs12580100 and $p = 3.0 \times 10^{-4}$ for rs2066808). Further supporting potential correlation of *RPS26* with psoriasis susceptibility, the authors examined psoriatic skin lesions, and found several highly significant eQTLs that colocalize with the psoriasis association signal for *RPS26* but not for *IL23A*.

3.4 Initial Hypothesis and experimental design

The purpose of my MSc. thesis was to test the hypothesis that alleles of exonic SNPs found at higher abundance in the polyribosomes also increase translation at the protein level. Because of its disease associations, RPS26 was selected among several genes in which we found association of exonic SNPs with ribosomal distribution patterns.

The *RPS26* SNPs most strongly associated with ribosomal distribution, rs17118262 (C/G) and rs1131017(C/G), are in very close proximity located 9 bases apart, are in strong linkage disequilibrium (D'=1) but have different allele frequencies (r^2=0.54). The rs1131017 SNP, which is a more common variant compared to the rs17118262, was identified in the initial Illumina transcriptome profile study as the variant being significantly associated with ribosomal distribution in the *RPS26* gene. However subsequent Sanger Sequencing of the ribosomal fractions of individuals who are heterozygotes for both SNPs revealed that the G allele of the rs17118262 is found at higher proportions in the heavy polysomes, suggesting that this variant also plays a significant role on ribosomal distribution. Due to the close proximity of the two 5'UTR *RPS26* SNPs and their strong LD, we decided to include both SNPs in our downstream investigations. The goal of this project was to study the effect both those SNPs have on translational efficiency.

3.4 A. Independent confirmation of allelic imbalance in ribosomal distribution for RPS26 mRNA

To independently confirm the translational imbalance found by ribosomal distribution in the three genotypes of the 40 individuals, we studied allelic distribution in single heterozygous samples (allelic translation). Single Nucleotide Primer Extension (SNuPE)

and Pyrosequencing of the different RNA fractions (total RNA, heavy and light fractions) were used to measure the differential abundance of the mRNA alleles in ribosomal fractions from heterozygous individuals.

3.4 B. Confirmation of translational effects at the protein level

After confirming the translational imbalance, we could assume that the mRNA allele found at a higher proportion in the polyribosomal fraction than in the soluble RNA will produce more protein product under the same translational environment. To test this hypothesis we applied both an *ex vivo* quantification and *in vitro* translation approach. By quantifying the *RPS26* protein derived from two independent sources, we could reliably confirm or reject our hypothesis.

4. Methods

4.1 Methods for independent confirmation of allelic imbalance in ribosomal distribution for *RPS26* mRNA

4.1 A. Polysome profiling

As stated before, a method to ascertain translational control of an mRNA transcript is by analyzing the redistribution of mRNA between ribosome-free and polysome-bound fractions. Ribosome loading of a transcript serves as an indicator for translation efficiency [5]; RNAs that are efficiently translated associate with multiple ribosomes, while less active ones associate with fewer or none ribosomes. Using ultracentrifugal fractionation according to size, clusters of multiple ribosomes (polysomes) are separated

from single ribosomes (monosomes) and soluble RNA (free RNA bound to none ribosome). The less active allele, associated with fewer ribosomes, is less heavy and thus remains in the upper fractions of the ultracentrifuge tube. On the other hand, most efficiently translated alleles are associated with multiple ribosomes and thus heavier in weight resulting in higher abundances in the bottom of the tube.

Preparation of the ribosomal fractions was part of the thesis of the previous graduate student, Dr Lu Yang. Briefly, human immortalized lymphoblastoid cell lines (LCL) were obtained from the Centre de l'Étude du Polymorphisme Humain (CEPH) CEU. The CEPH CEU samples are U.S. Utah residents with northern and western European ancestry and have been used for the international HapMap project; their genotypes are publicly available in the website *http://www.hapmap.org*. LCLs were cultured using RPMI-1640 containing 15% FBS, 1% L-glutamine, 1% penicillin and 1% streptomycin, in 37ºC 5%CO_2 incubator. $1x10^8$ cells from each LCL were incubated with 100 µg/µl cycloheximide to stop translation and fix the ribosomes on the mRNAs. Cells were lysed with hypotonic polysome lysis buffer (Tris-HCL pH 7.5, MgCl2, KCL, 100 µg/mL cycloheximide, dithiothreitol (DTT), 10% Triton X-100, and 10% sodium deoxycholate). The free-of-cellular debris supernatant was directly applied to 10%~50% sucrose gradients. After centrifugation in a Beckman SW41 rotor, polyribosomal fractions were collected using the Brandel fraction collector (Gaithersburg, Maryland) with real-time monitoring of UV optical density at 254 nm. RNA from each fraction was then extracted using Trizol (Invitrogen, CA, USA) followed by the phenol–chloroform method and quantified by optical density. The relative abundance of ribosomal RNA

subunits was assayed on the Agilent 2100 bioanalyzer with the RNA 6000 Nano LabChip® kit (Agilent Technologies, CA, USA).

However, due to the presence of varying amounts of ribosomal RNA in different fractions, the OD measurement was not an accurate reflection of abundance of total mRNA. Precise quantification of the mRNA present in each fraction was done using the DIG Oligonucleotide 3'-End Labeling Kit (Roche, Indiana). Aliquots from the ribosomal fractions were slot-blotted on a nylon membrane and quantified using an oligo(dT) probe labeled wtih digoxigenin. The final concentration was quantified by the densitometry image analysis (Bio-Rad GS-700 imaging densitometer and Bio-Rad Quantity One 4.2, Bio-Rad, France) against a standard curve of the commercial human pancreas Poly(A) RNA at the standard concentrations.

4.2 B. Reverse transcription of the ribosomal fractions

RNA fractions thus previously obtained, were used for part of my experiments. First, each sample of RNA (1 µg) was reverse-transcribed into cDNA by use of Superscript II Reverse Transcriptase (Invitrogen). Each reaction was carried out in total volume of 20 µl, containing 10 mM d-NTPs (deoxyribonucleotide triphosphate; HT Biotechnology) and 200 U/ml superscript in the presence of random hexamers (300 µg). RNA was initially denaturated at 70°C for 5 minutes and quick-chilled on iced water. The reaction mixture was then added and reverse transcription was performed at 37°C for 1 hour. cDNA samples of the cohort of 40 individuals from the CEU set of HapMap were produced with each individual having 3 different cDNA aliquots representing the total

RNA, the heavy RNA fraction and the light RNA fraction. cDNA samples were stored at -20°C prior to use as templates for SNuPE and pyrosequencing reactions.

4.2 C. Single nucleotide primer extension (SNuPE)

Single nucleotide primer extension analysis was done to independently study the translational imbalance suggested by the transcriptome profiling study of the ribosomal fractions. SNuPE is a technique that allowed us to accurately measure the relative abundance of each allele in a given ribosomal fraction [96]. The sequence of interest containing the target SNP is being amplified by polymerase chain reaction (PCR) using reverse-transcribed RNA as a template. The PCR products are then analyzed for the presence, absence and relative amounts of each allele of the SNP. A probe is designed to bind exactly 1 base before the polymorphism and then it gets extended by 1 base. Each dideoxynucleotide is labelled with a different fluorescent label and if complimentary to the template, it gets incorporated at the 3' end of the probe.

SNuPE was done prior to my arrival to the laboratory by the research assistant Luc Marchand. Using the *Generunner* program *(http//www.generunner.net)* appropriate *RPS26* probes were manually designed. Since SNuPE served as an independent confirmation of the association initially found by transcriptome profiling, there was no need to apply it to all the 40 LCLs samples of the CEU HapMap set. Representative heterozygous samples were selected. Based on the genotype of the samples, probes for each of the two SNPs were designed. Results were displayed as percentages of each allele (C or G) in the different ribosomal populations: heavy polysomal RNA fractions, light RNA fractions and total RNA.

4.2 D. Pyrosequencing

Pyrosequencing was used as an additional method to study the proportions of each allele in the ribosomal fractions from heterozygous individuals. As with SNuPE, this independently confirms, by a totally different approach, the association found by transcriptome profiling of the ribosomal fractions that compared ribosomal distribution across the three genotypes. Pyrosequencing is a sequencing method that relies on the principle of "sequencing by synthesis". The activity of the DNA polymerase, extending a biotinylated primer with the complementary sequence as template, is detected by a chemiluminesence signal. Light is produced when the appropriate nucleotide complements the corresponding base of the template. Solutions of the four nucleotides, adenine, guanine, thymine and cytosine, are added to the reaction and subsequently washed away. The pyrophosphate being released when the appropriate nucleotide is incorporated, produces chemiluminescent signal, allowing the determination of the sequence of the template.

Forward and reverse PCR (polymerase chain reaction) primer sequences for *RPS26* gene were designed using the PyroMark Assay Design Software 2.0 (Qiagen, CA, USA). Expected PCR product was 162 bases pair. The lyophilized primer sequences were purchased from IDT (Integrated DNA Technologies, Iowa, USA) and dissolved in nuclease-free water to make a stock solution of 100 μM. Working solution aliquots of 1:10 dilution were made and stored at -20°C. Conditions for PCR amplification of the DNA template were optimized using the PyroMark PCR Kit (Qiagen, CA, USA). As suggested by the manufacturer, PyroMark PCR Master Mix was used. It contains HotStarTaq DNA Polymerase and PyroMark PCR buffer with 3 mM $McCl_2$ and dNTPs.

Optimization reaction with a representative sample was carried out in total volume of 50 µl, containing 10 µM of each primer, PyroMark PCR Master Mix 2x and reverse-transcribed DNA (<500 ng). The sample tube was filled up with nuclease-free water in a total volume of 50 µl and placed in the thermal cycler. The amplification program consisted of the following steps: 15 min at 95°C (required for the activation of the HotStarTaq DNA Polymerase), 45 cycles of 20 seconds at 94°C, 20 seconds at 56°C and 10 seconds at 72°C, with a final step of 10 min at 72°C. After verification of the PCR product by agarose gel analysis, we proceeded with ordering the 5'biotinylated primer and the sequencing primer that were previously generated by the PyroMark Assay Design Software 2.0 (Qiagen, CA, USA). The pyrosequencing procedure requires one of the primers to be biotin-labeled to enable immobilization to streptavidin-coated beads during preparation of the single-stranded pyrosequencing template. The orientaton of the assay can either be forward or reverse. The sequencing primer is hybridized to the single-stranded PCR-amplified DNA template. It matches to the biotinylated strand and it was carefully checked for hairpin loops and duplexes with the biotinylated primer, as excess biotinylated primer could cause significant background in the pyrosequencing assay. The biotinylated primer was HPLC-purified as free biotin competes with the biotinylated PCR product for binding on streptavidin-coated beads.

In our experiment the 5'biotinylated primer was the reverse primer. Following the same procedure as in the optimization PCR, mixture reactions were prepared for all our samples (cDNAs from total RNA, heavy and light RNA fractions of 6 individuals with CG/CG genotype, 3 individuals with CG/CC genotype and 2 individuals with CC/CG genotype). Again PCR products were verified by agarose gel analysis and 35 µl of each

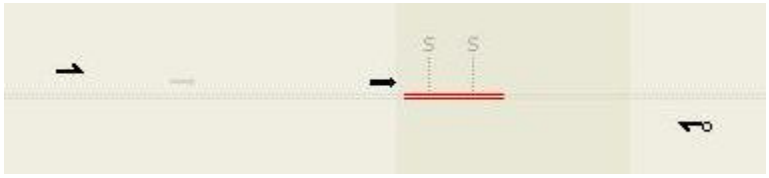
PCR sample was used for pyrosequencing in the PyroMark Q24 instrument (Qiagen, CA, USA).

Following the manufacturer's instructions, the biotinylated PCR products were immobilized on streptavidin-coated sepharose beads (2 µl of streptavidin beads per sample). Binding buffer (40 µl per sample) and high-purity water were added to a total volume of 80 µl per well, including the 35 µl of PCR product. The mixture was agitated at 1400rpm for 10 minutes allowing the sepharose beads to sediment. Samples were processed in the vacuum workstation with serial washes of 70% ethanol, 1x wash buffer, denaturation solution and high-purity water. 0.3 µM of sequencing primer was mixed with annealing buffer and 25 µl of the solution was added to each sample. Following instructions in the PyroMark Q24 Manual, samples were heated at 80°C for 2 minutes before placing them to the PyroMark Q24 instrument. PyroMark Gold Q24 Reagents Kit was used to load the appropriate amounts of nucleotides, lyophilized enzyme and substrate mixtures, as had been previously generated using the manufacturer's software program.

PyroMark Q24 software was used to create and load the run parameters of the *RPS26* assay. The AQ (allele quantification) analysis mode was selected. The sequence to be analyzed was pasted into the corresponding text box, following the IUPAC code (A-Adenine, C-Cytosine, G-Guanine, T-Thymine, M-C/A polymorphism, K-T/G polymorphism, W-T/A polymorphism, S-C/G polymorphism). The software automatically checks the validity of the sequence. Our sequence, 152 bases-pair long, contained the rs17118262 and rs1131017 SNPs of *RPS26*. *Figure 1* shows the exact location of the two SNPs and the region to be amplified by the forward, the reverse and

the sequencing primer. When setting up the AQ assay, a theoretical representation of the expected pyrosequencing peak pattern was given in a histogram with variable positions being highlighted with a blue-gray background color.

A.



B.

```
    AAGTGAGTTG CCGTTCTTGA A                                                TATAGGAGGG CCCTG
    |||||||||| |||||||||| |                                                |||||||||| |||||
143 AAGTGAGTTG CCGTTCTTGA AGCCCGTCTC CTAAGGATTC TCCCGGTGTC CGCGTAGGGA TCTCATGCTA TATAGGAGGG CCCTGCCAGG

233 CASCGTCTCC TSTCTCCGGT CCGTGCCTCC AAGATGACAA AGAAAAGAAG GAACAATGGT CGTGCCAAAA AG 305
                                                          ||||||| |||||||||| ||
                                                          GTTACCA GCACGGTTTT TC
```

Figure 1: Panel A shows the graph generated by the pyrosequencing software with the location of the two target SNPs (labelled as S) and the region amplified by the 3 primers; forward, reverse and sequencing primer. The forward primer is shown as half arrow on the left of the graph and the reverse biotinylated primer is marked with a small cycle on the right corner of the graph. The forward and reverse primers were designed to amplify the RPS26 gene in a PCR reaction. The biotinylated PCR products were then analyzed in the pyroMarkQ24 instrument. The region marked by the red lines shows the sequence that was analyzed; it begins with the sequencing primer shown as a bold arrow right before the two target SNPs. Panel B shows the exact nucleotide sequence of all the three primers and the RPS26 region that they bind.

The results of the run were displayed in specially designed graphs, called pyrograms, where the incorporated nucleotides are shown as separate peaks, indicating the allele

frequencies of the target *RPS26* SNPs. Allele frequencies above each variable position (SNP) are displayed in percentages. A representative pyrogram from our samples is shown in *figure 2*. The background color represents the quality assessment of the analysis result. The color code as defined in the PyroMark manual is as follows: blue-the analysis passed quality control, yellow–it needs to be checked, red-the analysis failed the quality control. Any samples not displayed in blue were repeated. All the samples were run in duplicates or triplicates ensuring reproducibility of our results. Analysis results report was generated for all the samples and pyrograms of the different ribosomal fractions of the same sample were compared. A clear trend favouring the G allele of the first SNP (rs17118262) in the heavy polysomes was noted. Statistical analysis of the allele frequencies in all the patients (double heterozygotes, homozygotes only for the first SNP and homozygotes only for the second SNP) was then carried to compare the allele frequencies between the heavy and light fractions in each sample.

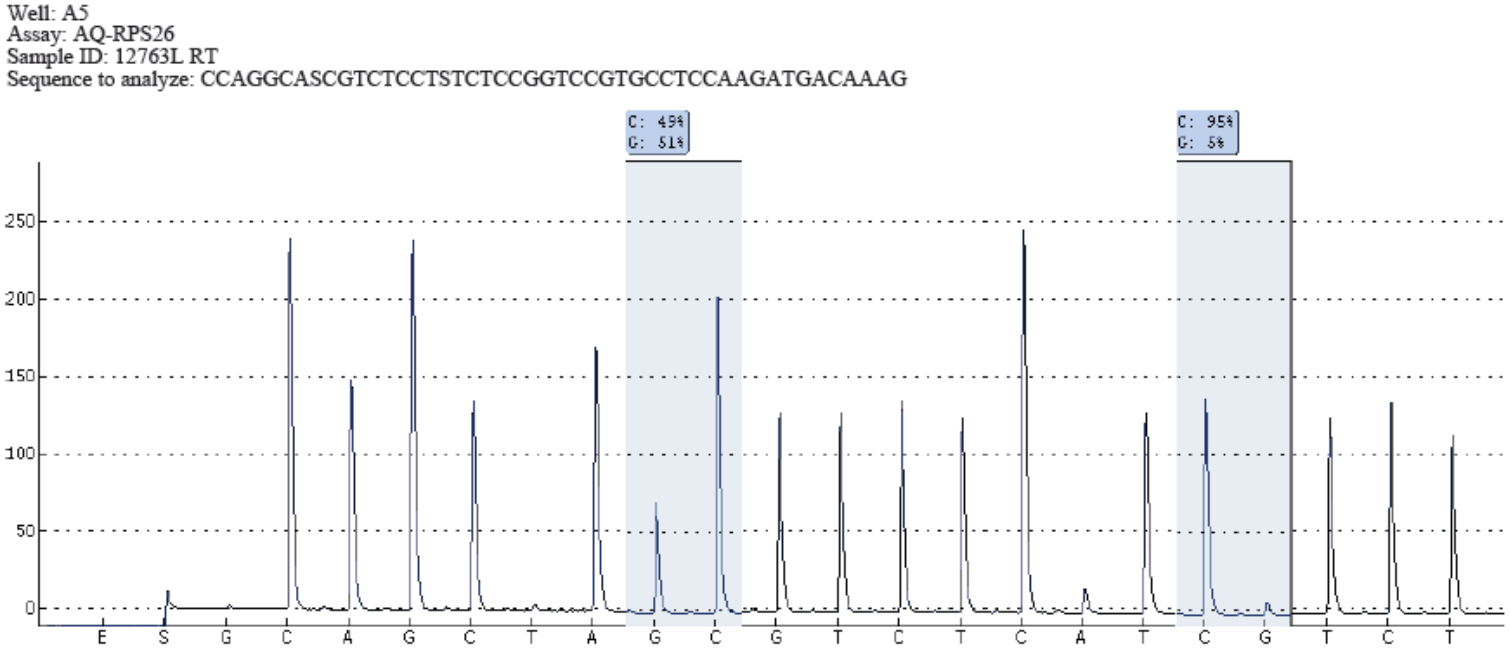


Figure 2: A representative pyrogram. The x-axis shows the sequence that is being analyzed. The incorporated nucleotides are shown as separate peaks. Allele frequencies of each variable position (SNP) are displayed as percentages in the blue boxes above the corresponding blue columns. The y-axis shows the height of the peaks representing the signal value. The minimum value for a peak to achieve "passed" quality assessment is 10. The genotype of the above sample is CG/CC as also confirmed by the HapMap data. In the first SNP, the percentages of the C and G alleles are similar in that particular ribosomal sample; however the height of the C allele is almost double the size of the G allele. Looking closer to the sequence, there is a C allele right after the SNP; thus the signals of the two consecutive C alleles are adding up in the peaks.

4.2 Methods for confirmation of translational effects at the protein level

To test the hypothesis that mRNA with the G allele produces more protein, we applied both an *ex vivo* quantification and *in vitro* translation approach.

4.2 A. Methods for ex vivo translation quantification

The purpose of this was to compare the amount of *RPS26* proteins across samples with different genotypes, after adjusting for the level of *RPS26* mRNA.

4.2 A1. Protein extraction

Total protein extract was obtained from the lymphoblastoid cell lines (LCLs) of the 40 CEU individuals that were initially used in the ribosomal fraction study. Cells were

centrifuged, washed with PBS and mixed with the appropriate volume of Cell Extraction Buffer (CEB) consisting of 98% TALON buffer, 1% Protease inhibitor and 1% PMSF (phenylmethylsulfonyl fluoride). Samples were subject to vortex every 10 minutes for 10 seconds to ensure complete lysis and after 30 minutes they were centrifuged at 20,000 g for 20 minutes. The supernatant was transferred to a clean tube and stored to -80°C until protein concentration was determined using the BCA (bicinchoninic acid) assay.

4.2 A2. Protein quantification with bicinchoninic acid assay (bca)

To precisely quantify the total protein extracted from the 40 LCLs of the CEU set of HapMap, we used the Pierce BCA Protein Assay Kit (Thermo Scientific, USA). Bicinchoninic Acid (BCA) allows highly accurate colorimetric detection of the cuprous cation. The method relies on the chelation of two molecules of BCA with one cuprous ion. This water-soluble complex produces a purple-colored reaction and exhibits a strong absorbance at 562 nm. This allows the spetrophotometric quantitation of protein samples in aqueous solutions. Three different factors are involved in the color formation in protein samples when assayed with BCA: the macromolecular structure of the protein, the peptide bonds and the presence of four particular aminoacids, cysteine, cystine, tryptophan, and tyrosine. Protein concentrations are determined and reported based on the known concentrations of a common protein, like the bovine serum albumin (BSA). A series of dilutions of known concentration of the standard protein (BSA) are prepared and assayed alongside the unknown protein samples. A standard curve is obtained based on the BSA standard dilutions correlating absorbance to protein concentration. Using this standard curve, we determine the concentration of our protein samples.

The experiment was carried out as following: a mixture of three reagents (Talon buffer 98%, PMSF 1%, Protein Inhibitor 1%) was initially made in a total volume of 60 µl per unknown protein sample. In parallel, 5 different BSA concentrations were generated by serially diluting 40 µl of BSA in 160 µl of the above mixture solution: 2,000 µg/µl, 1,000 µg/µl, 500 µg/µl, 250 µg/µl, and finally 0 µg/µl, of BSA. The 5 standard sample concentrations were run in triplicates and each unknown sample in duplicates. A working reagent BCA solution was then prepared by mixing 50 parts of BCA Reagent A with 1 part of BCA Reagent B (50:1, Reagent A:B). 200 µl of the working solution were added in each sample, both the standard and the unknown samples. After incubation at 37°C for 30 minutes, absorbance was measured at 490 nm on a plate reader. The 490 nm absorbance measurement of the blank standard replicates (BSA of 0 µg/µl concentration) was subtracted from the measurements of all the other individual standard and unknown sample replicates. A standard curve was then prepared by plotting the average blank-corrected absorbance values of each BSA standard sample versus its concentration. In all our experiments r^2 of the standard curve graph reached >0.98, allowing us to accurately determine the concentration of the unknown samples *(figure 3)*.

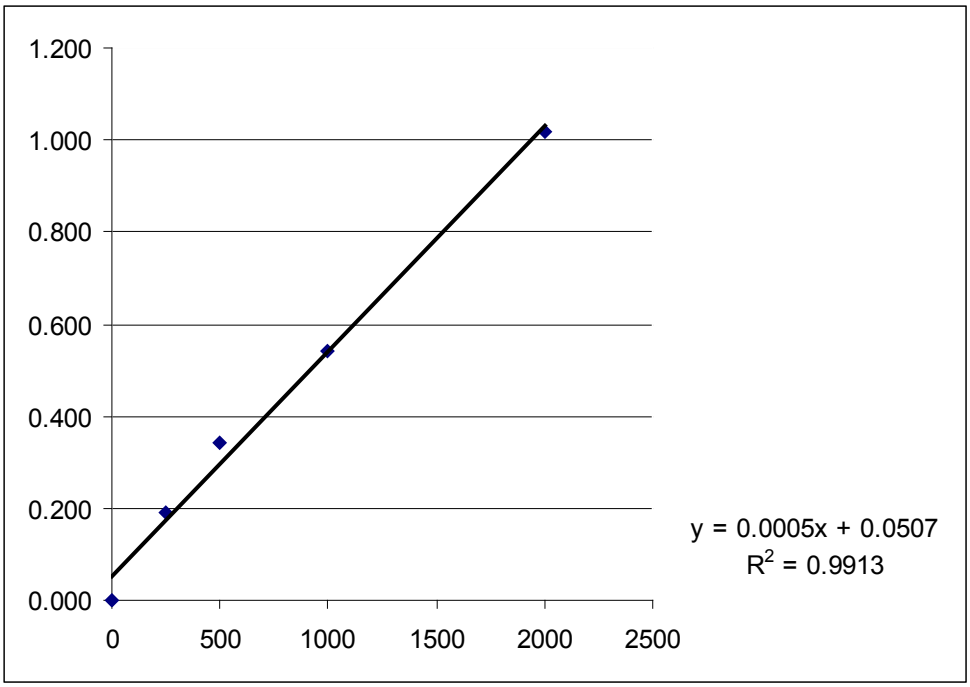


Figure 3: Bicinchoninic Acid Assay (BCA) standard curve is generated by plotting the average blank-corrected 490 nm absorbance values (y-axis) of each BSA standard sample versus its concentration (x-axis). Based on the graph, we determine the protein concentration of the unknown samples.

4.2 A3. Western Blotting

Western Blotting was performed to measure the relative amounts of the RPS26 protein present in the different samples. The quantified protein samples were first diluted to the same concentration to ensure equal loading. Equivalent amounts of loading buffer consisting of 30 mM Tris-HCl pH 6.8, 15% glycerol, 5% beta-mercaptoethanol and 2% of the anionic denaturing detergent sodium dodecyl sulfate (SDS) were added to each

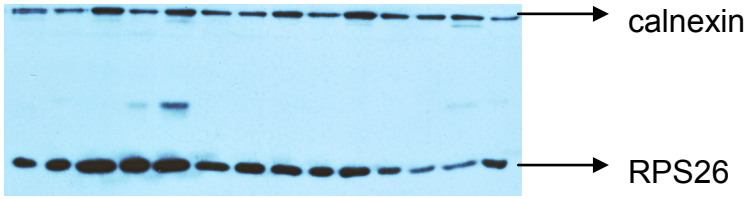
sample [97]. The mixture was boiled for 10 minutes at 100°C to allow protein denaturation. Protein unfolding is necessary since the domain that many antibodies recognize may reside within the 3D conformation of the protein. Samples were then vortexed and centrifuged at 10,000 rpm for 2 minutes and the full volume of the sample was loaded to a 15% polyacrylamide gel consisting of a separate and a stacking portion. The gels were prepared as following: solution A (acrylamide 40%), solution B (3M Tris-HCl, pH 6.8), 20% SDS, 10% APS, 30 µl temed. The gel ran for 2 hours at 100V in 1x running buffer, containing glycine, Tris base and SDS in pH 8.5, until the molecular dye (Kaleidoscope Polypeptide standards, Bio-Rad, CA, USA) reached the bottom of the gel.

Wet transfer was immediately done to avoid elution of the proteins from the gel. The PVDF (polyvinylidene fluoride) membrane was activated with methanol and placed next to the gel between absorbent materials ensuring no air bubbles have formed between the gel and membrane. The assembly is submerged in transfer buffer, consisting of glycine and Tris-base, to which an electrical field of 100V for 1 hour is applied causing the negatively-charged proteins to migrate towards the positively-charged electrode. The success of the transfer is assessed by visualizing the proteins in the PVDF membrane with Ponceau Red solution consisting of 2% Ponceau S in 30% trichloroacetic acid and 30% sulfosalicylic acid. The membrane was first washed with Tris Buffer Saline Tween 20 (TBST buffer consisting of Tris base, NaCl and tween 20, pH 7.6) to remove transfer buffer. The membrane now stained with Ponceau solution was incubated on an agitator for 5 min. Next, the membrane was washed with TBST buffer until the solution was clear and the protein bands well-defined.

Rabbit *RPS26* polyclonal antibody (ProteinTech Group, Illinois, USA) was diluted to 1:2000 and mixed with 5% milk blocking solution (5 g of non-fat milk per 100 ml TBST). Mouse calnexin antibody (BD Biosciences, California, USA) serving as the house-keeping gene, was also diluted to 1:2000 and mixed with 5% milk blocking solution. The membrane was incubated with both antibodies at 4°C overnight. Next morning, several TBST washes on the agitator were performed to remove residual primary antibody, and the membrane was subsequently incubated with goat-raised anti-mouse and anti-rabbit HRP-conjugated secondary antibodies at 1:2000 dilution for 1 hour at room temperature. TBST washes on the agitator for 30 minutes were again performed to remove the secondary antibodies.

Western Lightning Chemiluminescence ECL plus detection kit (PerkinElmer, MA, USA) was used to visualize the *RPS26* protein bands of the samples. After mixing equal volumes of the oxidizing reagent and the enhanced luminol reagent, 0.125 ml of Chemiluminescence reagent per cm^2 of membrane was applied. The membrane was incubated for 1 minute and immediately developed using manual x-ray films. ImageJ software was used to compare the density of the *RPS26* bands of the different samples. In each sample the density of the calnexin band (loading-control band) was used to scale the values for our protein of interest *(figure 4A).* The results were plotted in a graph *(figure 4B)* to detect potential correlation between the number of G alleles in the genotype (0, 1 or 2) and the level of *RPS26* (the genotypes of the 40 CEU individuals were already known from the HapMap project *http://hapmap.ncbi.nlm.nih.gov/*). However, the results had to be adjusted for the amount of RNA present in each sample.

A.



B.

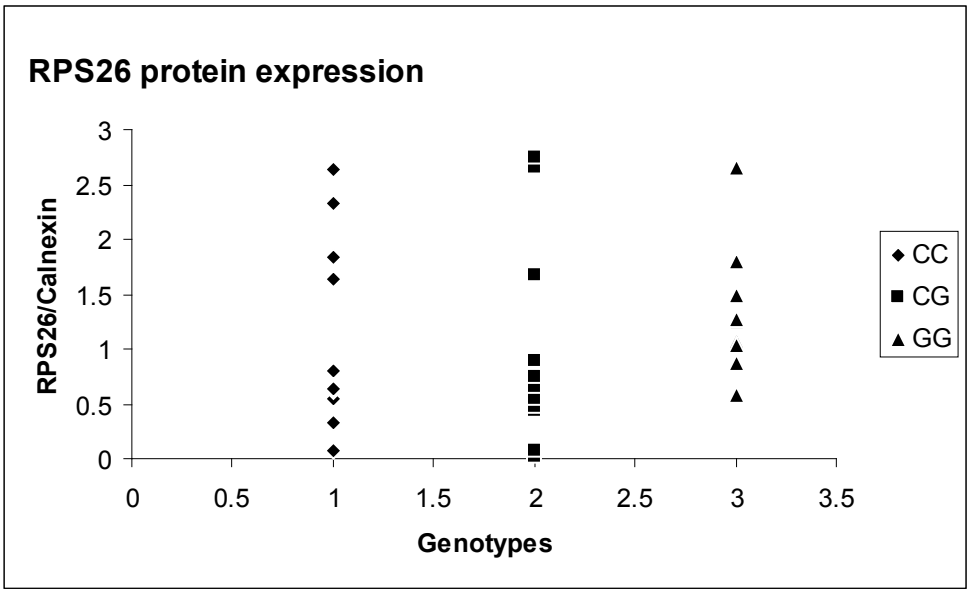


Figure 4: A representative Western Blot membrane (panel A) measuring the RPS26 protein product in the different genotypes with the ImageJ software. Calnexin served as the loading-control to scale the values for the RPS26 protein. After subtracting the density of the background, the ratio of RPS26/calnexin was calculated in each sample and the results were plotted in a graph (panel B), to detect potential correlation between the number of G alleles in the genotype (0, 1 or 2) and the level of RPS26 (the genotypes of the 40 CEU individuals were already known from the HapMap project http://hapmap.ncbi.nlm.nih.gov/).

4.2 A4. RNA extraction

To assess translation, protein results needed to be corrected for the amount of mRNA in each sample. In parallel to protein extraction from the 40 CEU LCLs, RNA was also obtained using the RNeasy Plus Mini Kit (Qiagen, CA, USA). Cells were lysed and homogenized in a highly denaturing guanidine-isothiocyanate-containing buffer that immediately inactivates RNases to ensure isolation of intact RNA. The lysate was then treated with a high-salt buffer and passed through a genomic DNA eliminator spin column (provided in the kit). Ethanol was added to the flow-through, allowing for appropriate RNA binding conditions and the sample was applied to an RNeasy spin column (also provided in the kit), where total RNA binds to the membrane and contaminants are efficiently washed away. 40 µl of nuclease-free water were then added, eluting purified RNA whose concentration was initially determined by optical density.

4.2 A5. Multiplex Ligation-dependent Probe Amplification (MLPA)

Multiplex Ligation-dependent Probe Amplification (MLPA) was performed to measure the RNA levels of *RPS26* in all the samples to accordingly adjust the protein *RPS26* levels. MLPA is a method that allows detection of difference quantities in genomic DNA or RNA sequences, enabling the study of copy number variations, allelic imbalance and degree of methylation. Moreover, with a careful design, SNPs can be interrogated as well.

MLPA is based on quantifying different sequences by multiplexing several PCR reactions, all using the same pair of primers that are added to the end of probes that cover adjacent segments of the sequence to be interrogated. Hybridization to the sample DNA

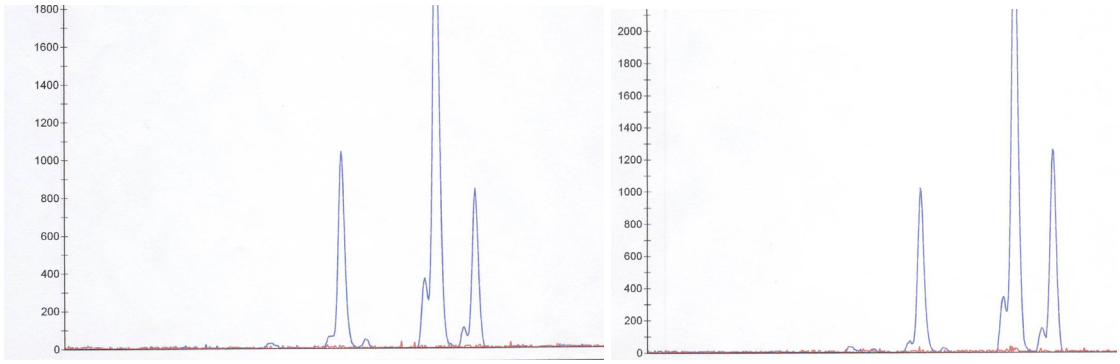
allows ligation of the two segments, thus providing templates that can be amplified with the pair of universal primers. Different probes are distinguished by size, the differences assured by the introduction of spacers.

Briefly, MLPA reaction consists of five steps; 1) initially DNA is denatured and incubated overnight with the MLPA probes, with each probe containing one of the PCR primer-oligonucleotides, 2) when the two probes are both hybridised, ligation occurs, 3) the ligated probes are then amplified in the subsequent PCR reaction, 4) the amplification products are separated by capillary electrophoresis; 5) and finally data are analyzed. Since only ligated probes are amplified, the number of probe ligation products reflects the number of the target sequences in the sample. Non-ligated probes only contain one primer sequence attached to each unligated segment; thus are not amplified and do not generate a signal. A peak pattern is obtained for each sample and is compared to that of reference control samples [98,99].

The MLPA experiment was carried out as following: 200 ng of cDNA from the 40 LCLs was generated as described above *(section 4.1 B. Reverse transcription of the ribosomal fractions)* and diluted in TE buffer (Tris-EDTA) in a total volume of 5 µl. Samples were denaturated at 98°C for 5minutes and subsequently allowed to cool down at room temperature. *RPS26*-specific probes were manually designed by the research assistant Luc Marchand. Every probeset contains the required primer pairs for all the locations that we want to interrogate. In our case, there were a total of 6 primers: 2 primers hybridising to the boundary of the 1st and 2nd *RPS26* exon, 2 primers hybridising to the boundary of the 3rd and 4th *RPS26* exon, and finally 2 primers for the housekeeping microglobulin M_2 gene. Every primer was resuspended in 100 µM with water, 1:100

dilution was prepared for each primer at a final concentration of 1 µM and 0.8 µl of each diluted primer were mixed in a final volume of 600 µl filled up with nuclease-free water. 1.5 µl of probemix was then added to each sample along with 1.5 µl of Salt solution (1.5 M KCl, 300 mM Tris pH 8.5, 1.5 mM EDTA). Samples were incubated at 95°C for 1 minute and left overnight (16 hours at least) at 60°C. Next morning, continuing to the second step of the MLPA experiment, the ligation reaction mixture was prepared by mixing 10x Taq ligase buffer, Taq ligase enzyme (4U) and nuclease-free water in a total volume of 32 µl. Samples were incubated at 54°C for 15 minutes and then at 98°C for 5 minutes. PCR reaction followed: 10x PCR buffer and 50 mM $MgCl_2$ were prepared in a total volume of 30 µl per sample. After heating the samples at 60°C a second PCR mix was added, consisting of 0.5 µM of each primer (Salsa left and right), 0.1 mM dNTPs and 0.5 µl Taq polymerase. The PCR amplification reaction was done as following: 33 cycles at 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute. 1 µl of each PCR product was then mixed with 10 µl formamide and 1 µl of a five dye-labeled size standard (GeneScan 600, Applied Biosystems, California, USA). After incubation at 80°C for 2 minutes and immediate placing on ice, samples were loaded on the ABI 310 Genetic Analyzer (Applied Biosystems, California, USA) in a loading voltage of 1.6Kv. A peak pattern for each sample was obtained *(figure 5A)* allowing measurement of the RNA level present in each sample. After normalizing for the house-keeping microglobulin B2M, the results were plotted against the number of G alleles in the genotype (0, 1 or 2) *(figure 5B)*.

A.



B.

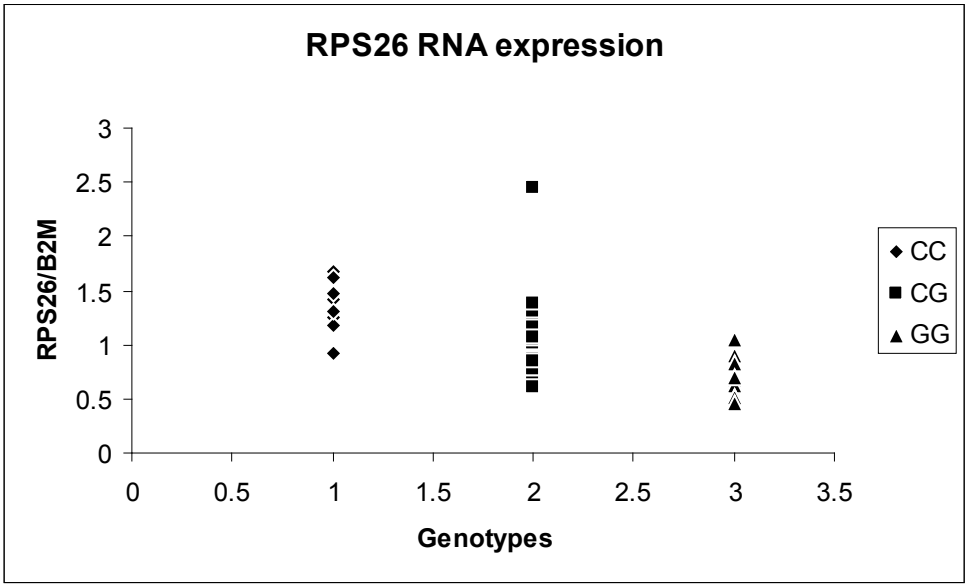


Figure 5: Multiplex Ligation-dependent Probe Amplification (MLPA) applied to measure the RPS26 levels present in each mRNA transcript of our cohort of 40 LCLs samples. Panel A shows two representative samples; the first one bearing CC/GG genotype and the second one CC/CC. Two different probe sets were used; the first probe hybridises to the boundary of the 1st-2nd RPS26 exon (first peak in the graph) and the second one to the boundary of the 3rd-4th exon (second peak). The third peak represents the housekeeping microglobulin M2 that was used for normalization of the protein expression. The ratio of

the first RPS26 peak versus the microglobulin peak was calculated, giving the normalized RPS26 RNA expression. The two graphs in panel A clearly show a distinct pattern: the RPS26 peak is higher than the microglobulin peak in the CC/GG individual, while this pattern is reversed in the CC/CC individual. In panel B, the results are plotted against the number of G alleles in the genotype (0, 1 or 2).

4.2 B. Methods for *in vitro* translation quantification

4.2 B1. Plasmid constructs

Plasmid constructs containing three different haplotype combinations were made to study *in vitro* the effects of the mRNA alleles on translational efficiency. Haplotype refers to the allelic combinations of the two adjacent SNPs found on a single chromosome. Given the fact that the rs17118262 SNP is a less common variant, we did not have enough representative samples per genotype for this SNP in our CEPH cohort. Particularly, what we were missing due to its rare frequency in nature was the G allele of this SNP. The 40 LCLs that were previously used in the *ex vivo* approach had only 2 samples with CG genotype in the rs17118262 SNP and none with GG genotype. Since the translational effects of the rs17118262 SNP could not be thoroughly investigated by the ex vivo approach – due to its rare frequency- we decided to study it in vitro by constructing the appropriate haplotypes.

Three different haplotypes were made: CC (C allele for the rs17118262 SNP and C allele for the rs1131017), CG (C allele for the rs17118262 SNP and G allele for the rs1131017) and finally GC (G allele for the rs17118262 SNP and C allele for the

rs1131017). Each haplotype was constructed twice, each time with a different tag epitope; *flag* or *cmyc*.

TA cloning was done using the pGEM vector (Promega, WI, USA). TA cloning is a technique that relies on hybridization of adenine (A) at the 3' end of a PCR product to complementary 3' thymine (T) of a linearized vector [100]. This work had started prior to my arrival to the laboratory. I took over and finished it. Briefly, polymerase chain reaction (PCR) was done using *RPS26* primers designed in *Generunner* to amplify both SNPs in the 5'UTR. The PCR products were sequenced and samples with the desired allele combinations for the two SNPs were selected for cloning. The rapid transformation procedure using ampicillin selection was done using the One Shot TOP10 Chemically Competent E.coli (Invitrogen, CA, USA) following the manufacturer's instructions. Bacteria grew on a pre-warmed selective agar plate overnight and 5 colonies were selected for further analysis of their insert.

The Qiagen Plasmid Mini Kit (Qiagen, CA, USA) was used following the manufacturer's instructions to prepare pure plasmid DNA. Pelleted bacteria were initially processed in an alkaline lysate followed by binding of the plasmid DNA to an anion-exchange resin under appropriate low-salt and pH conditions. A medium-salt wash buffer was then applied to remove RNA, proteins and dyes, and plasmid DNA was eluted in a high-salt buffer. Finally, isopropanol precipitation allowed pure plasmid DNA generation, suitable for further downstream investigations. Concentration of the eluted DNA was assessed by optical density.

The pGEM constructs of *RPS26* and either *flag* or *cmyc* epitope tag had a total size of 3907 base pairs (bp). The *RPS26* transcript was 723 bp. Using the nebcutter

software *(http://tools.neb.com/NEBcutter2)* two enzymes were selected: BamH1 and ScaI, with the sizes of the expected bands at 1,792 bp and 2,115 bp. Double digestion reaction was carried out in a total volume of 20 µl as following: 10X buffer3, 100x BSA (Bovin Serum Albumin), 0.5 µl of each enzyme, 0.5 µl of the each DNA plasmid filled up with water. Samples were incubated at 37°C for 90 minutes. Agarose gel electrophoresis was done to assess proper digestion. Samples were loaded to 1% agarose gels allowing direct comparison of the digested and undigested plasmid *(figure 6)*. Colonies in which the expected bands were seen, were subsequently analyzed by sequencing to ensure proper construction of the desired haplotypes.

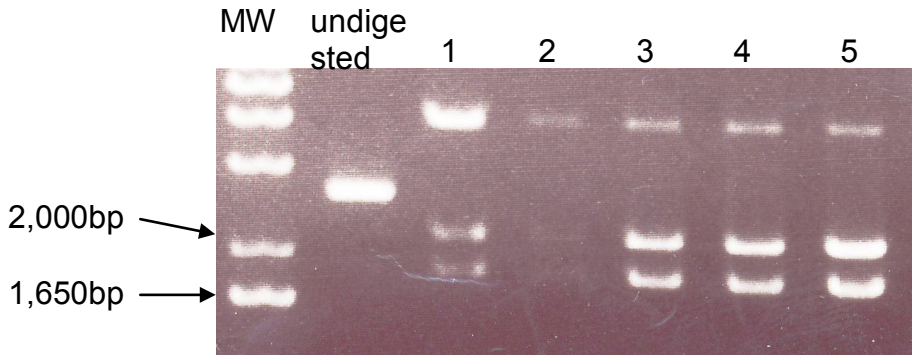


Figure 6: Initial screening of the colonies grown after the cloning procedure; for each plasmid construct, 5 colonies were selected and incubated with BamH1 and ScaI. The sizes of the expected bands are at 1,792 bp and 2,115 bp. The molecular ladder (MW) is shown in the first column followed by the undigested plasmid. The five numbered columns represent the five different colonies of this haplotype that were randomly selected for initial screening by double enzyme digestion. Colonies with successful digestion (in this case,

colonies 1, 3, 4, 5) were then analyzed by sequencing and one of them was used to generate large amounts of the plasmid construct.

Qiagen Maxi Preparation Kit (Qiagen, CA, USA) was then used to generate large amounts of each construct. The DNA plasmids were again verified by both double enzyme digestion and sequencing analysis and the corresponding bacterial colonies were stored at -80°C (1,5 ml of glycerol and 0.5 µl of bacteria). The eluted pure DNA plasmids of each haplotype (CCflag, CGflag, GCflag, CCmyc, CGmyc, GCmyc) were temporarily stored at -20°C.

4.2 B2. Plasmid linearization and DNA purification

The six plasmid constructs, each one containing a different haplotype-tag combination (CCflag, CGflag, GCflag, CCmyc, CGmyc, GCmyc) were linearized with BamH1 enzyme. The reaction mixture was carried out in total volume of 50 µl consisting of 10 mg plasmid DNA, 10x buffer 3, 100x BSA and nuclease-free water. The reactions were incubated at 37°C for 2 hours. The digestion was verified by agarose gel electrophoresis; direct comparison with the undigested plasmid was performed *(figure 7)*. If unlinearized plasmid was present *(figure 7)*, the samples were re-digested by adding 1 µl of additional BamH1 enzyme and incubating the reaction at 37°C for an additional 2 hours. Complete digestion was eventually verified and a clean non-degraded DNA fragment was seen in the agarose gel. The DNA templates were then purified using the QIAquick PCR Purification Kit (Qiagen, CA, USA), following the manufacturer's instructions. 5 volumes of binding buffer were added to 1 volume of DNA template and placed in QIAquick spin column

(provided in the kit). Samples were centrifuged at 13,000 rpm for 1 minute, the flow-through was discarded and wash buffer was then added. The same procedure was repeated and 10 mM Tris-Cl, ph 8.5 was finally added eluting purified DNA. The concentration of the purified linearized DNA plasmid was then calculated by optical density measurement. Integrity of the purified DNA fragment was assessed by agarose gel electrophoresis. The DNA templates for each sample before and after purification and before and after digestion by BamH1 were loaded in 0.8% agarose gel to ensure the presence of nondegraded DNA fragment at the expected size *(figure 8).*

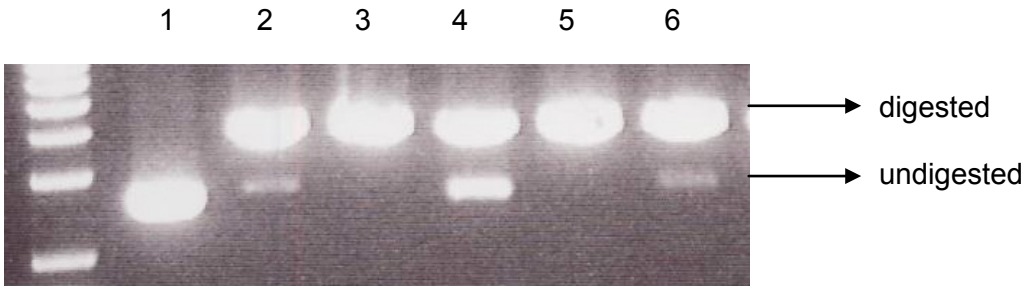


Figure 7: Linearization of the plasmid constucts with BamH1. Direct comparison with the undigested plasmid (column 1) was performed by agarose gel electrophoresis. If unlinearized plasmid was present (columns 2, 4, 6), the samples were re-digested as described in the text until complete digestion was eventually verified and a clean non-degraded DNA fragment was seen in the agarose gel.

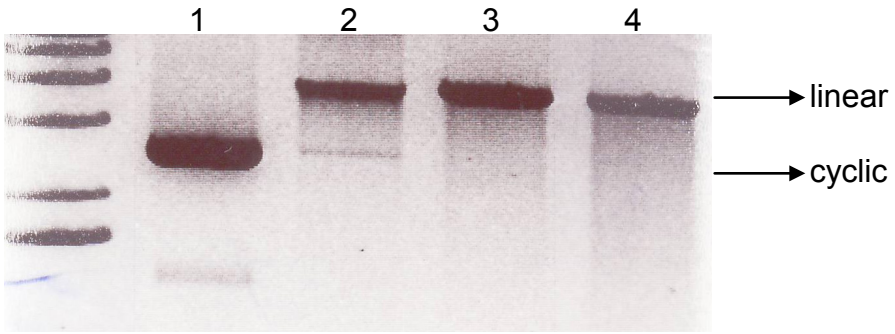


Figure 8: Agarose gel electrophoresis assessing the DNA plasmid constructs after BamH1 linearization and purification process. Initial plasmid construct (column 1) was linearized by BamH1 (columns 2, 3). After complete digestion occurred (column 3), the plasmid was purified (column 4) as described in the text.

4.2 B3. In vitro transcription

RNA from the six haplotype combinations was generated using an *in vitro* transcription kit (Ribomax Large Scale RNA Production–T7, Promega, WI, USA). The *in vitro* transcription reaction was carried out in a total volume of 100 µl as following: T7 transcription 5X buffer, 25 mM of each ribonucleotide (equal volumes of the 4 individual 100 mM rNTPs, ATP, CTP, GTP and UTP, were mixed in a total volume of 30 µl per reaction), 10 µl T7 enzyme mix, and 5 mg of linear DNA template plus nuclease-free water up to 40 µl per reaction. Linear DNA template from each one of the six haplotype combinations (CCflag, CGflag, GCflag, CCmyc, CGmyc, GCmyc) was used. The exact volume of the DNA was calculated to a final concentration of 5 mg/µl based on the optical density concentration generated after the purification of the linearized plasmid constructs. Reactions were set up at room temperature as spermidine, a component of the transcription

buffer, can precipitate DNA in colder temperatures. The reaction mixture for each sample was incubated at 37°C for 2 hours. After the transcription reaction was complete, samples were immediately subject to DNAase digestion to remove contaminating DNA template, thus allowing accurate determination of the RNA concentration.

4.2 B4. DNase digestion

The DNA-free kit by Ambion (Invitrogen, CA, USA) was used following the manufacturer's instructions. 0.1 volume of 10X DNase I Buffer and 1 µl of recombinanat DNase I were added to the previously *in vitro* transcribed RNA (100 µl). The solution was mixed gently and incubated at 37°C for 1 hour. 0.1 volume of DNase Inactivation Reagent was then added to subsequently remove the DNase and divalent cations from the sample. The mixture was incubated at room temperature for 3 minutes and centrifuged at 10,000 g for 2 minutes. During centrifugation, the DNAase Inactivation Reagent formed a pellet at the bottom of the tube. The supernatant that contains the clean RNA was carefully transferred to another tube.

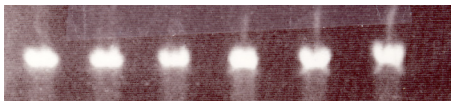
4.2 B5. Visualizing RNA by electrophoresis

The DNAase-treated *in vitro* transcripts were assessed in a denaturing gel *(figure 9A)* to verify the integrity of the full-length transcript after the DNase treatment. 19 µl of formaldehyde/formamide buffer was added to 1 µl of each RNA sample. Samples were denaturated at 65°C for 5 minutes. 3 µl of RNA loading buffer was added to each sample before loading them to the denaturing gel (e-gel EX 1% agarose, Invitrogen, CA, USA).

Denaturing gels were preferred as they provide the greatest resolution of the denaturated RNAs.

4.2 B6. RNA purification

The *in vitro* produced DNAase-treated RNAs were purified using the RNeasy Plus Mini Kit (Qiagen, CA, USA). Following the manufacturer's instructions 20 mg/µl of each RNA sample was passed through a genomic DNA eliminator spin column (provided in the kit) and mixed with binding buffer. This column in combination with the optimized high-salt buffer allows efficient removal of genomic DNA. 70% ethanol was added to the flow-through and the sample was placed to an RNeasy spin column (also provided in the kit) where total RNA binds to the membrane and contaminants are efficiently washed away. After serial centrifugations at 10,000 rpm for 1 minute and subsequent discards of the flow-through, 40 µl of nuclease-free water were added, eluting purified RNA whose concentration was initially determined by optical density measurement. The RNA samples were then diluted with nuclease-free water at the same final concentration. Gel electrophoresis was again done to ensure the integrity of the RNA transcripts after the purification process *(figure 9B)*. Denaturing gels were used as previously described to verify the presence of the expected non-degraded band for all six RNA transcripts.

A. 

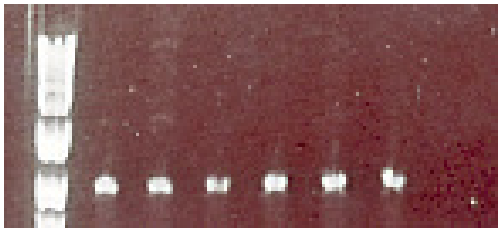
B. 

Figure 9: In vitro transcribed RNA samples from the six haplotypes after the DNasing process (panel A) and the purification process (panel B).

4.2 B7. RNA quantification

The Ribogreen RNA quantitation assay (Molecular Probes, OR, USA) was used to quantify the RNA samples. In our experimental design, achieving accurate and precise RNA concentration measurement was critical. In order to reach our goal of reliably comparing the RPS26 protein levels produced by the three different haplotypes (CC, CG and GC), we had to ensure that equal RNA amounts for all the three samples were used in the translation experiment. For this reason, ribogreen quantification was used instead of the suggested ultraviolet light absorbance. The RiboGreen reagent is a proprietary, unsymmetrical cyanine dye that is incubated with the samples for 5 minutes [101].

RNA concentrations are determined and reported based on the known concentrations of a standard DNA sample. A series of dilutions of known concentration of the standard sample are prepared and assayed with the ribogreen reagent alongside the unknown samples. The fluorescence produced by each sample is then measured using a

fluorescence microplate reader at 485/535 nm wave-length. A standard curve is obtained correlating fluorescence measurement to RNA concentration. Using the formula generated by the standard curve, the RNA concentration of the samples is determined *(figure 10).* In all our experiments, the r^2 of the ribogreen standard curves reached >0.989. All the three RNA haplotypes (CC, CG, GC) in every experiment exhibited a similar concentration of ≈ 250 ng/ul, ensuring that equivelant amounts of RNAs were used for the in vitro translation.

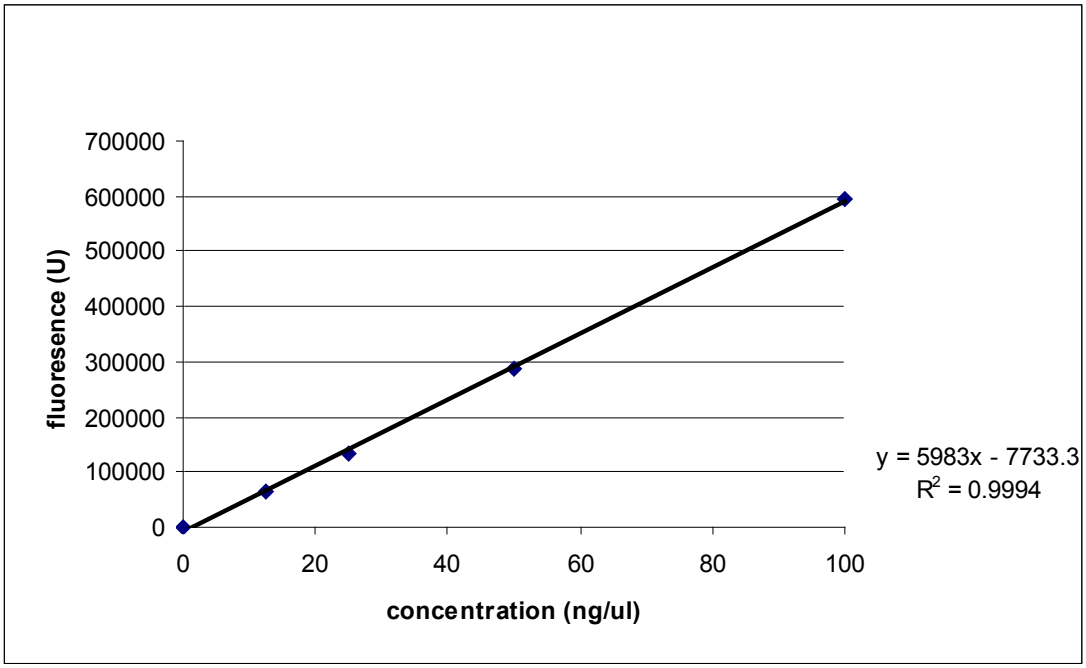


Figure 10: Quantification of the in vitro produced RNAs by the ribogreen assay. The assay displays a linear correlation between RNA concentration of the standard samples (y-axis) and the corresponding fluorescence measurement (x-axis). Using the formula generated by the graph, and based on the fluorescence each unknown sample produces, its concentration is accurately determined.

4.2 B8. Sequencing of cDNAs derived from the in vitro produced RNAs

Before proceeding to the *in vitro* translation experiment, and after having purified and accurately concentrated the RNA transcripts, the sequence of the RNA samples was confirmed by direct Sanger sequencing analysis. cDNAs for each sample were generated. In each sample 1 µg of RNA was reverse-transcribed into cDNA by use of Superscript II Reverse Transcriptase (Invitrogen, CA, USA). Each reaction was carried out in total volume of 20 µl, containing 10 mM d-NTPs (deoxyribonucleotide triphosphate; HT Biotechnology) and 200 U/ml Superscript in the presence of random hexamers (300 µg). RNA was initially denaturated at 70°C for 5 minutes and quick-chilled on iced water. The reaction mixture was then added and reverse transcription was performed at 37°C for 1 hour. *Generunner* program was used to design forward *RPS26* primer that amplifies our region of interest, containing both 5'UTR SNPs. 10 mM of the primer was used for each sequencing reaction. Sanger Sequencing confirmed that the *in vitro* produced RNAs derived from the 6 plasmid constructs (CCflag, CGflag, GCflag, CCmyc, CGmyc, GCmyc) contained the corresponding haplotype combinations.

4.2 B9. In vitro translation

The *in vitro* synthesis of proteins in cell-free extracts is frequently used to characterize mRNA products in variety of applications. *In vitro* translation systems are commonly preferred over *in vivo* gene expression particularly in cases where the protein is rapidly degraded by intracellular proteases, or when its over-expression is toxic to the host cells.

In our case, *in vitro* translation served as a totally independent approach to study translational efficiency and compare the results with the *ex vivo* protein studies.

Extracts from rabbit reticulocyte lysate is a commonly used *in vitro* translation system originally described by Pelham and Jackson [102]. Reticulocytes represent immature red blood cells that even if they have lost their nuclei, they still have enough mRNA and complete translation machinery. The vast majority of the protein made in reticulocytes is haemoglobin. To eliminate the endogenous globin mRNA, the *in vitro* rabbit reticulocyte system is being processed with Ca^{2+} -dependent micrococcal nuclease; thus reducing background translation. A variety of factors are added to the rabbit reticulocyte lysate system to enhance in vitro translation: hemin to prevent inhibition of the initiation factor eIF-2α, phosphocreatine kinase and phosphocreatine to generate the energy required by the system, and finally calf liver tRNAs to balance the accepting tRNA populations.

We used the Flexi rabbit reticulocyte lysate system (Promega, WI, USA) to produce *in vitro RPS26* protein from the three epitope-tagged allelic constructs of *RPS26* each one containing a different RNA haplotype (CC, CG, GC).

The translation reaction was initially optimized for Mg^{2+} and K^{+} concentrations. Given that different RNAs have different salt requirements that can lead to significant variations in translation, much optimization was required to finally achieve the optimal salt concentration for our *in vitro*-generated *RPS26* RNA transcripts. By directly contacting the company, we verified that each lysate batch we used, had the same endogenous Mg^{2+} concentration, ensuring exactly same conditions in every different experiment that was performed.

To reduce contamination of the samples by RNases and to prevent subsequent degradation, RNasin Ribonuclease Inhibitor (Promega, WI, USA) was added in each reaction.

The *in vitro* translation reactions were carried out as following: Reagents were rapidly thawed by hand-warming from -80°C where they were stored and immediately placed on ice. Reaction mixture for each sample (each of the six different haplotypes) contained: amino acid mixture minus leucine 1 mM, amino acid mixture minus methionine 1mM, potassium chloride 2.5 M, DTT 100 mM, RNasin Ribonuclease Inhibitor 40 U/µl, RNA substrate 1 mg, rabbit reticulocyte lysate 16.5 µl and nuclease-free water to a final volume of 25 µl. Reaction components were assembled in a 0.5 ml microcentrifuge tube and gently mixed by pipetting the lysate and stirring the reaction with the pipette tip. A control reaction containing water instead of RNA transcript was included. Three or more replicates for each different haplotype were made in every experiment. Each replicate was a different reaction performed in a separate tube with equal amounts of the corresponding RNA. Translation reactions were then incubated at 30°C for 90 minutes.

After completion of the translation reactions, equal amounts (25 µl) of loading buffer were added to each tube. Loading buffer contains 2% anionic denaturing detergent sodium dodecyl sulfate (SDS), 5% beta-mercaptoethanol, 15% glycerol and 3 M Tris-HCl pH 6.8 and is used to denaturate the protein's tertiary structure. Samples were then incubated at 100°C for 10 minutes and stored at -20°C until western blotting was performed within the following 3 days.

4.2 B10. Western Blotting

Western Blotting was done to quantify the *RPS26* protein produced by the different haplotypes. Samples were thawed from -20°C where they had been temporarily stored after the *in vitro* translation experiment. After vortex and centrifugation at 10,000 rpm for 2 minutes, 15 µl of each sample were loaded in the acrylamide gel. Western blotting was performed as previously described in the *ex vivo* experiments section.

Incubation with either *flag* or *cmyc* antibody allowed the detection of the *RPS26* band. Monoclonal *flag* M2-Peroxidase (HRP) antibody produced in mouse was purchased from Sigma-Aldrich, St Louis, USA. The antibody was diluted at 1:1000 with Tris Buffered Saline (TBS) containing 0.05 M Tris, pH 7.4, 0.15 M NaCl. Monoclonal *cmyc* HRP antibody produced in mouse was purchased from Invitrogen, CA, USA. The *cmyc* antibody was diluted at 1:2000 with TBS. The membrane where the *flag* constructs were loaded was incubated with *flag* antibody while a separate membrane with the *cmyc* constructs was incubated with the *cmyc* antibody, both at 4°C overnight. Next morning, several TBST washes on the agitator were performed to remove the residual antibody. Since both primary antibodies were HRP-conjugated, there was no need for incubation with a secondary antibody.

Western Lightning Chemiluminescence ECL Plus detection kit (PerkinElmer, MA, USA) was used to visualize the *RPS26* bands-tagged either with the *flag* or *cmyc* antibody. To confirm specificity of the *flag*- or *cmyc*- detected bands, the membranes were stripped and subsequently incubated with *RPS26* antibody *(figure 11)*. The stripping process was carried out as following: the membranes were placed in a sealable bag and incubated with 10 ml of stripping buffer (62.5 mM Tris, 2% SDS, 100 mM 2-

mercaptoethanol, pH 6.9) at 50°C for 30 minutes to remove the primary antibodies, *anti-flag* or *anti-cmyc*. *RPS26* antibody was prepared in 1:2000 dilution as described previously in the ex vivo section. *RPS26* protein is a ribosomal protein; thus it is naturally expressed in the rabbit reticulocyte system we used. The negative control reaction in the *in vitro* translation experiment, is expected to show an *RPS26* band; it represents the endogenous *RPS26* protein.

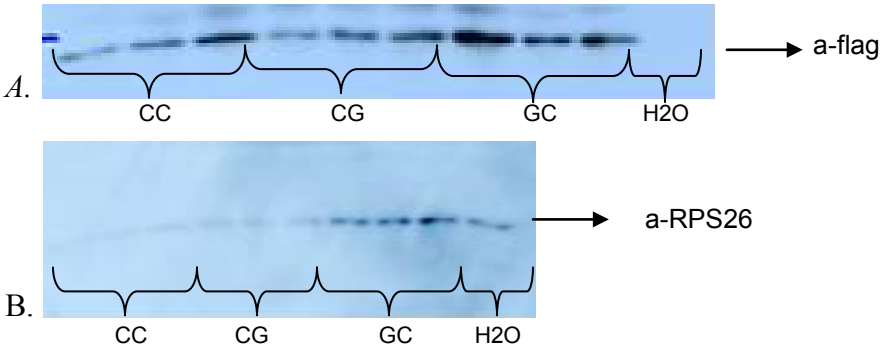


Figure 11: Western Blot of the flag constructs (CC, CG and GC) in triplicates. The same membrane is incubated with anti-flag antibody (panel A) and anti-RPS26 (panel B). The negative control reaction (labelled as H2O) has no flag tag and therefore no band is detected when the membrane is incubated with anti-flag. However, negative control reaction contains endogenous ribosomal proteins naturally expressed in the rabbit reticulolysate system; thus incubation with anti-RPS26 detects a band, as it was expected. Despite the fact that consistency is not perfect among the triplicates, there is a clear trend favouring the GC haplotype noted with both antibodies.

ImageJ software analysis was used to compare the Western bands reflecting protein production among the three haplotypes. In every experiment each haplotype (CC, CG and CG) was done in three or four replicates, with each replicate representing a separate *in vitro* translation reaction. After subtracting the background of the film, the mean densitometry value of the replicates per haplotype was calculated in each film and compared to the mean value of the other two haplotypes in the same film. Replicates of the same haplotype in different films were not directly compared to eliminate potential cofounding factors in the different western experiments (eg exposure time of the film).

5. Results

The initial Illumina transcriptome profile study *(HumanRef-8 v3.0 array)* of the ribosomal fractions of 40 individuals from the CEU set of the HapMap project identified the rs1131017 SNP as the variant being significantly associated with ribosomal distribution in the *RPS26* gene. However subsequent Sanger Sequencing revealed that the G allele of the neighbouring rare variant, rs17118262, is found at higher proportions in the heavy polysomes, suggesting that both 5'UTR SNPs affect ribosomal distribution and therefore translational efficiency. Our studies included both variants.

5.1 Independent confirmation of allelic imbalance in ribosomal distribution for RPS26 mRNA

Both pyrosequencing and SNuPE approaches confirmed that the same allele (G allele) of the rs17118262 SNP is found at higher proportions at the heavy polysomes. Individuals

who are heterozygotes for both SNPs showed no difference in the expression levels of the mRNA alleles in the total RNA and the light RNA fractions while the heavy RNA fractions showed preferential distribution of the G allele versus the C allele in the first SNP (rs17118262).

We had previously noticed a specific pattern of allele combinations in each chromosomal copy in all the double heterozygote samples; the G allele of the first SNP is always followed by the C allele in the second SNP. Accordingly, the other chromosome has C allele in the first SNP followed by G allele in the second SNP. Thus the percentages of the G allele in the first SNP and the C allele in the second SNP have to be similar since both are located in the same chromosome (*figure 12*). A representative pyrosequencing graph from a double heterozygous patient is shown in *figure 13*.

The translational imbalance favouring the G allele of the first SNP was confirmed by analyzing patients who are heterozygotes only for the second SNP (rs1131017) and patients heterozygotes only for the first SNP (rs17118262). When homozygosity in the first SNP was seen (CC/CG), the previously observed translational imbalance was lost; both the heavy and the light ribosomal fractions as well as the total RNA showed identical expression levels of the two mRNA alleles *(figure 14)*. On the other hand, single heterozygotes for the first SNP (CG/CC) showed consistently higher expression of the G allele in the heavy fractions while there was no difference in the levels of the two alleles in the light fractions *(figure 15)*.

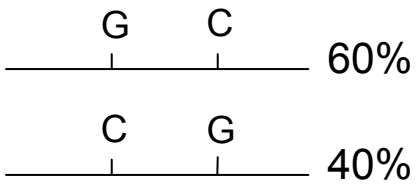


Figure 12: A specific combination of alleles was noticed in all the double heterozygote samples; the G allele of the first SNP is always followed by the C allele in the second SNP. Accordingly, the other chromosome has C allele in the first SNP followed by G allele in the second SNP. Thus the percentages of the G allele in the first SNP and the C allele in the second SNP have to be similar since both are located in the same chromosome. The same principle applies to the expression levels for the C allele in the first SNP and the G allele in the second SNP; their percentages have to be similar.

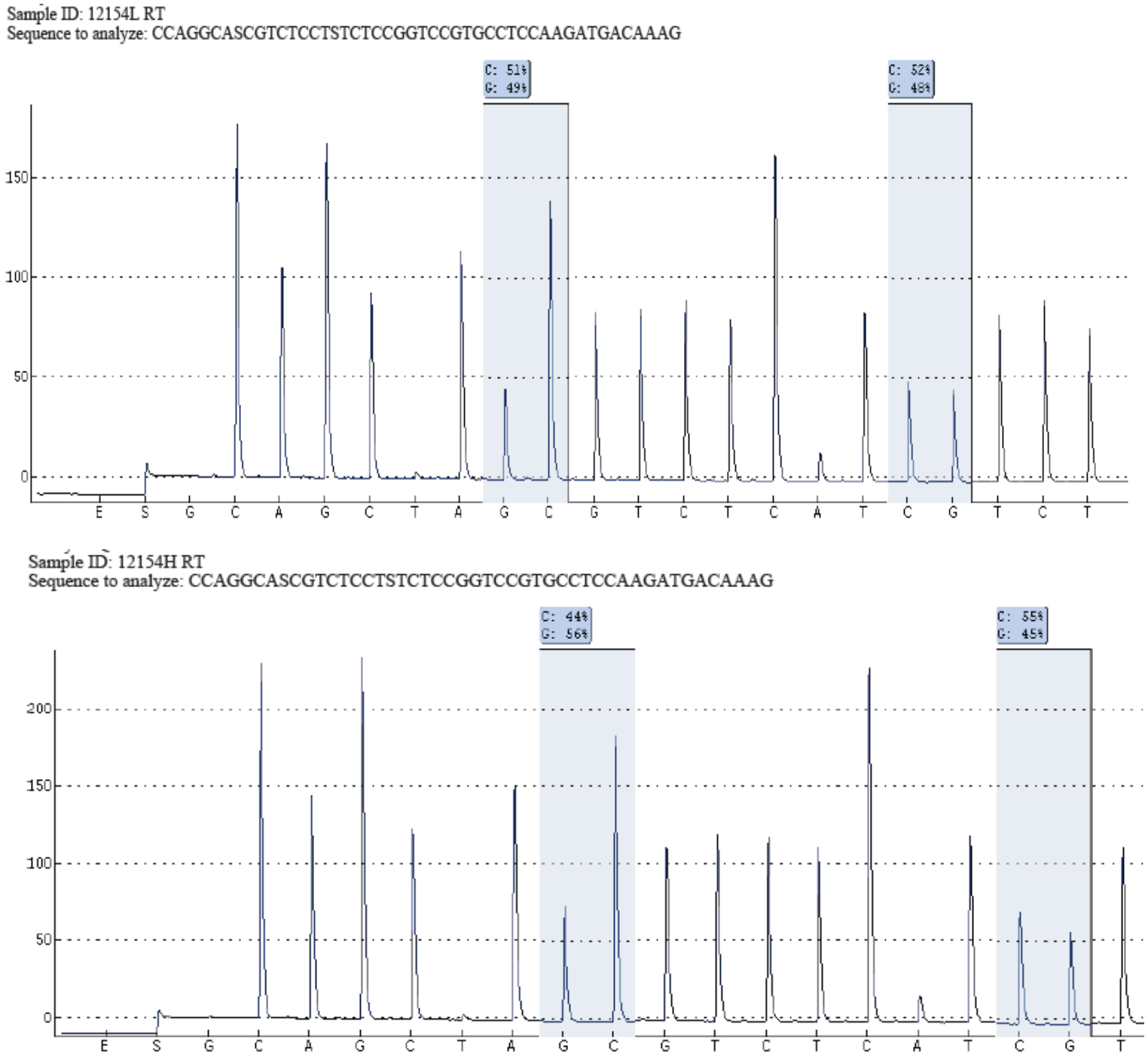


Figure 13: Pyrograms of the light (panel A) and heavy (panel B) ribosomal fractions in a representative double heterozygote individual (CG/CG). There is no difference in the expression levels of the mRNA alleles in the light RNA fraction (pyrogram A) while the heavy RNA fraction (pyrogram B) shows preferential distribution of the G allele versus the

C allele in the first SNP. The expression level of the G allele in the first SNP (56%) is similar to the expression level of the C allele in the second SNP (55%) as expected (figure 12).

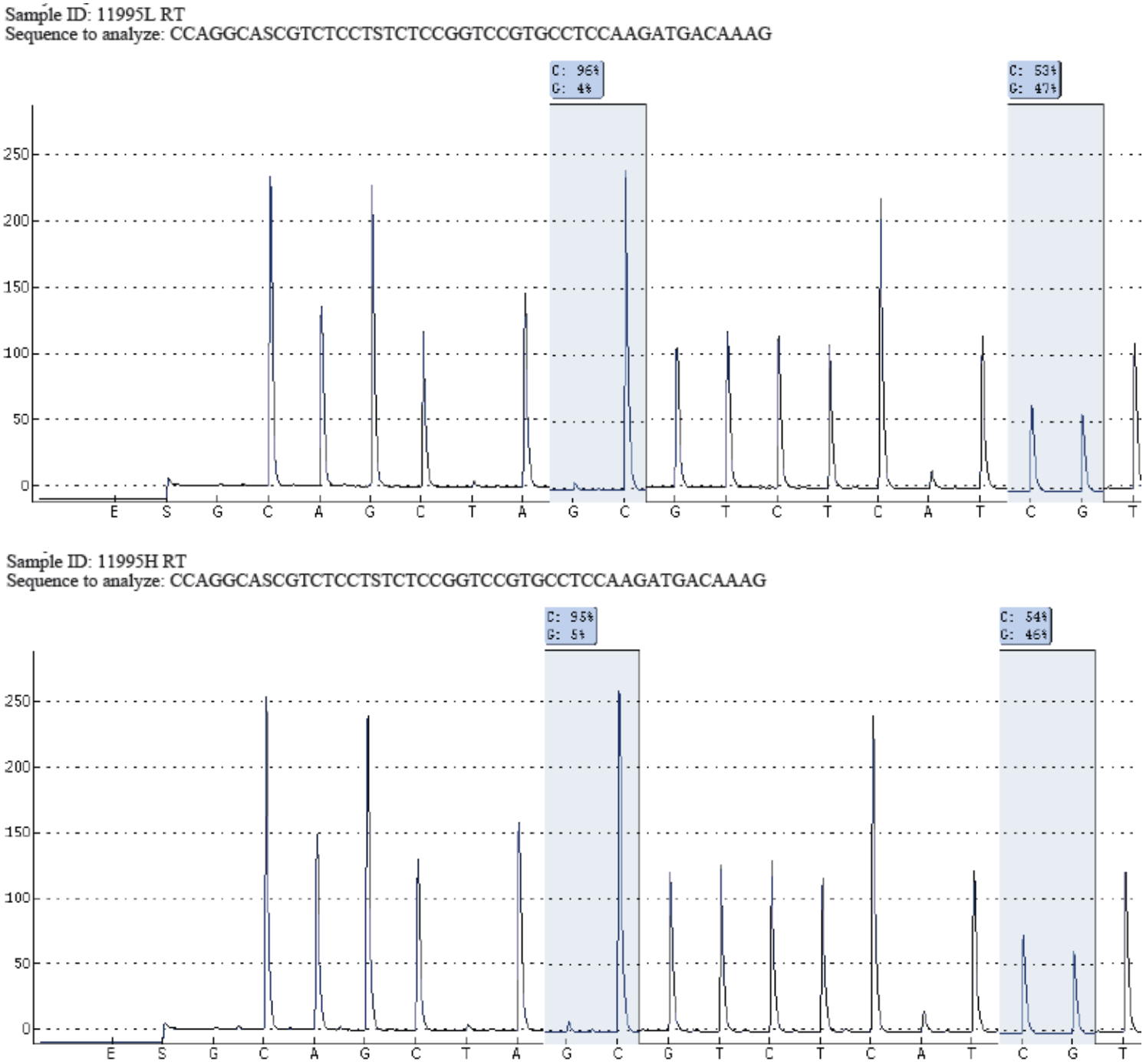


Figure 14: Pyrograms of the light and heavy ribosomal fractions in a representative

CC/CG individual. The previously observed translational imbalance is lost when there is homozygosity in the first SNP (rs17118262); both the heavy and the light ribosomal fractions show identical expression levels of the two mRNA alleles.

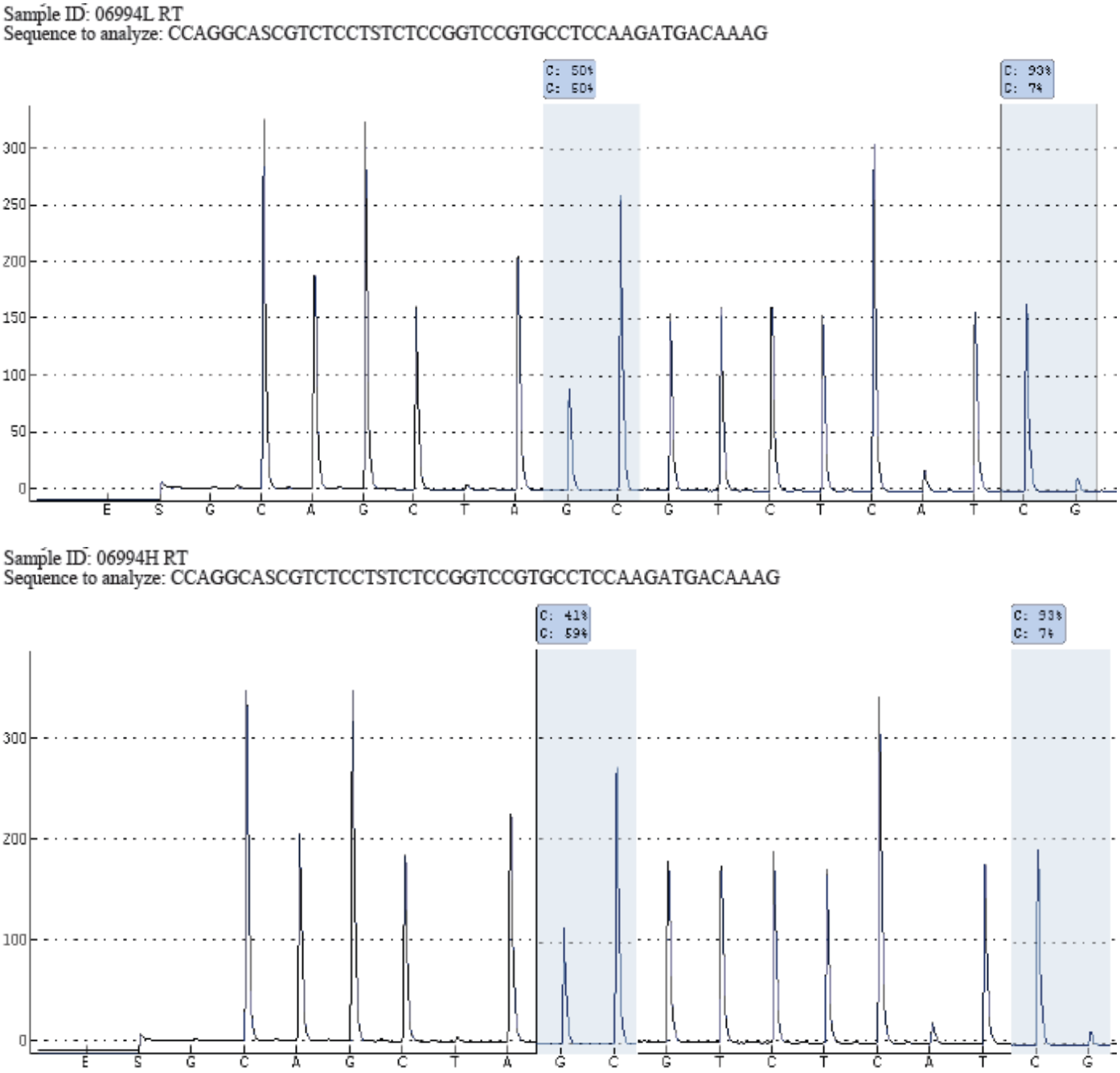


Figure 15: Pyrograms of the light and heavy ribosomal fractions in a representative

CG/CC individual. There is consistently higher expression of the G allele in the heavy fractions while there is no difference in the levels of the two alleles in the light fractions.

The pyrosequencing results from all the samples per genotype were plotted in separate graphs (one graph per genotype); the x axis represents the ribosomal populations, the heavy and light fractions, and the y-axis represents the manually calculated ratio of the G allele versus the C allele (the percentage of the G allele in each pyrogram was divided by the percentage of the C allele in the same SNP). Statistical analysis was applied revealing statistical significance (paired t-test=0.00045) in the distribution of the G allele versus the C allele (ratio>1) in the double heterozygote individuals (CG/CG). Statistical significance was lost when the same statistical analysis was applied to compare the G/C allele ratio of the heavy versus the light fractions in individuals bearing the CC/CG genotype *(figure 16)*.

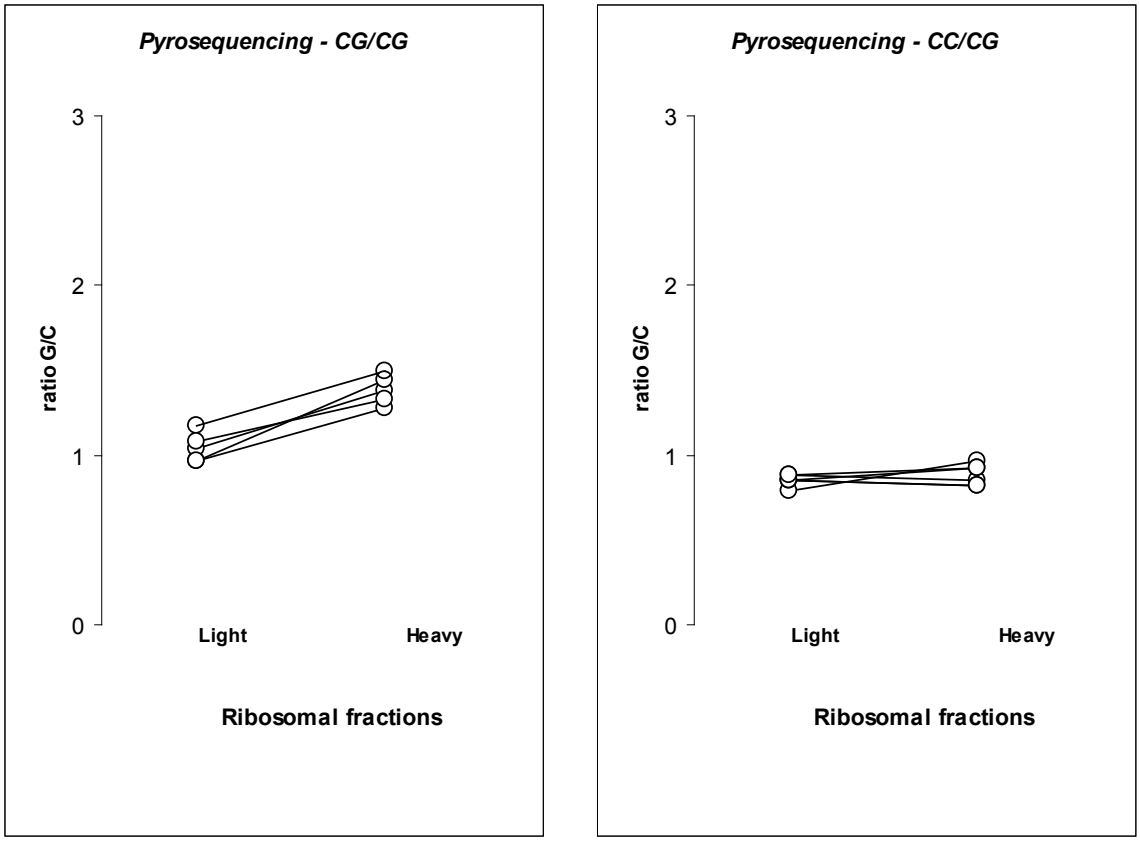


Figure 16: Graphs plotting the pyrosequencing data; the x axis represents the ribosomal populations, the heavy and light fractions, and the y-axis represents the manually calculated ratio of the G allele versus the C allele. Statistical significance was revealed in the distribution of the G allele versus the C allele (ratio>1) in the double heterozygote individuals (CG/CG genotype) when comparing the heavy polysomal fractions to the light ribosomes (paired t-test=0.00045) (graph A). Statistical significance was lost when the same statistical analysis was applied to compare the G/C allele ratio of the heavy versus the light fractions in individuals bearing the CC/CG genotype (graph B).

Summarizing the results of Sanger sequencing, SNuPE and pyrosequencing, all the 3 approaches showed that the same allele (G allele) of the rs17118262 is found at higher proportions at the heavy polysomes. We assumed this is happening due to more efficient translation and we hypothesized that the mRNA allele found at a higher proportion in the polyribosomal fraction than in the soluble RNA will produce more protein product under the same translational environment. The absence of imbalance in the light fraction is probably due to the much smaller proportion of mRNA in the heavy fraction, consequent to which the concentration of the G allele in that fraction results in a relatively small decrease in the light fraction that was lost in the noise of measurement error.

5.2 Confirmation of translational effects at the protein level

To test the hypothesis that mRNA with the G allele produces more protein, we applied both an *ex vivo* and *in vitro* translation approach that would allow us to quantify *RPS26* protein derived by two independent sources.

5.2 A. Results for ex vivo translation quantification

The graph plots with the *RPS26* RNA quantification data *(figure 5B)* and the *RPS26* protein quantification data *(figure 4B)* from the 40 LCLs samples were analyzed based on their genotype. Combining the graph plots of the protein and RNA results, we compared the mRNA-normalized protein levels of *RPS26* between the three genotypes *(figure 17)*. Statistically significant correlation (p = 0.005) was only found when data were plotted against the number of G alleles (0, 1, or 2) in the genotype of the common SNP, rs1131017, that was initially identified in the microarray chip results of the transcriptome

profiling study as the variant being most significantly associated with ribosomal distribution.

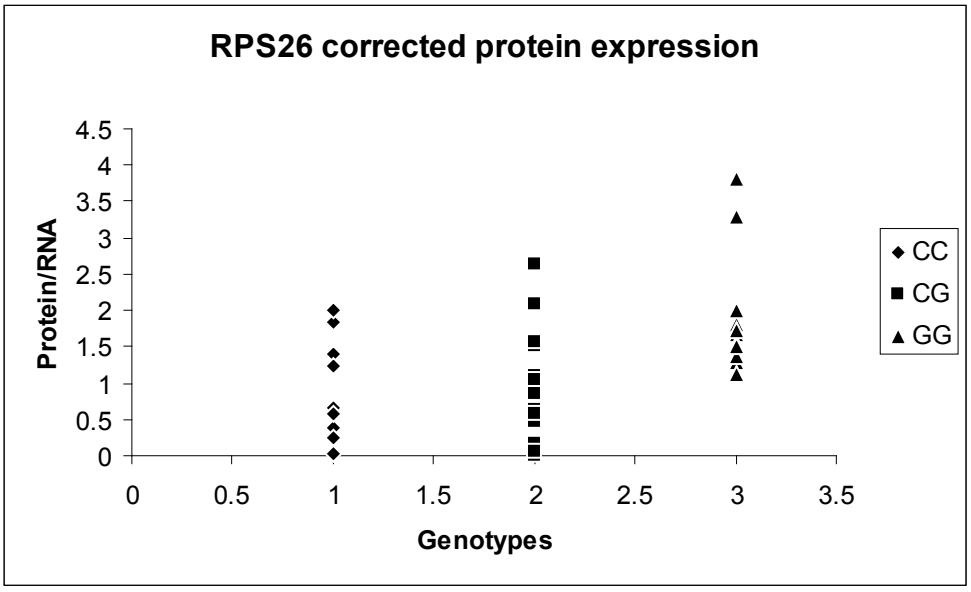


Figure 17: Combining the graph plots of the protein (figure 4B) and RNA results (figure 5B), we compared the mRNA-normalized protein levels of RPS26 between the three genotype. Statistically significant correlation (p = 0.005) was reached when data were plotted against the number of G alleles (0, 1, or 2) in the genotype of the common SNP, rs1131017.

The *ex vivo* results show that G allele of the rs1131017 produces significantly more protein product regardless of the genotype of the other SNP (rs17118262). When the data were plotted taking into account the genotype of the rs17118262 SNP, statistical significance could not be reached.

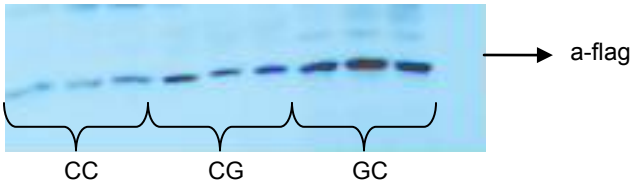
According to those results the rs1131017 variant has a statistically significant effect on translational efficiency that is independent of the rs17118262 SNP. However,

special consideration needs to be given to the fact that the rs17118262 SNP is a low-frequency variant; there were not enough representative samples per genotype. The GG combination of alleles for the rs17118262 SNP does not exist in our HapMap cohort and the CG combination for the rs17118262 with C homozygosity in the rs1131017 (CG/CC) exists only in 2 samples. Since the translational effects of the rs17118262 SNP could not be thoroughly investigated by the *ex vivo* approach – due to its rare frequency- we decided to study it in vitro by constructing the appropriate haplotypes.

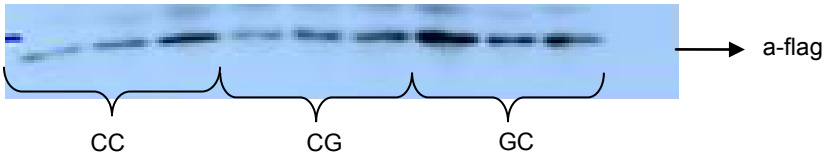
5.2 B. Results for in vitro translation quantification

We constructed *in vitro* three different haplotype combinations: CC (C allele for the rs17118262 SNP and C allele for the rs1131017), CG (C allele for the rs17118262 SNP and G allele for the rs1131017) and finally GC (G allele for the rs17118262 SNP and C allele for the rs1131017). Quantification of the bands detected by the *flag* antibody revealed that the GC haplotype produces significantly more protein. Using the ImageJ software analysis, densitometry value for each individual replicate was calculated. After subtracting the background density, the average value for each haplotype within the same blot was calculated and compared to the average value of the other two haplotypes. Replicates of the same haplotype in different western experiments were not directly compared to avoid cofounding factors from the different experiments. Statistical analysis was applied to reveal potentially significant correlation between haplotype combinations and *RPS26* protein product, as measured by the densitometry analysis. GC haplotype was found to produce significantly more protein, *p=0.05 (figure 18).*

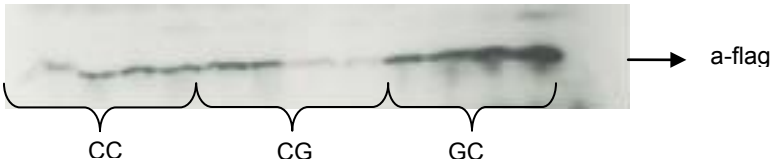
A.


a-flag
CC
CG
GC

B.


a-flag
CC
CG
GC

C.


a-flag
CC
CG
GC

D.

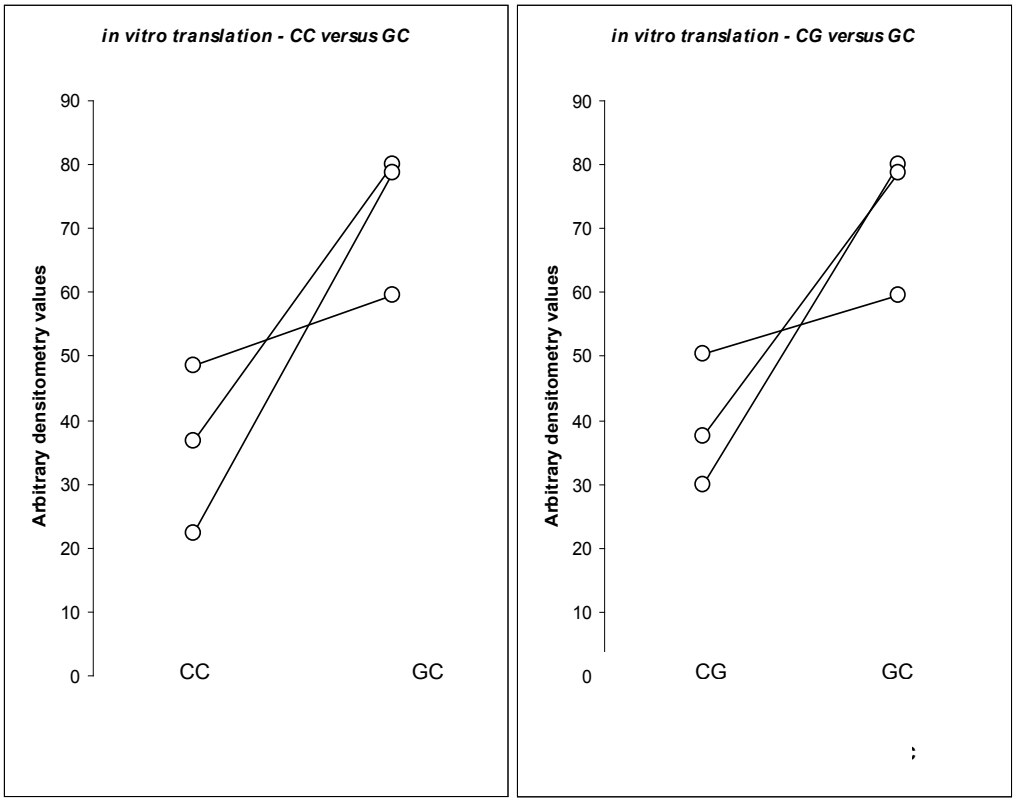

in vitro translation - CC versus GC
Arbitrary densitometry values
90
80
70
60
50
40
30
20
10
0
CC
GC
in vitro translation - CG versus GC
Arbitrary densitometry values
90
80
70
60
50
40
30
20
10
0
CG
GC

Figure 18: Comparison of the RPS26 protein product in the flag constructs (CC, CG, GC) in three different western blot experiments, all incubated with flag antibody. In each experiment, multiple replicates of each haplotype were done (panels A-C). ImageJ software enabled the quantitative comparison of the RPS26 bands among the different haplotypes. The mean value of each haplotype in the same blot was calculated and compared with the mean value of the other two haplotypes (panel D). Paired t-test was done to compare GC with each of the other two haplotypes (CC and CG). GC was found to produce more protein product (p = 0.05).

Consistency among the different replicates of the same haplotype was not perfect in all our western experiments. Part of this inconsistency can be explained by the nature of the *in vitro* translation experiments. Firstly, the rabbit reticulocyte lysate we used is not a homogeneous solution, but a rather heterogeneously-prepared lysate with ribosomal molecules composing the translationary machinery. On the other hand, the different replicates of each haplotype represent completely different reactions performed in separate tubes; thus it is impossible to predict that the exact same amounts of the lysated ribosomal molecules will be added in each tube. Therefore it was expected to have a range of densitometry values rather than bands of perfectly similar density.

When the G allele in the first SNP (rs17118262) is replaced by the C allele (CC and CG haplotypes), less protein is produced, suggesting that it is the G allele in the rs17118262 variant that exerts a significant effect on translational efficiency. However this effect seems to be dependent on the allele found in the other SNP (rs1131017). The CC and CG haplotypes do not produce the same amount of protein. Investigating whether

there is a statistically significant difference between the protein products of the CC and CG haplotypes, would allow us to examine the translational effect of the G allele in the rs1131017 SNP, thus elucidating the *ex vivo* results, where the G allele of the rs1131017 was found to produce significantly more protein product.

To prove this, we repeated the *in vitro* translation reactions of the CC and CG haplotypes in multiple different experiments. Both *flag* and *cmyc* constructs were used. *Figure 19* shows a representative western blot of the *cmyc* constructs initially incubated with the *cmyc* antibody; the film generated was quite blurry and difficult to quantify, while the same membranes after stripping and incubation with the endogenous *RPS26* antibody yielded a much cleaner film, suitable for quantification with the Image J software. As described above, the densitometry value for each individual replicate was calculated. After subtracting the background density and the endogenous *RPS26* band detected in the negative control reaction, the average value for each haplotype within the same blot was calculated and compared to the average value of the other haplotype in the same blot.

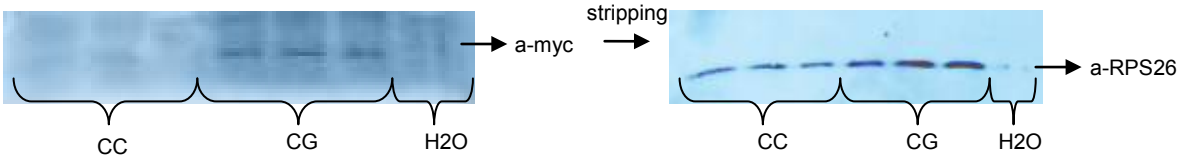


Figure 19: CC and CG cmyc constructs incubated with cmyc antibody; the film generated was quite blurry and difficult to quantify, while the same membrane after stripping and incubation with the endogenous RPS26 antibody yielded a much cleaner film.

Flag constructs were also incubated with the *RPS26* antibody *(figure 20)*. Following the same quantification analysis, the average densitometry value for each haplotype within the same blot was calculated and compared to the average value of the other haplotype. The quantification results of seven different western experiments were plotted in a graph; each membrane contained at least 3 replicates of each haplotype, CC and CG, with either the *flag* or *cmyc* constructs. Since the quantification results were based on incubation with the *RPS26* antibody and not the tag antibodies (*flag or cmyc)*, the *flag* and *cmyc* constructs for each haplotype are essentially the same and thus can be directly compared. Each tag is 8 to 9 aminoacids at the end of the *RPS26* protein, making no difference to the amount of protein that is produced.

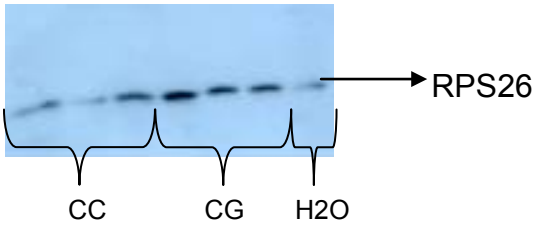


Figure 20: CC and CG flag constructs incubated with the RPS26 antibody. Quantification analysis was performed as described in the text and the average densitometry value for each haplotype within the same blot was calculated and compared to the average value of the other haplotype.

Paired t-test was applied to compare the CC and CG haplotypes in seven different western experiments with each membrane containing at least 3 replicates of each haplotype; statistical significance (t-test=0.02) favouring the CG haplotype was revealed

(figure 21). Therefore, it is shown that the G allele of the second SNP (rs1131017) has a significant effect on translational efficiency. This observation confirms the *ex vivo* results, where the G allele of the common variant (rs1131017) was found to be most strongly associated with translational efficiency.

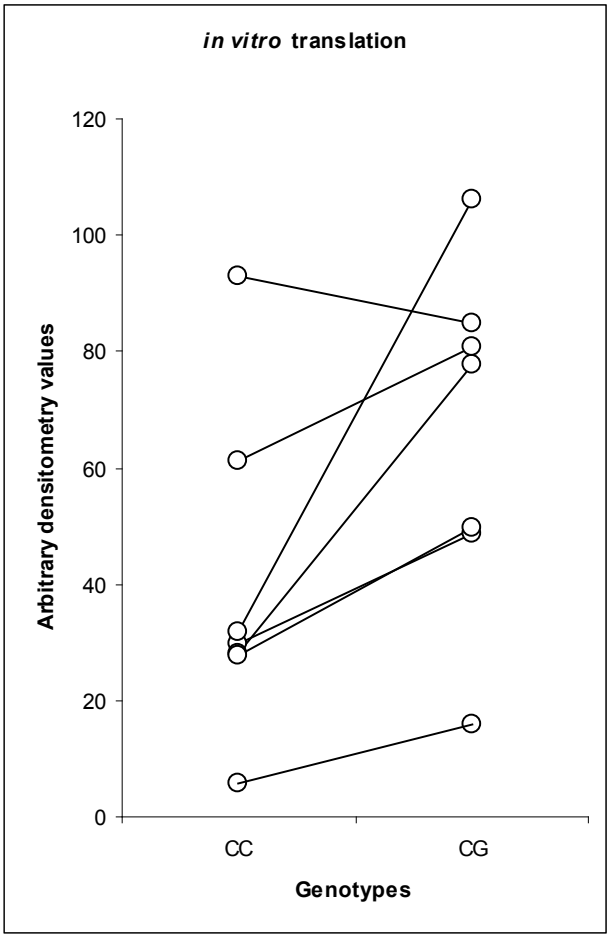


Figure 21: Comparison of CC versus the CG haplotype. A graph is plotted with the quantification data from the western blots shown in figures 19 and 20, all incubated with the RPS26 antibody. Each membrane contained at least 3 replicates for each haplotype (CC or CG) with either the flag or myc construct. Statistical significance favouring the CG haplotype was found (paired t-test=0.0204).

We focused on comparing the CC and CG haplotype to study the effect that the common SNP (rs1131017) has on translational efficiency. In this way we resolved the potentially contradicted results of the *ex vivo* and *in vitro* translation approach. The G allele of the common variant (rs1131017) was found to produce significantly more protein both in the *ex vivo* and *in vitro* approach (when comparing the CC and CG haplotypes).

Pyrosequencing and SNuPE independently confirmed the results of the *in vitro* approach; when comparing *in vitro* all the three haplotypes (CC, CG and GC) it was clearly shown that the GC haplotype produces significantly more protein, thus pointing to the first SNP (rs17118262) as the significant variant for translational efficiency. Both pyrosequencing and SNuPE showed that the G allele of the rs17118262 SNP is found at higher proportions at the heavy polysomes. Since the heavy polysomes are known to represent the most active part of the tranlsationary machinery, the G allele is predicted to produce more protein product.

Taking into account all the above data, it seems that both SNPs (the common rs1131017 and the rare rs17118262) exert a significant effect on the amount of RPS26 protein that is produced. It is difficult to quantify the translational effect coming from each individual SNP. However it is clearly shown that having one allele over the other makes a significant difference on the amount of protein being produced. In our study, it is the G allele in both SNPs that is associated with more protein product.

6. Discussion

In the past few years, the scientific community has witnessed enormous advances in the genotyping technologies that made feasible and affordable complicated studies of genome-

wide gene expression analysis. As a result, a large number of disease-associated loci have been identified in numerous complex human diseases, including type 1 and type 2 diabetes, cardiovascular diseases, psychiatric disorders and autoimmune diseases [103]. Despite the spectacular success in identifying new susceptibility loci, the underlying genetic mechanisms that mediate the reported association between a gene and the risk it confers to a particular disease, remain largely unknown [84,104].

Genome-wide association studies (GWAS) identify loci that have extended linkage disequilibrium (LD) for more than a few hundred kilobases and thus may include several different genes. A study illustrating this complexity is the T1D GWAS by our group [6]; the novel T1D susceptibility locus on chromosome 12q13 involves SNPs on a tight LD block extending over hundreds of kilobases that encompasses more than ten genes, including *RPS26*. There may be multiple variants in multiple genes in the same LD block that can explain the association because of LD. A gene closely located to the most associated variant in the block is not necessarily functionally involved in the particular disease [83]. Therefore it becomes challenging to clarify the causative gene within a locus. However, identifying genes with functional genetic variations offers significant help. A gene that has no functional variation, cannot be the cause of a genetic disease.

Most papers studying the function of DNA variations focus either on how the protein function is altered after an amino acid substitution caused by a SNP or on how the mRNA levels are changed by a regulatory SNP [105-107]. However, it is possible that neither an amino acid nor an mRNA change will be found in DNA variations, when the association is due to a translational effect. Due to the lack of appropriate methodology,

the effects that DNA variations exert on gene translation have been poorly investigated. However, this phenomenon might represent a significant portion of the gene expression regulation. Numerous studies in the literature, some of them were discussed in the introduction section, provide strong clinical correlations revealing the significance of the translational effect in the human genome caused by common DNA polymorphisms.

Despite the importance of this potential mechanism of genetic translational control, there is no systematic high throughput method to investigate the SNPs' effect on gene translation. Our group has previously applied a novel methodology to detect in genome-wide scale exonic polymorphisms that alter translational efficiency. Using ribosomal association as a proxy for translational efficiency 17,495 autosomal genes were screened by transcriptome profiling of ultracentrifugal ribosomal fractions of cell lines from 40 individuals to identify transcripts whose relative distribution between heavy polysomes (active translation) and soluble RNA (inactive) was skewed by single-nucleotide polymorphisms. One of the strongest hits was *RPS26* gene; its ribosomal distribution was strongly associated with the rs1131017 SNP, at a significance level that met the Bonferroni threshold of 4×10^{-9}, for the number of gene-SNP pairs examined. The exact same SNP also had the most significant association with T1D at the 12q13 locus in our genome-wide association study (GWAS) [6], also found independently in the GWAS by the Wellcome Trust Case-Control Consortium. Therefore, *RPS26* seemed the ideal candidate for downstream functional studies, given our laboratory's strong interest in the pathogenesis of T1D.

The central purpose of my thesis was to independently confirm at the protein level the allelic imbalance in ribosomal distribution detected in the *RPS26* gene. Our

downstream studies included both the rs1131017 initially identified in the transcriptome profile study as well as the rs17118262 variant that was revealed in the subsequent re-sequencing analysis. Ribosomal fractions of individuals who are heterozygotes for both SNPs showed that the G allele of the rs17118262 is found at higher proportions in the heavy polysomes, suggesting that this variant also plays a significant role on ribosomal distribution. Due to the close proximity of the two 5'UTR *RPS26* SNPs and their strong LD, we decided to include both SNPs in our investigations.

Each of the two SNPs rs1131017 (G/C) and rs17118262 (G/C) have two alleles. Therefore, four haplotypes are expected in general population, i.e. 1) rs1131017G-rs17118262G, 2) rs1131017G-rs17118262C, 3) rs1131017C-rs17118262G, and 4) rs1131017C-rs17118262C. If the two SNPs are not in LD, the probability of each haplotype in the general population is equal to the product of the frequencies of each SNP allele. If there is LD, the probability of a specific haplotype will be higher or lower than the expected random frequency [108]. Physical proximity of the two SNPs along with recombination events (D') and allele frequencies (r^2) are the main factors determining LD. When there is no recombination between the two SNPs (D'=1), we can only see three of the four haplotypes at most. Furthermore if the frequencies of the two SNPs are exactly the same (r^2=1), we can only see two haplotypes since one allele of one SNP always co-occurs with a specific allele of the other SNP. In our case, the two SNPs are located 9 bases apart, are in strong linkage disequilibrium (D'=1) but have different allele frequencies (r^2=0.54). The haplotype that is rarely found in the population is the GG haplotype. The GC haplotype is also relatively rare with only 2 individuals in our study

cohort. This natural limitation made us construct *in vitro* the rare allele combination in order to fully investigate the translational effects of the two SNPs.

Another limitation of our study lies in the biological role of *RPS26*. *RPS26* is part of the ribosomal translational machinery. Mutations in ribosomal proteins may lead to defects in ribosome synthesis, thus affecting ribosome quality which in turn may alter translational efficacy [72]. However, our study did not investigate the translational efficiency of the *RPS26* gene itself. It rather focused on the quantitative differences of the protein product among the different haplotype combinations of the same gene. We did not compare the translational efficiency of the *RPS26* gene versus other genes, but the effects of two polymorphisms on the same gene product, which happened to be *RPS26*. *RPS26* was initially selected for 2 critical reasons: firstly, its strong ribosomal distribution association ($p = 10^{-10}$) and secondly, the GWAS of our group had revealed that the rs1131017 has the most significant association with T1D at the 12q13 locus.

The question of whether this translation effect explains the T1D association remains open but we believe that it is unlikely, as Plagnol et al. showed a residual effect on T1D susceptibility after accounting for the effect of rs1131017. However it is possible, in the same recombination interval, to have two or even more variants independently conferring disease risk. A more important argument is the presence of a transcriptional effect that almost exactly counterbalances the translational effect we have found. The end result is that the amount of protein produced does not differ among the three genotypes (*Figure 4B*) and the effect on protein production can only be seen after adjusting for the amount of mRNA. The existence of two effects by the same genetic variant, exactly counterbalancing each other may be a coincidence, or it may reflect the high evolutionary conservation of precise

gene-dosage control required for *RPS26*. *RPS26* haploinsufficiency (50% gene expression) causes dominantly inherited Blackfan-Diamond anemia, consistent with the importance of maintaining the exact level of expression (most loss-of function mutations are recessive, because loss of 50% of expression level of most genes does not cause disease).

7. Conclusion

Both SNPs (the common rs1131017 and the low-frequency rs17118262) exert a significant effect on the amount of *RPS26* protein that is produced. While it is difficult to quantify the translational effect coming from each individual SNP, we clearly showed that having one allele over the other makes a significant difference on the amount of protein being produced. In our study, it is the G allele in both SNPs that is associated with more protein product.

Adding to the well-studied polymorphic transcriptional effects, the effects of exonic single-nucleotide polymorphisms (SNPs) on translational efficiency may be equally important in determining protein expression levels and therefore contributing significantly to the disease risk. *RPS6* in type 1 diabetes is an excellent example of such an association. Even if the particular variant I studied in my thesis may not explain the T1D association, my work has contributed to the development of a tool that can be applied genome-wide to elucidate mechanisms of action of complex-trait loci.

8. Future directions

It would be interesting to extend the approach to other selected genes, giving priority to genes mapping to confirmed complex-trait loci for which no mechanism is known (the

majority, at present). To see whether the effect is specific to the LCL translational machinery, we will repeat these studies in fibroblasts. It will be interesting to see whether the effect is reproducible for genes not normally expressed in fibroblasts. Finally, a scale-up of the genome-wide ribosomal screen will eventually be tried; from 42 subjects in the initial study, to 300, and from microarrays to transcriptome sequencing.

9. Acknowledgments

It is with great pleasure to thank all the people who made this thesis possible.

I am heartily thankful to my supervisor, Dr. Constantin Polychronakos, whose encouragement, guidance and support from the initial to the final level enabled me to develop and complete this project. His deep knowledge in the fields of human genetics and endocrinology inspired me to follow a career in the physician/scientist pathway focusing on the genetics of pediatric endocrinology diseases.

I owe my deepest gratitude to the research assistant of our laboratory, Luc Marchand. He taught me numerous techniques necessary for the experimental design of this project. He carefully designed and successfully executed the SNuPE and MLPA experiments. He always offered generously his technical expertise in multiple aspects of this project.

I am grateful to Dr Lu-Yang, former graduate student of Dr Polychronakos' laboratory. Dr Lu-Yang prepared the ribosomal fractions, as part of her M.Sc. thesis project. Despite she is not working any more to our laboratory, she always offered me valuable theoretical guidance in many occasions during the completion of this project.

My special thanks and sincere gratitude to my very first bench-teacher, Rosemary Grabs, our laboratory's senior technician. This thesis would not have been possible without her helpful insights and comments in multiple aspects throughout my masters.

I would like to show my gratitude to my academic advisors and members of my committee meeting, Dr Nancy Braverman and Dr Indra Gupta.

I would like to thank the Montreal Children's Hospital Research Institute for the financial support throughout the completion of my thesis.

Finally, my very special thank you to my family in Greece, my parents and my brother, and my close friends, Gerasimos Antypas and Irene Liodaki, whose spiritual support during my stay in Montreal was invaluable.

10. References

1. Vafiadis P BS, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus Nat Genet 15: 289-292.

2. Esposito AM, Mateyak M, He D, Lewis M, Sasikumar AN, et al. (2010) Eukaryotic polyribosome profile analysis. J Vis Exp.

3. Hittel D, Storey KB (2002) The translation state of differentially expressed mRNAs in the hibernating 13-lined ground squirrel (Spermophilus tridecemlineatus). Arch Biochem Biophys 401: 244-254.

4. Sheets MD, Fritz B, Hartley RS, Zhang Y (2010) Polyribosome analysis for investigating mRNA translation in Xenopus oocytes, eggs and embryos. Methods 51: 152-156.

5. del Prete MJ, Vernal R, Dolznig H, Mullner EW, Garcia-Sanz JA (2007) Isolation of polysome-bound mRNA from solid tissues amenable for RT-PCR and profiling experiments. RNA 13: 414-421.

6. Hakonarson H, Qu HQ, Bradfield JP, Marchand L, Kim CE, et al. (2008) A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. Diabetes 57: 1143-1146.

7. Consortium. WTCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.

8. Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem 266: 19867-19870.

9. Cazzola M, Skoda RC (2000) Translational pathophysiology: a novel molecular mechanism of human disease. Blood 95: 3280-3288.

10. Kronja I, Orr-Weaver TL (2011) Translational regulation of the cell cycle: when, where, how and why? Philos Trans R Soc Lond B Biol Sci 366: 3638-3652.

11. Lackner DH, Bahler J (2008) Translational control of gene expression from transcripts to transcriptomes. Int Rev Cell Mol Biol 271: 199-251.

12. Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361: 13-37.

13. Sonenberg N, Hinnebusch AG (2007) New modes of translational control in development, behavior, and disease. Mol Cell 28: 721-729.

14. Parsyan A, Svitkin Y, Shahbazian D, Gkogkas C, Lasko P, et al. (2011) mRNA helicases: the tacticians of translational control. Nat Rev Mol Cell Biol 12: 235-245.

15. Richter JD, Sonenberg N (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature 433: 477-480.

16. Sachs AB, Davis RW (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell 58: 857-867.

17. Wells SE, Hillner PE, Vale RD, Sachs AB (1998) Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell 2: 135-140.

18. Namen AE, Lupton S, Hjerrild K, Wignall J, Mochizuki DY, et al. (1988) Stimulation of B-cell progenitors by cloned murine interleukin-7. Nature 333: 571-573.

19. Bamford RN, Battiata AP, Burton JD, Sharma H, Waldmann TA (1996) Interleukin (IL) 15/IL-T production by the adult T-cell leukemia cell line HuT-102 is associated with a human T-cell lymphotrophic virus type I region /IL-15 fusion message that lacks many

upstream AUGs that normally attenuates IL-15 mRNA translation. Proc Natl Acad Sci U S A 93: 2897-2902.
20. Arrick BA, Lee AL, Grendell RL, Derynck R (1991) Inhibition of translation of transforming growth factor-beta 3 mRNA by its 5' untranslated region. Mol Cell Biol 11: 4306-4313.
21. Harigai M, Miyashita T, Hanada M, Reed JC (1996) A cis-acting element in the BCL-2 gene controls expression through translational mechanisms. Oncogene 12: 1369-1374.
22. Ghilardi N, Wiestner A, Skoda RC (1998) Thrombopoietin production is inhibited by a translational mechanism. Blood 92: 4023-4030.
23. Ghilardi N, Skoda RC (1999) A single-base deletion in the thrombopoietin (TPO) gene causes familial essential thrombocythemia through a mechanism of more efficient translation of TPO mRNA. Blood 94: 1480-1482.
24. Rao CD, Pech M, Robbins KC, Aaronson SA (1988) The 5' untranslated sequence of the c-sis/platelet-derived growth factor 2 transcript is a potent translational inhibitor. Mol Cell Biol 8: 284-292.
25. Kim SJ, Park K, Koeller D, Kim KY, Wakefield LM, et al. (1992) Post-transcriptional regulation of the human transforming growth factor-beta 1 gene. J Biol Chem 267: 13702-13707.
26. Grens A, Scheffler IE (1990) The 5'- and 3'-untranslated regions of ornithine decarboxylase mRNA affect the translational efficiency. J Biol Chem 265: 11810-11816.
27. Klausner RD, Rouault TA, Harford JB (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. Cell 72: 19-28.
28. Theil EC (1994) Iron regulatory elements (IREs): a family of mRNA non-coding sequences. Biochem J 304 (Pt 1): 1-11.
29. Allerson CR, Cazzola M, Rouault TA (1999) Clinical severity and thermodynamic effects of iron-responsive element mutations in hereditary hyperferritinemia-cataract syndrome. J Biol Chem 274: 26439-26447.
30. Singh S, Bevan SC, Patil K, Newton DC, Marsden PA (2005) Extensive variation in the 5'-UTR of Dicer mRNAs influences translational efficiency. Biochem Biophys Res Commun 335: 643-650.
31. Sedman SA, Gelembiuk GW, Mertz JE (1990) Translation initiation at a downstream AUG occurs with increased efficiency when the upstream AUG is located very close to the 5' cap. J Virol 64: 453-457.
32. McClelland S, Shrivastava R, Medh JD (2009) Regulation of Translational Efficiency by Disparate 5' UTRs of PPARgamma Splice Variants. PPAR Res 2009: 193413.
33. Jin Q, Agrawal L, Meyer L, Tubiana R, Theodorou I, et al. (2008) CCR5Delta32 59537-G/A promoter polymorphism is associated with low translational efficiency and the loss of CCR5Delta32 protective effects. J Virol 82: 2418-2426.
34. Cho MC, Lee K, Paik SG, Yoon DY (2008) Peroxisome Proliferators-Activated Receptor (PPAR) Modulators and Metabolic Disorders. PPAR Res 2008: 679137.
35. Willson TM, Brown PJ, Sternbach DD, Henke BR (2000) The PPARs: from orphan receptors to drug discovery. J Med Chem 43: 527-550.
36. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, et al. (1996) Identification of a major co-receptor for primary isolates of HIV-1. Nature 381: 661-666.
37. O'Brien SJ, Moore JP (2000) The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. Immunol Rev 177: 99-111.

38. Mummidi S, Ahuja SS, McDaniel BL, Ahuja SK (1997) The human CC chemokine receptor 5 (CCR5) gene. Multiple transcripts with 5'-end heterogeneity, dual promoter usage, and evidence for polymorphisms within the regulatory regions and noncoding exons. J Biol Chem 272: 30662-30671.
39. Pelletier J, Sonenberg N (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature 334: 320-325.
40. Stoneley M, Paulin FE, Le Quesne JP, Chappell SA, Willis AE (1998) C-Myc 5′ untranslated region contains an internal ribosome entry segment. Oncogene 16: 423-428.
41. Paulin FE, Chappell SA, Willis AE (1998) A single nucleotide change in the c-myc internal ribosome entry segment leads to enhanced binding of a group of protein factors. Nucleic Acids Res 26: 3097-3103.
42. Bonnal S, Boutonnet C, Prado-Lourenço L, Vagner S. IRESdb: the Internal Ribosome Entry Site database. Nucleic Acids Res. 2003 Jan 1;31(1):427-8.
43. Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283-292.
44. Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15: 8125-8148.
45. Kozak M (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. J Cell Biol 115: 887-903.
46. Kozak M (1997) Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J 16: 2482-2492.
47. Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, et al. (1994) A cell cycle regulator potentially involved in genesis of many tumor types. Science 264: 436-440.
48. Serrano M, Hannon GJ, Beach D (1993) A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature 366: 704-707.
49. Liu L, Dilworth D, Gao L, Monzon J, Summers A, et al. (1999) Mutation of the CDKN2A 5′ UTR creates an aberrant initiation codon and predisposes to melanoma. Nat Genet 21: 128-132.
50. De Angioletti M, Lacerra G, Sabato V, Carestia C (2004) Beta+45 G --> C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence. Br J Haematol 124: 224-231.
51. Prevot D, Darlix JL, Ohlmann T (2003) Conducting the initiation of protein synthesis: the role of eIF4G. Biol Cell 95: 141-156.
52. de Moor CH, Meijer H, Lissenden S (2005) Mechanisms of translational control by the 3′ UTR in development and differentiation. Semin Cell Dev Biol 16: 49-58.
53. Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. Nat Rev Mol Cell Biol 5: 827-835.
54. Zhang X, Virtanen A, Kleiman FE (2010) To polyadenylate or to deadenylate: that is the question. Cell Cycle 9: 4437-4449.
55. Kuersten S, Goodwin EB (2003) The power of the 3′ UTR: translational control and development. Nat Rev Genet 4: 626-637.
56. Sheets MD, Wickens M (1989) Two phases in the addition of a poly(A) tail. Genes Dev 3: 1401-1412.
57. Mendez R, Richter JD (2001) Translational control by CPEB: a means to the end. Nat Rev Mol Cell Biol 2: 521-529.

58. Cosson B, Gautier-Courteille C, Maniey D, Ait-Ahmed O, Lesimple M, et al. (2006) Oligomerization of EDEN-BP is required for specific mRNA deadenylation and binding. Biol Cell 98: 653-665.
59. Sayed D, Abdellatif M (2011) MicroRNAs in development and disease. Physiol Rev 91: 827-887.
60. Lujambio A, Lowe SW (2012) The microcosmos of cancer. Nature 482: 347-355.
61. Small EM, Olson EN (2011) Pervasive roles of microRNAs in cardiovascular biology. Nature 469: 336-342.
62. Nothnick W (2012) The role of microRNAs in the female reproductive tract. Reproduction.
63. Im HI, Kenny PJ (2012) MicroRNAs in neuronal function and dysfunction. Trends Neurosci.
64. Mazumder B, Li X, Barik S (2010) Translation control: a multifaceted regulator of inflammatory response. J Immunol 184: 3311-3319.
65. Kontoyiannis D, Pasparakis M, Pizarro TT, Cominelli F, Kollias G (1999) Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. Immunity 10: 387-398.
66. Orkin SH, Cheng TC, Antonarakis SE, Kazazian HH, Jr. (1985) Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. EMBO J 4: 453-456.
67. Rund D, Dowling C, Najjar K, Rachmilewitz EA, Kazazian HH, Jr., et al. (1992) Two mutations in the beta-globin polyadenylylation signal reveal extended transcripts and new RNA polyadenylylation sites. Proc Natl Acad Sci U S A 89: 4324-4328.
68. Filipenko ML, Vinichenko NA, Karpova GG, Mertvetsov NP, Amaldi F (1998) Isolation, structural analysis and mapping of the functional gene of human ribosomal protein S26. Gene 211: 287-292.
69. Stahl J, Kobetz ND (1984) Affinity labelling of rat liver ribosomal protein S26 by heptauridylate containing a 5'-terminal alkylating group. Mol Biol Rep 9: 219-222.
70. Malygin AA, Graifer DM, Bulygin KN, Zenkova MA, Yamkovoy VI, et al. (1994) Arrangement of mRNA at the decoding site of human ribosomes. 18S rRNA nucleotides and ribosomal proteins cross-linked to oligouridylate derivatives with alkylating groups at either the 3' or the 5' termini. Eur J Biochem 226: 715-723.
71. Wu M, Tan HM (1994) The Saccharomyces cerevisiae homologue of ribosomal protein S26. Gene 150: 401-402.
72. Ivanov AV, Malygin AA, Karpova GG (2005) Human ribosomal protein S26 suppresses the splicing of its pre-mRNA. Biochim Biophys Acta 1727: 134-140.
73. Brosche M, Strid A (1999) The mRNA-binding ribosomal protein S26 as a molecular marker in plants: molecular cloning, sequencing and differential gene expression during environmental stress. Biochim Biophys Acta 1445: 342-344.
74. Vincent S, Marty L, Fort P (1993) S26 ribosomal protein RNA: an invariant control for gene regulation experiments in eucaryotic cells and tissues. Nucleic Acids Res 21: 1498.
75. Filipenko ML, Vladimirov SN, Muravlev AI, Karpova GG, Mertvetsov NP (1994) [Cloning cDNA of human S26 ribosomal protein and determination of its primary structure]. Bioorg Khim 20: 644-649.

76. Bortoluzzi S, d'Alessi F, Romualdi C, Danieli GA (2001) Differential expression of genes coding for ribosomal proteins in different human tissues. Bioinformatics 17: 1152-1157.
77. Doherty L, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, et al. (2010) Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. Am J Hum Genet 86: 222-228.
78. Strittmatter AW, Fischer C, Kleinschmidt M, Braus GH (2006) FLO11 mediated filamentous growth of the yeast Saccharomyces cerevisiae depends on the expression of the ribosomal RPS26 genes. Mol Genet Genomics 276: 113-125.
79. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, et al. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 39: 857-864.
80. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, et al. (2007) A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 448: 591-594.
81. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107.
82. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. Nat Genet 39: 1202-1207.
83. Polychronakos C, Li Q (2011) Understanding type 1 diabetes through genetics: advances and prospects. Nat Rev Genet 12: 781-792.
84. Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, et al. (2011) Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet 7: e1002367.
85. Garcia C, Feve B, Ferre P, Halimi S, Baizri H, et al. (2010) Diabetes and inflammation: fundamental aspects and clinical implications. Diabetes Metab 36: 327-338.
86. Nakatani Y, Kaneto H, Hatazaki M, Yoshiuchi K, Kawamori D, et al. (2006) Increased stress protein ORP150 autoantibody production in Type 1 diabetic patients. Diabet Med 23: 216-219.
87. Plagnol V, Smyth DJ, Todd JA, Clayton DG (2009) Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. Biostatistics 10: 327-334.
88. Vlachos A, Muir E (2010) How I treat Diamond-Blackfan anemia. Blood 116: 3715-3723.
89. Lipton JM, Ellis SR (2010) Diamond Blackfan anemia 2008-2009: broadening the scope of ribosome biogenesis disorders. Curr Opin Pediatr 22: 12-19.
90. Draptchinskaia N, Gustavsson P, Andersson B, Pettersson M, Willig TN, et al. (1999) The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. Nat Genet 21: 169-175.
91. Cmejla R, Ludikova B, Sukova M, Blatny J, Pospisilova D (2011) Can mutations in the ribosomal protein S26 (RPS26) gene lead to Klippel-Feil syndrome in Diamond-Blackfan anemia patients? An update from the Czech Diamond-Blackfan Anemia registry. Blood Cells Mol Dis 46: 300-301.
92. Tassabehji M, Fang ZM, Hilton EN, McGaughran J, Zhao Z, et al. (2008) Mutations in GDF6 are associated with vertebral segmentation defects in Klippel-Feil syndrome. Hum Mutat 29: 1017-1027.

93. Stuart PE, Nair RP, Ellinghaus E, Ding J, Tejasvi T, et al. (2010) Genome-wide association analysis identifies three psoriasis susceptibility loci. Nat Genet 42: 1000-1004.
94. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, et al. (2009) Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat Genet 41: 199-204.
95. Ellinghaus E, Ellinghaus D, Stuart PE, Nair RP, Debrus S, et al. (2010) Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2. Nat Genet 42: 991-995.
96. Singer-Sam J, LeBon JM, Dai A, Riggs AD (1992) A sensitive, quantitative assay for measurement of allele-specific transcripts differing by a single nucleotide. PCR Methods Appl 1: 160-163.
97. Bolt MW, Mahoney PA (1997) High-efficiency blotting of proteins of diverse sizes following sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Anal Biochem 247: 185-192.
98. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, et al. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30: e57.
99. Sellner LN, Taylor GR (2004) MLPA and MAPH: new techniques for detection of gene deletions. Hum Mutat 23: 413-419.
100. Zhou MY, Gomez-Sanchez CE (2000) Universal TA cloning. Curr Issues Mol Biol 2: 1-7.
101. Jones LJ, Yue ST, Cheung CY, Singer VL (1998) RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization. Anal Biochem 265: 368-374.
102. Pelham HR, Jackson RJ (1976) An efficient mRNA-dependent translation system from reticulocyte lysates. Eur J Biochem 67: 247-256.
103. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90: 7-24.
104. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747-753.
105. Burke DF, Worth CL, Priego EM, Cheng T, Smink LJ, et al. (2007) Genome bioinformatic analysis of nonsynonymous SNPs. BMC Bioinformatics 8: 301.
106. Kiekens R, Vercauteren A, Moerkerke B, Goetghebeur E, Van Den Daele H, et al. (2006) Genome-wide screening for cis-regulatory variation using a classical diallel crossing scheme. Nucleic Acids Res 34: 3677-3686.
107. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, et al. (2005) Mapping determinants of human gene expression by regional and genome-wide association. Nature 437: 1365-1369.
108. Boehnke M (2000) A look at linkage disequilibrium. Nat Genet 25: 246-247.