NOTE TO USERS

This reproduction is the best copy available.

UMI®

High throughput study of the translational effect of human single nucleotide polymorphisms

By

Yang Lu

Department of Experimental Medicine

McGill University, Montreal

August, 2008

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

©Copyright by Yang Lu (2008)

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-67031-6
Our file Notre référence
ISBN: 978-0-494-67031-6

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Table of Contents

1. Abbreviations.....1

2. Abstracts.....3

2.1 English Version.....3

2.2 French Version (Résumé)... 4

3. Introduction and Literature Review.....5

3.1 The association between genetic variations and complex traits.....5

3.2 The Gene Regulators In Disease (GRID) project.....6

3.3 Translational control.....8

3.4 Genetic variations with translational effects.....14

3.5 Objective of this project.....18

4. Methods.....20

4.1 Subjects.....20

4.2 Polysome fractionation.....21

4.3 RNA extraction and assessment.....22

4.4 Precise quantification of mRNA in the fractions.....23

4.5 Translational imbalance (TI) assay.....24

4.5.1 Resequencing.....24

4.5.2 Microarray.....26

4.6 Data analysis.....38

5. Results.....40

5.1 Polysome fractionation 40

5.2 The Agilent assay of fraction RNA samples.....41

5.3 Expression profiles of six genes in the fractions.....43

5.4 Oligo-dT assay to quantify fraction poly(A) RNA.....45

5.5 The fractionated RNA sample bank.....47

5.6 TI assay on the *Illumina* beadarray.....49

5.7 Confirmation of the *Illumina* results by resequencing.....53

6. Discussion.....60

6.1 A novel approach to cover an aspect of functional genomics.....60

6.2 The exclusion of non-causative SNPs.....61

6.3 The analysis of the TI mechanism 67

7. Conclusion.....71

8. Acknowledgement.....73

9. References.....74

Abbreviations

AI, allelic imbalance;

ASOs, allele-specific oligos;

cDNA, complementary DNA;

CEPH, Centre de l'Étude du Polymorphisme Humain;

CPE, cytoplasmic polyadenylation element;

CPSF, cleavage and polyadenylation specificity factor;

DIG, digoxigenin;

DNA, deoxyribonucleic acid;

DTT, dithiothreitol;

EDEN, embryonic deadenylation element;

eIFs, eukaryotic initiation factors;

FBS, Fetal Bovine Serum;

FRP, ferritin repressor protein;

GD, Graves' disease;

gDNA, genomic DNA;

GRID, Gene Regulators in Disease project;

HBB, the hemoglobin, beta gene;

IFIH1, the interferon induced with helicase C domain 1 gene;

IL10RA, the interleukin 10 receptor, alpha gene;

IRE, iron-responsive element;

IREBP, IRE-binding protein;

IRES, internal ribosome entry site;

IRF, iron regulatory factor;

IRPs, iron-regulatory proteins;

ITGB1BP1, integrin beta 1 binding protein 1;

LCL, lymphoblastoid cell lines;

LD, linkage disequilibrium;

LSOs, locus-specific oligos;

LTB4R, the including leukotriene B4 receptor gene;

MFE, minimum free energy;

miRNA, microRNA;

NAHR, non-allelic homologous recombination;

NHEJ, non-homologous end-joining;

nsSNP, nonsynonymous SNP;

OD, optical density;

PABP, poly(A)-binding protein;

PBS, phosphate buffered saline;

RNA, ribonucleic acid;

SNP, Single-nucleotide polymorphism;

sSNP, synonymous SNP;

T1D, type 1 diabetes;

TAP2, the transporter 2, ATP-binding cassette, sub-family B;

TI, translational imbalance;

uAUGs, upstream AUGs;

UCPAS, upstream core polyadenylation signal sequence;

uORFs, upstream open-reading frames;

VNTR, rariable number of tandem repeats;

WBCs, white blood cells;

X-DC, X-linked dyskeratosis congenital.

1.1 Abstract

Introduction: As a part of the Gene Regulators in Disease project (GRID), this study aims to create a novel high throughput method to discover the genetic effect on gene translation, taking advantage of the rationale that efficiently translated mRNAs associate with multiple ribosomes, while less active ones with fewer or none.

Methods: Lymphoblastoid cell lines (LCLs) from 44 HapMap European individuals were used for polyribosomal fractionation and establishing the sample bank for the future study. The fractionated mRNA samples of 10 out of the 44 individuals were run on an *Illumina* GoldenGate Beadarray to detect allelic imbalance (developed by the group of T.J. Hudson and T.M. Pastinen).

Results: This study established a high-quality RNA bank, including 1,100 RNA fraction samples. By the *Illumina* chip, translational imbalance was detected in 75 out of 1483 (5.06%) assays, and 63 out of 269 (23.4%) genes. The translational effect was well replicable by the resequencing method.

Conclusion: This study found that genetic effect on gene translation is a common mechanism of expression regulation. Our best hit found in the integrin beta 1 binding protein 1 gene (*ITGB1BP1*) highlights the role of mRNA 3'UTR secondary structure in gene translation.

Keywords: Gene translation, High throughput genotyping, Human genetics, Polyribosome, RNA, Single nucleotide polymorphism

1.2 Résumé

Introduction: Cette étude qui s'inscrit dans le cadre du projet sur les régulateurs des gènes dans les maladies (GRID) vise à développer une nouvelle méthode de détection à haut débit des effets génétiques sur la traduction des gènes en prenant avantage du fait que lorsque traduits efficacement, les ARNm s'associent à de multiples ribosomes alors que ces associations sont plus rares ou inexistantes lorsque les ARNm sont moins actifs.

Méthodes: Des lignées cellulaires lymphoblastoïdes (LCL) dérivées de 44 individus Européens caractérisés dans HapMap on été utilisées pour le fractionnement des ribosomes. Les échantillons d'ARNm fractionné de dix individus ont été analysés sur la plateforme matricielle *GoldenGate* Beadarray d'Illumina pour la détection de déséquilibres alléliques (développé par le groupe de T.J. Hudson et T.M. Pastinen).

Résultats: Cette étude a permit d'établir une banque d'ARN de grande qualité, incluant 1100 échantillons d'ARN fractionné. Grâce aux puces d'Illumina, un déséquilibre de traduction a été détecté dans 75 tests sur 1483 (5,06%) et 63 gènes sur 269 (23,4%). Ces effets sur la traduction ont été confirmés par la méthode de séquençage.

Conclusion: Cette étude a démontré que les effets génétiques sur la traduction sont des mécanismes fréquents dans la régulation de l'expression génique. Notre meilleur candidat, trouvé dans le gène ITGB1BP1, met en lumière le rôle d'une structure secondaire de la région non-traduite en 3' de l'ARNm dans la traduction génique.

3. Introduction and Literature Review

3.1 The association between genetic variations and complex traits

Deoxyribonucleic acid (DNA) is a long chain of nucleotides, and carries the genetic information for unicellular organisms, plants, and animals. There are four types of nucleotides, and each contains one of the four base molecules, i.e. Adenine (A), Cytosine (C), Guanine (G), or Thymine (T). In the human nuclear genome, DNA sequences exist in the form of a double helix DNA structure. Two complementary DNA sequences are connected by hydrogen bonds through A-T pairing or C-G pairing. About 3 billion DNA base pairs (Human Genome Sequencing 2004) are found on 22 pairs of autochromosomes and two sex chromosomes (X and Y) in the human. Human genetic variation is the difference between genomes of individuals. The most common type of DNA variation is single nucleotide polymorphisms (SNPs) due to a change in which one nucleotide is replaced by another in the DNA sequence. SNPs with frequencies >1% occur every 100 to 300 bases along the human genome. Besides SNPs, copy number variations (i.e. deletions, inversions, insertions and duplications) from non-allelic homologous recombination (NAHR), non-homologous end-joining (NHEJ), variable number of tandem repeats (VNTR), and retro-transposition, are attracting more and more attention in recent years (Kidd, Cooper et al. 2008).

There is increasing evidence that genetic variations play an important role in the determination of individual susceptibility to complex disease traits (Couzin and

Kaiser 2007; Wellcome Trust Case Control Consortium 2007). Allele-specific effects on gene expression appear relatively common, as determined by several recent studies (Ge, Gurd et al. 2005; Pastinen, Ge et al. 2005; Dixon, Liang et al. 2007; Zhang, Duan et al. 2008). The role of regulatory polymorphisms in determining susceptibility to a number of complex disease traits is supported by multiple reports of robustly replicated associations of phenotypes with DNA variation that does not involve protein sequence change. Direct observation of such effects include, for example, studies in my supervisor's laboratory on type 1 diabetes (T1D) regarding the role of a variation at the VNTR of the *INS* gene (Vafiadis, Bennett et al. 1997; Barratt, Payne et al. 2004), and a regulatory polymorphism of *CTLA4*, the gene encoding cytotoxic T lymphocyte antigen (Ueda, Howson et al. 2003; Anjos, Tessier et al. 2004; Anjos, Shao et al. 2005). Therefore, identification of human sequence polymorphisms that regulate gene expression is a key to understanding human genetic diseases (Hudson 2003). SNPs on transcribed sequences have been used as markers to discriminate allele-specific ribonucleic acid (RNA) levels as a powerful approach in screening for regulatory polymorphisms (Ge, Gurd et al. 2005; Pastinen, Ge et al. 2005).

3.2 The Gene Regulators In Disease (GRID) project

The Gene Regulators in Disease project (GRID, http://www.regulatorygenomics.org/index.html) is funded by Genome Canada and Genome Quebec since April 1, 2006. Its objective is the systematic screening of human genetic variations that affect gene expression and the identification of the

cascade of biological steps that modify the genetic risk of common complex diseases. The discoveries of this project will form a solid basis for the development of new diagnostic and therapeutic targets.

SNPs can change a gene's RNA levels through different mechanisms: (1) For a SNP at the gene promoter region, it may abolish the promoter function, thus impair the gene transcription (Menzaghi, Paroni et al. 2006); (2) For a SNP at the gene transcription region, it may change the mRNA secondary structure and mRNA stability (Wang, Johnson et al. 2005); (3) For a nonsense SNP at the gene coding region producing a premature stop codon, a specific mRNA degradation mechanism known as nonsense-mediated decay can result (Richards, Laidlaw et al. 2007). The RNA allelic imbalance (AI, i.e. two alleles have different RNA levels) method developed by Pastinen et al. (Ge, Gurd et al. 2005; Pastinen, Ge et al. 2005) has been successfully used to identify genetic variations that change the gene RNA level.

My supervisor's laboratory has recently demonstrated computational evidence of widespread effects of synonymous polymorphisms in the third codon position (Qu, Lawrence et al. 2006). They found excess of C-->T over G-->A polymorphisms at the fourfold degenerate sites. This finding supports for the importance of polymorphisms affecting RNA function post-transcriptionally. SNPs may change the levels of different mRNA isoforms, resulting from genetic effects on alternative promoters, exon skipping, alternative splicing, intron

retention, or alternative polyadenylation. J. Majewski has used the exon tiling microarray method to analyze the mRNA levels of different isoforms successfully (Kwan, Benovoy et al. 2007; Kwan, Benovoy et al. 2008). Effect of RNA polymorphism on translation, by contrast, has received little attention to date.

3.3 Translational control

The gene translation follows three stages, i.e. initiation, elongation, and termination (Fig.1). (1) Translation initiation is the process of forming 80S initiation complex. It commences with the 40S ribosomal subunit carrying methionine-tRNA binding to the mRNA 5' end (the m^7G cap), with the help of eukaryotic initiation factors (eIFs) and the poly(A)-binding protein (PABP) (Wells, Hillner et al. 1998). Binding of both the 5' and the 3' end results in RNA circularization. Next, the 40S subunit scans the mRNA from the 5' to the 3' direction until it finds the AUG codon in the translation start site. Right then, the 60S subunit binds to the 40S subunit, and forms the 80S initiation complex (Pestova and Hellen 2006). (2) The 80S ribosome has 3 functional binding sites, i.e. the aminoacyl (A) site, the peptidyl (P) site, and the exit (E) site. After the initiation step, the 80S ribosome begins the repetitive cycles of peptide elongation. In each cycle, an aminoacyl-tRNA enters the A site; a peptide bond forms between the peptidyl-tRNA at the P site and the new amino acid residue, and the tRNA at the P site moves to the E site and then is released; the ribosome moves to the next mRNA codon, and the peptidyl-tRNA moves from the A site to the P site (Spahn, Gomez-Lorenzo et al. 2004). (3) At the termination stage, one of the three

stop codons, UAA, UGA, or UAG, enters the A site, and the complete peptide is released (Merrick 1992). Translational control is complicated and can be regulated at multiple levels during initiation, elongation, and termination, and even after termination. Most well-studied mechanisms of translational control occur in the translation initiation stage (Sonenberg and Hinnebusch 2007).

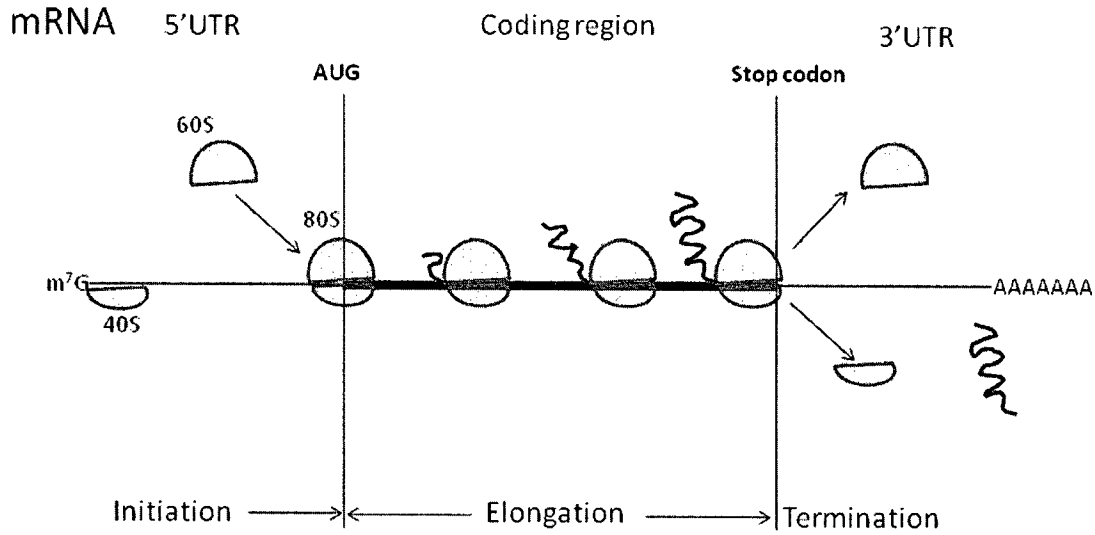

Fig. 1. The three stages of gene translation, i.e. initiation, elongation, and termination: (1) Translation initiation forms the 80S initiation complex; (2) The 80S ribosome is responsible for peptide elongation; (3) At the stop codon, the 80S ribosome releases the peptide, and splits to 60S subunit and 40S subunit.

3.3.1 Recognition of the m^7G cap

The initiation of translation requires the help of numerous eIFs. The eIF4E recognizes the m^7G cap (Pestova and Kolupaeva 2002; Sonenberg 2008). The eIF4G interacts with eIF4E and PABP as the scaffold protein to circularize the

mRNA by bringing the 3' UTR in close proximity to the 5' end of the mRNA, and to facilitate the binding of the 40S ribosome subunit (Wells, Hillner et al. 1998; Gebauer and Hentze 2004).

3.3.2 The 5'UTR in translational control

After binding with the m^7G, the 40S subunit scans the mRNA to find the translation start site. Multiple *cis*-acting factors in the 5'UTR region can influence the 40S subunit scanning. Upstream AUGs (uAUGs) and upstream open-reading frames (uORFs) in the mRNA 5'UTR can enable the formation of 80S ribosome, thus reducing the number of 40S subunit reaching the authentic or main AUG start codons (Wang and Rothnagel 2004). The inhibitory effect of translation by blocking the scanning of the 40S subunit to the downstream coding region depends on the distance between uORF and the downstream coding region, i.e. a shorter distance has stronger inhibitory effect (Vattem and Wek 2004). Non-specifically, the mRNA 5'UTR secondary structure inhibits the efficiency of the 40S scanning for the translation start site (Lee, Guertin et al. 1983). Specifically, some particular 5'UTR secondary structures can regulate translation by RNA-protein interaction. One example is the iron-responsive element (IRE), a specific translational inhibition element found in the 5'UTR of the ferritin mRNA (Leibold and Guo 1992). IREs have hairpin structure, and can be recognized by iron-regulatory proteins (IRPs), such as IRE-binding protein (IREBP), iron regulatory factor (IRF), or ferritin repressor protein (FRP) (Hentze and Kuhn 1996). The RNA-protein interaction of a m^7G cap-proximal IRE and IRP control

the mRNA translation efficiency by preventing the 40S subunit from binding to the m^7G cap (Muckenthaler, Gray et al. 1998). Also, the trans-acting protein IRPs, e.g. IRF-2, may decrease mRNA stability (Hentze and Kuhn 1996).

The internal ribosome entry site (IRES) is a mRNA 5'UTR sequence found in viral RNAs, which can regulate translation by initiating translation independent of m^7G (Pelletier and Sonenberg 1988). In eukaryotes, the empirical evidence for the existence of IRES is still controversial (Kozak 2003). It is worth noting that, because of the important roles of 5'UTR in translational control, mRNA isoforms with different 5' UTRs, as the result of alternative promoter, exon skipping, alternative splicing, or intron retention, may have different translational efficiency.

3.3.3 The Kozak sequence

Once the 40S subunit finds the AUG start codon, the 80S ribosome forms and peptide synthesis starts. The translation site around the start codon is also an important factor of translation efficiency. A common sequence of the translation start site is gccRCCAUGG, known as Kozak sequence (Kozak 1991). The A nucleotide of the AUG start codon is referred 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 AUG is essential. The G at position +4 (G^{+4}) and R^{-3} are the most conserved. In vertebrate mRNAs, 46% have G^{+4}, and 97% have R^{-3} (Kozak 1991). The replacement of G^{+4} (Kozak 1997) or R^{-3} (Kozak 1986) with another nucleotide has dramatic impact on translation efficiency. The other positions are less

conserved. The strength of the Kozak consensus sequences ranges from 'strong' to 'weak' and in descending order are $A^{-3}+G^{+4} > G^{-3}+G^{+4} > A^{-3}+A^{+4} > G^{-3}+A^{+4} > U^{-3}+G^{+4} > U^{-3}+A^{+4}$ (Wang and Rothnagel 2004).

3.3.4 The coding sequence in translational control

Little is known about the role of the coding sequence in translational control. A recent study from my supervisor's group observed asymmetrical distribution of synonymous SNPs (sSNPs) on the two DNA strands across the human genome. An excess of C-->T over G-->A polymorphisms was found in non-CpG site fourfold degenerate (FFD) site SNPs but was absent from iSNPs and CpG site FFD SNPs. This finding suggests widespread selective pressure due to functional effects of sSNPs (Qu, Lawrence et al. 2006). An empirical observation suggests that the coding sequence may regulate gene translation through the change of the mRNA secondary structure (Nackley, Shabalina et al. 2006).

3.3.5 The 3'UTR in translational control

Although translation begins at the 5' end of the mRNA, the 3'UTR plays an important regulatory role that governs the spatial and temporal gene expression (Kuersten and Goodwin 2003; Gebauer and Hentze 2004). The translational control of the 3'UTR is facilitated through mRNA circularization by the interaction of eIF4G and PABP, which brings the 3' UTR in close proximity to the translation initiation site (Prévôt, Darlix et al. 2003). There is a dramatic correlation between the length of poly(A) tail and translational efficiency (the rate

and the amount of protein production), i.e. longer poly(A) tail corresponds to higher and shorter poly(A) tail to lower translational efficiency (de Moor, Meijer et al. 2005). The 3' UTR contains the functional elements that regulate the length of the poly(A) tail, e.g. the cytoplasmic polyadenylation element (CPE), and the embryonic deadenylation element (EDEN). The CPE consists of UUUUUAU or other similar sequences, and binds with CPE binding protein to promote the extension of the poly(A) tail (Simon, Tassan et al. 1992). The core EDEN sequence consists of a repetition of U(A/G) dinucleotides, and binds to the EDEN-specific RNA-binding protein (EDEN-BP) to mediate embryonic deadenylation of mRNAs (Paillard, Omilli et al. 1998).

The upstream core polyadenylation signal sequence (UCPAS) can modulate the efficiency of gene translation by regulating the formation of the poly-A tail at the mRNA 3'UTR. UCPAS is located at the 3' UTR on the 5' side in proximity to the pre-mRNA cleavage site, and has the consensus hexamer sequence AAUAAA (Zhao, Hyman et al. 1999). By binding with cleavage and polyadenylation specificity factor (CPSF), UCPAS controls the cleavage of 3' flanking region of pre-mRNA and the synthesis of Poly(A) tail (Takagaki and Manley 1997). In addition, the 3' UTR contains other *cis*-acting elements that control mRNA stability and localization (Chen, Ferec et al. 2006), and that regulate translation initiation and elongation (Kuersten and Goodwin 2003; de Moor, Meijer et al. 2005). As a research highlight in recent years, microRNAs

(miRNAs) bind to the 3' UTR of target mRNAs, and promote mRNA degradation or repress translation initiation (Du and Zamore 2007).

3.4 Genetic variations with translational effects

It is obvious that genetic variation in any of the sequence elements enumerated above could affect translational efficiency. To date, most known genetic variations with translational effects are rare mutations causing Mendelian diseases (Scheper, van der Knaap et al. 2007), which will be introduced in details in the subsequent sections. The role of common DNA variations on gene translation in common complex diseases has received surprisingly little attention, likely because of the lack of appropriate methodology for high-throughput screening at the nucleotide (rather than at the protein) level.

3.4.1 5' UTR variants with known functional effects

The uORF is an important element of cis translational regulation. A 516G→T mutation at the 5'UTR of the thrombopoietin gene (*THPO*) mRNA truncates a uORF with potent inhibition on the mRNA translation. By the disruption of the inhibition effect, this gain-of-function mutation increase the production of thrombopoietin dramatically. Consequently, the over-expression of *THPO* increases platelet production, and causes hereditary thrombocythemia (Kikuchi, Tayama et al. 1995; Ghilardi, Wiestner et al. 1999). On the other hand, a loss-of-function mutation producing an upstream AUG or uORF can inhibit the mRNA translation. The case can be seen in the finding of the G→T mutation at the

5'UTR of the cyclin-dependent kinase inhibitor 2A gene (*CDKN2A*) mRNA, which produces a strong matched Kozak consensus sequence 34nt upstream of the wild type translation start site (Liu, Dilworth et al. 1999). The wild type sequence is 5'-gcgGAGAGGG-3', and the G→T mutation produces 5'-gcgGAGATGG-3', which strongly matches the Kozak consensus sequence 5'-gccRCCATGG-3'. The *CDKN2A* gene encodes a low-molecular-weight protein p16, which is a CDK4 kinase inhibitor (Kamb, Gruis et al. 1994). As a tumor suppressor, p16 can inhibit cell growth by arresting cells at G_1 phase (Serrano, Hannon et al. 1993). The impaired production of the wild type p16 protein may increase the susceptibility to melanoma (Liu, Dilworth et al. 1999).

The IRE is a 30nt motif at the 5' UTR of the ferritin mRNA, and the 3' UTR of transferrin receptor mRNAs. The consensus structure of IRE is a stem loop (Hentze and Kuhn 1996). There is a bulge in the middle of the stem. The 6nt loop at the end of the stem is highly conserved with the consensus sequence of 5'-CAGUGX-3', where X can be any nucleotide except G (Jaffrey, Haile et al. 1993; McCallum and Pardi 2003). Mutations change the stem loop structure of IRE of the ferritin mRNA, and disrupt the interaction of IRE-IRF, which consequently impairs the control of the ferritin translation and causes anomalous expression of ferritin, a genetic syndrome known as hereditary hyperferritinemia-cataract syndrome (Allerson, Cazzola et al. 1999).

Mutations may impair the IRES-dependent translation, e.g. the finding of the mutations of the dyskeratosis congenita 1, dyskerin gene (*DKC1*), which cause X-linked dyskeratosis congenita (X-DC)(Yoon, Peng et al. 2006). On the other hand, a single nucleotide change may also increase the function of an IRES element, as the finding in the v-myc myelocytomatosis viral oncogene homolog (avian) gene (*MYC*) (Paulin, Chappell et al. 1998). However, the consensus sequence of IRES is unclear.

Allele-specific alternative splicing can lead to the production of mRNA isoforms with different 5'UTR. Because of the importance of the functional elements in the mRNA 5'UTR, these isoforms differing only in their 5'UTR may have different translational efficiency. A typical case can be seen in the *THPO* gene. The G→C transversion at position +1 of intron 3 of the *THPO* gene leads to the production of an mRNA isoform with a shortened 5'UTR. The translational inhibitory elements are missing in the isoform with shortened 5'UTR. This gain-of-function mutation by splicing effect is also a cause of hereditary thrombocythemia (Wiestner, Schlemper et al. 1998).

3.4.2 The Kozak consensus sequence

Genetic variations in the Kozak sequence have also been reported in the study of human genetic diseases. A (G→C) mutation, found in a thalassaemia intermedia family, maps to position -6 of the Kozak sequence of the β-hemoglobin gene, i.e. 5'-(G→C)ACACCATGG-3' (De Angioletti, Lacerra et al. 2004). Although the

position -6 of the Kozak sequence is not highly conserved, De Angioletti et al. found that the mutation decreased the translational efficiency of the β-globin chain by about 30%. By this slight effect on gene function, the mutation alone is not disease-causing, but can worsen the clinical phenotype of thalassaemia intermedia when it coexists with a thalassaemia mutation (De Angioletti, Lacerra et al. 2004).

A common C/T SNP (NCBI dbSNP ID rs1883832, average heterozygosity=0.344) maps to the position -1 of the Kozak sequence of the CD40 molecule, TNF receptor superfamily member 5 gene (*CD40*) mRNA. It is a strong matched Kozak sequence, i.e. 5'-GC(C/T)ATGG-3', whereas the SNP is a nonconservative change in the Kozak sequence. Tomer et al. reported that this SNP was associated with Graves' disease (GD) (Tomer, Concepcion et al. 2002). However, the association has not been replicated by later studies (Heward, Simmonds et al. 2004; Hsiao, Tien et al. 2008) and remains controvensial.

3.4.3 The coding region

Genetic variations in the coding region may change the translational efficiency by the effect on mRNA secondary structure. An example is the finding in the *COMT* gene. *COMT* encodes the enzyme Catechol-O-Methyltransferase, which modulates pain sensitivity (Mannisto and Kaakkola 1999; Diatchenko, Slade et al. 2005). The haplotypes of three synonymous SNPs, rs4633, rs4818, and rs4680, have different mRNA secondary structures. The C-C-G haplotype has a local

stem-loop mRNA secondary structure which is more stable than the other haplotypes. This stable stem-loop decreases the efficiency of protein expression, and corresponds to the clinical phenotype of high pain sensitivity (Nackley, Shabalina et al. 2006).

3.4.4 The 3'UTR

Genetic variations in the UCPAS consensus sequence, the hexamer AAUAAA, may lead to alternative splicing of the pre-mRNA 3' end, and produce alternative mRNA isoforms with different 3' UTR structure and poly(A) tail, therefore different translational properties. By impairing the normal production of the major transcript and the corresponding peptide isoform, variations of the hexamer AAUAAA may cause genetic diseases (Chen, Ferec et al. 2006). The hemoglobin, beta gene (*HBB*) encodes β-globin. Absence of β-globin by mutations in the *HBB* gene causes β^0-thalassemia (Pirastu, Galanello et al. 1987). Reduced amounts of β-globin by mutations in the *HBB* gene causes β^+-thalassemia. A number of studies have found that mutations in the UCPAS consensus sequence can impair the translation of *HBB*, and cause β^+-thalassemia (Orkin, Cheng et al. 1985; Rund, Dowling et al. 1992).

3.5 Objective of this project

The studies discussed above highlight the potential translational effect of common DNA polymorphisms in the human genome. Despite the importance of this potential mechanism of genetic translational control, up to now, there is no high

throughput method to investigate the SNPs' effect on gene translation. As part of the GRID project, the objective of our project is the systematic screening of human genetic variations that affect gene expression through effects on translational efficiency using a unique approach. Our hypothesis is a substantial portion of human phenotypic variation may be due to the effects of common polymorphisms on RNA translational efficiency. The method is developed on the basis of the high throughput AI detection developed by Pastinen et al. (Ge, Gurd et al. 2005; Pastinen, Ge et al. 2005). To search for exonic polymorphisms that alter translational efficiency, I took advantage of the fact that efficiently translated RNAs associate with multiple ribosomes, while less active ones with fewer or none. Clusters of ribosomes (polysomes) can be separated from single ribosomes and soluble RNA by ultracentrifugal fractionation according to size. In the presence of translational imbalance (TI, i.e. two alleles have different translational efficiency), the less active allele is associated with fewer ribosomes, thus is lighter in weight and has less sedimentation rate. By ultracentrifuge, the less translated allele will be found in higher abundance in the upper fractions, with reversal of the ratio towards the lower ones.

4. Methods

4.1 Subjects

Human immortalized lymphoblastoid cell lines (LCL) were acquired from the Centre de l'Étude du Polymorphisme Humain (CEPH) CEU. The CEPH CEU samples have been used for the international HapMap project. These LCLs have publicly available genotypes for ~ four millions SNPs up to March 2008 (http://www.hapmap.org/). The CEPH CEU samples are U.S. Utah residents with ancestry from northern and western Europe. All together, 90 individuals from 30 parent-child trios were provided by the Coriell Institute for Medical Research (http://ccr.coriell.org/Sections/Collections/NHGRI/hapmap.aspx?PgId=266&coll=HG, New Jersey, USA). The cells were cultured using RPMI-1640 containing 15% FBS, 1% L-glutamine, 1% penicillin and 1% streptomycin, in 37°C 5%CO_2 incubator. The Research Ethics Board of the Montreal Children's Hospital and other participating centers approved the study.

Our study needs to investigate AI or TI on hetezygous SNPs. For a rare SNP, there will be very few or no hetezygous sample. To supplement this limit, our lab has established a LCL bank, consisting of more than 300 individuals from T1D families. These individuals have been genotyped with 550,000 SNPs genome-widely using the *Illumina* Infinium™ II HumanHap550 BeadChip technology (*Illumina*, San Diego CA)(Gunderson, Steemers et al. 2005; Steemers, Chang et al. 2006). I have used these LCLs to map genetic effect of gene expression

(Hakonarson, Grant et al. 2007; Qu, Lu et al. 2007; Qu, Marchand et al. 2007). The immortalization of LCLs from fresh peripheral blood lymphocytes follows the standard EBV transformation protocol with minor revisions (Pattengale, Smith et al. 1973).

4.2 Polysome fractionation

RNA molecules were fractionated on the basis of the number of ribosomes they were associated with, by sedimentation velocity in viscous media (10%~50% sucrose gradients). The larger aggregates of the more actively translated RNA move faster than smaller aggregates, while RNA not associated with ribosomes remains at the top. From each LCL, $1x10^8$ cells were incubated with 100ug/μl cycloheximide (to stop translation and fix the ribosomes on the mRNAs) for 5 mins at 37℃, and then were washed with phosphate buffered saline (PBS) containing 100μg/ml cycloheximide. Then, cells were collected by centrifugation at 1,000 rpm for 10 min. Cell pellets were lysed with hypotonic polysome lysis buffer [5 mmol/L Tris-HCL (pH 7.5), 2.5 mmol/L MgCl2, 1.5 mmol/L KCL, 100μg/mL cycloheximide , 2 mmol/L dithiothreitol (DTT), 10% Triton X-100, and 10% sodium deoxycholate], then transferred to a prechilled Eppendorf tube. After 2 min of incubation on ice with occasional vortexing, the extracts were centrifuged for 2 mins at 13,000 rpm at 4℃ to remove cellular debris. The supernatant was directly loaded on a 10%–50% linear sucrose gradient [20 mmol/L HEPES-KOH (PH 7.6), 100 mmol/L KCl, 5 mmol/L $MgCl_2$] and then centrifuged in a Beckman SW41 rotor for 120 minutes at 35,000 rpm at 4℃. As

stated earlier, mRNA directing enhanced protein synthesis is expected to be more abundant in the larger polysome fractions, whereas mRNA subjected to repressed translation should be more abundant in fractions spanning low-number polysomes, monosomes, and free ribosome subunits. Polyribosomal fractions were collected using the Brandel fraction collector (Gaithersburg, Maryland) with real-time monitoring of UV optical density at 254 nm using an Isco type 11 optical unit connected with a recorder. There were 24 fractions/cell line. This project obtained technical support and protocols from Dr. Nahum Sonenberg's laboratory (Kahvejian, Svitkin et al. 2005), and advice from Dr. Francis Robert.

4.3 RNA extraction and assessment

After the fractions were collected, RNA from each fraction was extracted using standard Trizol (Invitrogen, California) followed by the phenol–chloroform method. RNA was dissolved in 15µl of DEPC H_2O. The total RNA concentration of each fraction was quantified by optical density (OD), and the relative abundance of ribosomal RNA subunits assayed on the Agilent 2100 bioanalyzer with the RNA 6000 Nano LabChip® kit (Agilent Technologies, Santa Clara CA). RT-PCR amplification of multiple genes known to be expressed in LCLs had been done to assess the mRNA distribution in the fractions, including *IFIH1* (interferon induced with helicase C domain 1), *LTB4R* (leukotriene B4 receptor), *IL10RA* (interleukin 10 receptor, alpha), *CD40*, *TAP2* (transporter 2, ATP-binding cassette, sub-family B), and the Cyclophilin housekeeping gene. Except Cyclophilin, the other genes have important immune functions and may be

involved in the pathogenesis of type 1 diabetes. Genetic effects on these genes' expression are the interest of my supervisor's group.

4.4 Precise quantification of mRNA in the fractions

Because of the fractionation procedure, abundance of total mRNA is no longer reflected by the OD measurement, the bulk of which is due to ribosomal RNA. Therefore, the mRNA in each fraction needs to be precisely quantified. To quantify the poly(A) RNA in the polysome fractions, fraction aliquots were slot-blotted on a nylon membrane and quantified using an oligo(dT) probe labeled wtih digoxigenin, using the DIG Oligonucleotide 3'-End Labeling Kit (Roche, Indiana). This assay involves the following 5 steps:

a) Labeling the 3' end of the nucleotide probe with DIG;
b) Immobilization of target fraction mRNA on a nylon membrane;
c) Prehybridization and hybridization DIG-labeled probe onto the membrane;
d) Binding of anti-DIG antibody conjugates ;
e) The hybridization signal was detected by an enzyme-linked immunoassay with anti-DIG-AP antibody conjugate, and a chemiluminescence reaction;
f) The final concentration of Poly(A) RNA 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.

4.5 Translational imbalance (TI) assay

Two methods were used to measure the relative proportions of each allele of mRNA SNPs: the resequencing method and the *Illumina* Golden Gate beadarray assay.

4.5.1 Resequencing

Fractionated mRNA samples can be assayed for AI at any transcribed SNP, by RT-PCR and sequencing. However, this is too slow a method for a large-scale genome screen. Therefore, the re-sequencing method was mainly used to study candidate genes and to confirm the beadarray results. To determine the allele levels, heterozygous CEU individuals were selected based on the genotyping data of HapMap. Allelic expression was analyzed by sequencing in both directions. Sequencing was done in parallel for amplified genomic DNA (gDNA) and complementary DNA (cDNA) samples using the 3700 DNA Analyzer (Applied Biosystems, Foster City, CA). To design the PCR/sequencing primers, these steps were followed: (1) Obtain the DNA or RNA reference sequence from the NCBI database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore); (2) Using the SNPmasker 1.0 program (http://bioinfo.ebc.ee/snpmasker/) to mask all SNPs and repeat regions in the given sequences, and to keep from designing primers from these regions (Andreson, Puurand et al. 2006); (3) PCR/sequencing primers for the gDNA and the cDNA were designed by the Primer3 program (http://primer3.sourceforge.net) (Rozen and Skaletsky 2000); (4) Secondary structure analysis of the oligos using the Gene Runner program

(http://www.generunner.net/, Hastings Software). Repeat Step 2-4 until appropriate primers are obtained. The cDNA synthesis was carried out using Superscript II RNase H Reverse Transcriptase (Invitrogen, California). The PCR amplification was done in a 25μl reaction system using the Taq DNA polymerase kit (Invitrogen, California) on the Gene Amp PCR system 9700 (Applied Biosystems).

Two approaches were used to quantify the allelic ratio: (1) To assess the allelic ratio at the mRNA levels, we used allelic proportions in heterozygous gDNA RT-PCR product to establish 1:1 stoichiometry and to correct for relative expression levels of the two alleles in the unfractionated mRNA samples; (2) To assess the allelic ratio in the polysome fractions, allelic proportions in heterozygous unfractionated mRNA RT-PCR product were used to establish 1:1 stoichiometry and to correct for relative expression levels of the two alleles in the various fractions of the mRNA samples. The software PeakPicker developed by Bing Ge and Tomi Pastinen can quantify the allele ratio by correcting for adjacent peaks (Ge, Gurd et al. 2005). The AI pattern was depicted in Fig.2.

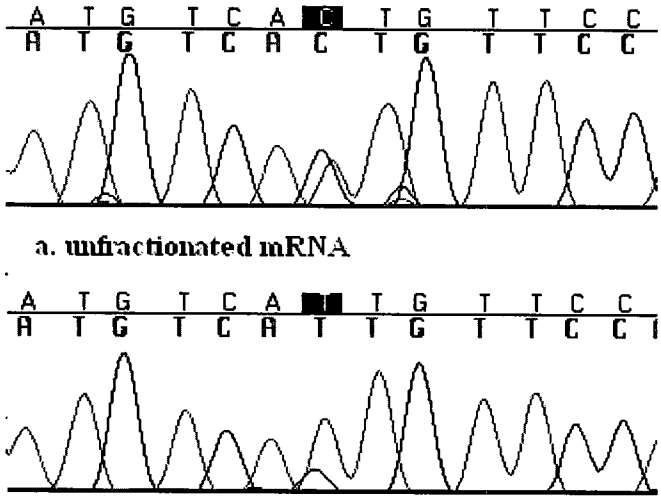

FIG. 2. A typical allelic imbalance (AI) pattern by the re-sequencing method. The nucleotide letter corresponding to the heterozygous SNP is highlighted in black. With the allelic ratio of unfractionated mRNA (a) as 1:1 stoichiometry control, the allelic ratio of fraction mRNA (b) was calculated (in this case, C:T=1: 4.3).

4.5.2 Microarray

For a much-higher throughput, our fractions were run on the *Illumina* Golden Gate genotyping array platform (*Illumina*, San Diego, CA), using a modification allowing relative quantification of the two alleles to detect AI, developed by the group of T.J. Hudson and T.M. Pastinen (Pastinen and Hudson 2004).

4.5.2.1 Samples

Ten unrelated CEPH CEU individuals, i.e. NA07357, NA11992, NA11993, NA12003, NA12043, NA12145, NA12155, NA12156, NA12761, and NA12815, were assayed on an *Illumina* Golden Gate array. For each individual, 6 samples were run on the array: (1) The genomic DNA; (2) The unfractionated mRNA in duplicate assays; and (3) Four mRNA fractions, corresponding to pre-80S mRNA, light polysome mRNA, medium polysome mRNA, and heavy polysome mRNA, each in triplicate.

4.5.2.2 The *Illumina* Golden Gate beadarray

(1) The plate design: The *Illumina* Golden Gate beadarray is a flexible, customized technology. The beadarray assay included 375 SNPs from 272 genes (Table 1). These genes were from the ENCODE regions (http://www.hapmap.org/downloads/encode1.html.en), chromosome 21, and genes with high variance in their expression among individuals (Serre, Gurd et al. 2008). The ENCODE regions had been extensively resequenced for SNP discovery and linkage disequilibrium (LD) mapping. In addition, these genes were known to be transcribed in LCLs by expression microarray assay. There were 371 SNPs from 269 genes which had at least one heterozygote, i.e. informative for the allelic translation assay. For the 371 SNPs, there were 10 5'UTR SNPs, 47 nonsynonymous SNPs (nsSNPs), 79 synonymous SNPs (sSNPs), and 235 3'UTR SNPs. The chromosome distribution of the 371 SNPs is shown in Fig.3.

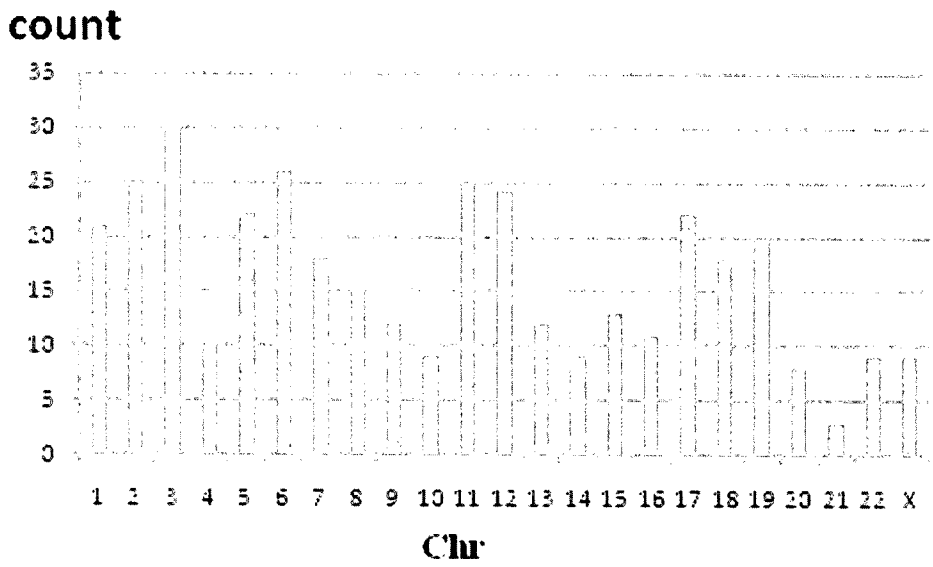

FIG. 3. The chromosome distribution of the 371 SNPs assayed for the allelic translation using the *Illumina* Golden Gate beadarray.

Table 1 The SNPs in the *Illumina* Golden Gate beadarray for the translational assay

rs	chr	pos	gene	region	HET n*
rs1203651	1	13,981,700	*PRDM2*	sSNP	2
rs10889205	1	40,478,312	*RLF*	nsSNP	3
rs12855	1	51,212,680	*CDKN2C*	3'UTR	4
rs7308	1	51,593,577	*EPS15*	3'UTR	9
rs17567	1	51,599,508	*EPS15*	nsSNP	6
rs7374	1	55,088,909	*DHCR24*	3'UTR	8
rs8535	1	111,587,451	*CHI3L2*	3'UTR	10
rs1056825	1	111,587,452	*CHI3L2*	3'UTR	12
rs1056831	1	111,587,534	*CHI3L2*	3'UTR	9
rs6573	1	112,056,911	*RAP1A*	3'UTR	2
rs14804	1	115,051,365	*NRAS*	3'UTR	3
rs878471	1	148,814,370	*MCL1*	3'UTR	7
rs11552229	1	149,050,608	*ARNT*	3'UTR	6
rs2228099	1	149,075,512	*ARNT*	sSNP	7
rs8401	1	151,800,521	*S100A2*	3'UTR	2
rs3753565	1	184,583,110	*TPR*	nsSNP	5
rs4245739	1	202,785,464	*MDM4*	3'UTR	4
rs4252745	1	202,785,809	*MDM4*	3'UTR	3
rs11240353	1	203,322,923	*RBBP5*	3'UTR	4
rs3024496	1	205,008,486	*IL10*	3'UTR	5
rs873	1	234,207,361	*NID*	3'UTR	4
rs10593	2	9,463,480	*ITGB1BP1*	3'UTR	3
rs4798	2	9,464,255	*ITGB1BP1*	3'UTR	3
rs1049500	2	10,498,417	*ODC1*	sSNP	2
rs162549	2	38,148,959	*CYP1B1*	3'UTR	7
rs2855658	2	38,150,393	*CYP1B1*	3'UTR	6
rs1800935	2	47,876,618	*MSH6*	sSNP	3
rs848291	2	58,242,199	*FANCL*	sSNP	5
rs2271627(rs6886)	2	85,475,569	*CAPG*	nsSNP	3
rs3731828	2	85,659,776	*VAMP8*	sSNP	9
rs1058588	2	85,662,381	*VAMP8*	3'UTR	10
rs1010	2	85,662,492	*VAMP8*	3'UTR	9
rs1304037	2	113,248,706	*IL1A*	3'UTR	6
rs17561	2	113,253,693	*IL1A*	nsSNP	5
rs1071676	2	113,303,903	*IL1B*	3'UTR	6
rs4556933	2	158,152,134	*ACVR1C*	sSNP	5
rs2066459	2	190,416,956	*PMS1*	nsSNP	1
rs7224	2	201,426,154	*CLK1*	3'UTR	2
rs1061157	2	203,129,443	*BMPR2*	sSNP	3
rs1048829	2	203,138,700	*BMPR2*	3'UTR	4
rs3731696	2	203,140,048	*BMPR2*	3'UTR	4
rs2229571	2	215,353,708	*BARD1*	nsSNP	5
rs2070096	2	215,353,789	*BARD1*	sSNP	6
rs207906	2	216,721,145	*XRCC5*	sSNP	3
rs1051677	2	216,778,492	*XRCC5*	3'UTR	2
rs1051685	2	216,778,620	*XRCC5*	3'UTR	6

Continuation of Table 1

rs7626117	3	10,115,670	*FANCD2*	3'UTR	2
rs1642742	3	10,166,942	*VHL*	3'UTR	4
rs1051208	3	12,600,746	*RAF1*	3'UTR	2
rs2470352	3	14,161,830	*XPC*	3'UTR	6
rs2228000	3	14,174,888	*XPC*	nsSNP	5
rs1046048	3	38,499,745	*ACVR2B*	sSNP	5
rs2953	3	41,256,391	*CTNNB1*	3'UTR	8
rs3448	3	49,371,754	*RHOA*	3'UTR	9
rs1061474	3	50,119,954	*RBM5*	sSNP	5
rs1043261	3	53,874,315	*IL17RB*	nsSNP	4
rs3774729	3	63,957,121	*ATXN7*	nsSNP	4
rs6441516	3	101,949,707	*TFG*	sSNP	8
rs7297	3	101,950,365	*TFG*	3'UTR	9
rs2305035	3	106,921,715	*CBLB*	sSNP	5
rs2681417	3	123,307,886	*CD86*	nsSNP	2
rs1129055	3	123,321,008	*CD86*	nsSNP	7
rs1131274	3	135,359,104	*RYK*	3'UTR	5
rs2227928	3	143,764,301	*ATR*	nsSNP	4
rs3182285	3	150,231,069	*SMARCA3*	3'UTR	5
rs2290725	3	150,242,014	*SMARCA3*	sSNP	5
rs568408	3	161,196,160	*IL12A*	3'UTR	2
rs3772172	3	171,560,406	*SKIL*	5'UTR	5
rs1131542	3	173,706,619	*TNFSF10*	3'UTR	3
rs1131535	3	173,706,768	*TNFSF10*	3'UTR	7
rs1131532	3	173,706,996	*TNFSF10*	sSNP	3
rs11919795	3	185,373,561	*DVL3*	3'UTR	6
rs6141	3	185,572,959	*THPO*	3'UTR	5
rs1047148	3	187,990,450	*RFC4*	3'UTR	8
rs406271	3	197,261,372	*TFRC*	3'UTR	7
rs17788379	3	197,261,607	*TFRC*	3'UTR	2
rs2245466	4	39,875,240	*RHOH*	5'UTR	4
rs2305948	4	55,674,314	*KDR*	nsSNP	4
rs1042040	4	56,955,054	*PPAT*	3'UTR	6
rs3733326	4	56,955,990	*PPAT*	3'UTR	3
rs4150052	4	78,298,795	*CCNG2*	sSNP	1
rs958	4	87,156,897	*MAPK10*	3'UTR	6
rs10965	4	89,217,106	*PKD2*	3'UTR	5
rs1042891	4	110,829,534	*CASP6*	3'UTR	9
rs3747676	4	124,037,523	*FGF2*	3'UTR	5
rs1049216	4	185,787,082	*CASP3*	3'UTR	6
rs702689	5	56,213,199	*MAP3K1*	nsSNP	9
rs832582	5	56,213,499	*MAP3K1*	nsSNP	6
rs832583	5	56,213,973	*MAP3K1*	sSNP	6
rs3730089	5	67,623,903	*PIK3R1*	nsSNP	6
rs300239	5	73,966,506	*ENC1*	sSNP	5
rs1805355	5	80,001,784	*MSH3*	sSNP	2
rs184967	5	80,185,736	*MSH3*	nsSNP	8
rs1056503	5	82,684,732	*XRCC4*	sSNP	2
rs2229085	5	108,161,865	*FER*	sSNP	3
rs2229992	5	112,190,752	*APC*	sSNP	6

Continuation of Table 1

rs397768	5	112,209,474	*APC*	3'UTR	5
rs3317	5	112,240,049	*C5orf18*	3'UTR	12
rs2545166	5	112,241,703	*C5orf18*	3'UTR	7
rs1549181	5	112,390,738	*MCC*	3'UTR	4
rs1051643	5	126,199,897	*LMNB1*	3'UTR	5
rs1059110	5	138,294,444	*CTNNA1*	sSNP	4
rs1059829	5	151,022,221	*SPARC*	3'UTR	10
rs1059279	5	151,022,371	*SPARC*	3'UTR	4
rs1054204	5	151,022,742	*SPARC*	3'UTR	7
rs299290	5	162,835,093	*HMMR*	nsSNP	1
rs17759	5	177,570,487	*HNRPAB*	3'UTR	11
rs9605	5	179,595,856	*MAPK9*	3'UTR	11
rs1013062	6	18,345,639	*DEK*	sSNP	7
rs1045537	6	26,204,726	*HFE*	3'UTR	5
rs2857713(rs2229094)	6	31,648,534	*LTA*	nsSNP	4
rs805256(rs14365)	6	31,743,688	*CSNK2B*	sSNP	4
rs522162	6	32,027,895	*RDBP*	3'UTR	1
rs760070	6	32,027,934	*RDBP*	3'UTR	1
rs8084	6	32,519,012	*HLA-DRA*	sSNP	10
rs7194	6	32,520,457	*HLA-DRA*	3'UTR	11
rs16871435	6	32,735,729	*HLA-DQB1*	3'UTR	3
rs9273960(rs1049133)	6	32,737,824	*HLA-DQB1*	sSNP	7
rs11244	6	32,888,701	*HLA-DOB*	3'UTR	10
rs2070121	6	32,889,531	*HLA-DOB*	nsSNP	1
rs8807	6	33,140,828	*HLA-DPA1*	3'UTR	6
rs7905	6	33,140,952	*HLA-DPA1*	3'UTR	2
rs9277534	6	33,162,784	*HLA-DPB1*	3'UTR	3
rs9277535	6	33,162,838	*HLA-DPB1*	3'UTR	5
rs210135(rs210135)	6	33,648,670	*BAK1*	3'UTR	2
rs8510	6	36,186,157	*MAPK14*	3'UTR	2
rs3025039	6	43,860,513	*VEGF*	3'UTR	5
rs4712138	6	56,571,368	*DST*	nsSNP	4
rs2172710	6	91,282,457	*MAP3K7*	3'UTR	2
rs2235481	6	146,098,540	*EPM2A*	5'UTR	7
rs220721	6	160,248,609	*MAS1*	sSNP	4
rs998075	6	160,388,267	*IGF2R*	sSNP	7
rs1040	6	169,358,271	*THBS2*	3'UTR	4
rs8770	6	170,728,455	*PDCD2*	3'UTR	13
rs1805321	7	5,993,513	*PMS2*	nsSNP	0
rs2066853	7	17,345,634	*AHR*	nsSNP	1
rs10259620	7	27,168,813	*HOXA9*	3'UTR	3
rs3735135	7	39,994,008	*CDC2L5*	nsSNP	2
rs10228436	7	55,205,761	*EGFR*	3'UTR	3
rs9530	7	65,063,328	*GUSB*	nsSNP	13
rs1202283	7	86,920,227	*ABCB4*	sSNP	5
rs3750105	7	94,132,617	*PEG10*	3'UTR	2
rs13073	7	94,134,704	*PEG10*	3'UTR	5
rs7810469	7	94,135,749	*PEG10*	3'UTR	1
rs6706	7	100,308,979	*TRIP6*	3'UTR	6
rs4710	7	115,684,627	*TES*	sSNP	4

Continuation of Table 1

rs2896181	7	115,685,119	*TES*	3'UTR	4
rs8713	7	115,987,032	*CAV1*	3'UTR	3
rs1049334	7	115,987,615	*CAV1*	3'UTR	5
rs6867	7	115,987,758	*CAV1*	3'UTR	3
rs1061285	7	128,640,543	*SMO*	3'UTR	7
rs4731575	7	128,855,382	*KIAA0828*	3'UTR	8
rs1665105	7	128,856,213	*KIAA0828*	3'UTR	2
rs2250788	8	11,389,464	*BLK*	5'UTR	4
rs1044011	8	12,986,079	*DLC1*	3'UTR	3
rs1047275	8	22,936,106	*TNFRSF10B*	3'UTR	5
rs3182143	8	29,263,763	*DUSP4*	5'UTR	4
rs1800392	8	31,093,498	*WRN*	sSNP	5
rs1346044	8	31,144,195	*WRN*	nsSNP	5
rs10503929	8	32,733,524	*NRG1*	nsSNP	5
rs1043782	8	38,187,480	*BAG4*	3'UTR	8
rs3242	8	41,238,710	*SFRP1*	3'UTR	3
rs1061302	8	91,027,597	*NBS1*	sSNP	12
rs1063045	8	91,064,194	*NBS1*	5'UTR	6
rs3812471	8	125,567,727	*RNF139*	sSNP	5
rs7460	8	141,738,041	*PTK2*	3'UTR	7
rs2294008	8	143,758,932	*PSCA*	3'UTR	8
rs9071	8	145,721,313	*LRRC14*	3'UTR	12
rs2230724	9	5,071,779	*JAK2*	sSNP	3
rs7048717(rs2230724)	9	5,071,779	*JAK2*	sSNP	3
rs3088440	9	21,958,158	*CDKN2A*	3'UTR	3
rs11515	9	21,958,198	*CDKN2A*	3'UTR	5
rs4142496(rs5812)	9	35,792,582	*NPR2*	sSNP	5
rs17062695	9	71,030,837	*TJP2*	sSNP	2
rs3812536	9	71,059,292	*TJP2*	3'UTR	9
rs16909910(rs1805155)	9	97,278,199	*PTCH*	sSNP	5
rs4978877	9	111,970,683	*AKAP2*	3'UTR	1
rs11681	9	111,972,149	*PALM2-AKAP2*	3'UTR	2
rs8083	9	129,591,999	*CDK9*	3'UTR	5
rs1056209	9	132,752,410	*ABL1*	3'UTR	2
rs11254401	10	17,228,823	*DNMT2*	3'UTR	1
rs3817405	10	20,546,423	*PLXDC2*	nsSNP	4
rs1042058	10	30,768,106	*MAP3K8*	sSNP	7
rs1131510	10	35,339,090	*CUL2*	3'UTR	3
rs1871446	10	62,223,768	*CDC2*	3'UTR	3
rs1468063	10	90,765,270	*FAS*	3'UTR	6
rs14401	10	112,035,160	*MXI1*	3'UTR	7
rs17658	10	112,036,139	*MXI1*	3'UTR	10
rs2459216	10	126,076,315	*OAT*	3'UTR	1
rs8839	11	1,730,711	*CTSD*	3'UTR	2
rs8234	11	2,826,683	*KCNQ1*	3'UTR	4
rs10798	11	2,826,740	*KCNQ1*	3'UTR	7
rs2239731	11	6,933,563	*ZNF215*	sSNP	6
rs1042359	11	8,202,756	*LMO1*	sSNP	5
rs1049403	11	9,567,380	*WEE1*	3'UTR	2

Continuation of Table 1

rs4447177	11	22,602,600	*FANCF*	3'UTR	1
rs7357	11	33,686,836	*CD59*	3'UTR	6
rs842	11	33,687,981	*CD59*	3'UTR	3
rs8193	11	35,207,893	*CD44*	3'UTR	5
rs13347	11	35,209,847	*CD44*	3'UTR	3
rs1061810	11	43,834,509	*HSD17B12*	3'UTR	2
rs1057233	11	47,333,023	*SPI1*	3'UTR	4
rs4246215	11	61,320,874	*FEN1*	3'UTR	6
rs947894(rs1695)	11	67,109,264	*GSTP1*	nsSNP	11
rs4891	11	67,110,545	*GSTP1*	sSNP	12
rs14983	11	101,896,634	*MMP7*	3'UTR	2
rs4445669	11	114,550,446	*IGSF4*	3'UTR	8
rs488219	11	118,125,906	*DDX6*	3'UTR	8
rs487728	11	118,126,704	*DDX6*	3'UTR	10
rs1047417	11	118,677,745	*CBL*	3'UTR	5
rs11217234	11	118,683,147	*CBL*	3'UTR	5
rs506504	11	125,030,404	*CHEK1*	nsSNP	3
rs8705	11	127,834,122	*ETS1*	3'UTR	11
rs4937333	11	127,835,729	*ETS1*	3'UTR	12
rs4980809	12	276,552	*JARID1A*	sSNP	8
rs7310449	12	892,375	*RAD52*	3'UTR	5
rs7301931	12	892,444	*RAD52*	3'UTR	5
rs3217926	12	4,281,943	*CCND2*	3'UTR	3
rs1049612	12	4,283,022	*CCND2*	3'UTR	3
rs3217933	12	4,283,260	*CCND2*	3'UTR	6
rs2302371	12	6,728,386	*MLF2*	sSNP	2
rs2301262	12	6,926,120	*PTPN6*	5'UTR	4
rs1058028	12	11,935,461	*ETV6*	3'UTR	6
rs7330	12	12,766,183	*CDKN1B*	3'UTR	12
rs921	12	14,986,428	*ARHGDIB*	3'UTR	4
rs4703	12	14,986,824	*ARHGDIB*	sSNP	4
rs8664	12	15,664,687	*EPS8*	3'UTR	4
rs12587	12	25,250,094	*KRAS*	3'UTR	6
rs13096	12	25,251,107	*KRAS*	3'UTR	5
rs9266	12	25,253,483	*KRAS*	3'UTR	8
rs1137282	12	25,254,043	*KRAS*	3'UTR	4
rs2230375	12	26,624,328	*ITPR2*	sSNP	5
rs2291264	12	26,702,043	*ITPR2*	sSNP	3
rs1900941	12	26,759,588	*ITPR2*	sSNP	6
rs4251545	12	42,466,561	*IRAK4*	nsSNP	2
rs11183605	12	45,445,234	*SLC38A4*	3'UTR	3
rs769412	12	67,519,481	*MDM2*	sSNP	4
rs4135113	12	102,900,822	*TDG*	nsSNP	0
rs14035	12	129,927,193	*RAN*	3'UTR	7
rs546782	13	21,173,945	*FGF9*	3'UTR	3
rs144848	13	31,804,728	*BRCA2*	nsSNP	5
rs1801406	13	31,809,887	*BRCA2*	sSNP	4
rs1047775	13	42,579,426	*DNAJD1*	3'UTR	3
rs5351	13	77,373,313	*EDNRB*	sSNP	8
rs2389910	13	96,915,164	*RAP2A*	3'UTR	9

Continuation of Table 1

rs12873919	13	96,917,524	*RAP2A*	3'UTR	14
rs17655	13	102,326,002	*ERCC5*	nsSNP	6
rs10131	13	107,657,846	*LIG4*	3'UTR	3
rs2289047	13	109,205,815	*IRS2*	3'UTR	6
rs2289046	13	109,205,906	*IRS2*	3'UTR	4
rs1061386	13	110,171,368	*ING1*	3'UTR	6
rs1803479	14	22,103,829	*DAD1*	3'UTR	2
rs4981429(rs7621)	14	22,113,841	*DAD1*	3'UTR	5
rs1051101	14	22,127,923	*DAD1*	5'UTR	3
rs11569620	14	34,940,706	*NFKBIA*	3'UTR	2
rs10782383(rs1050851)	14	34,942,676	*NFKBIA*	sSNP	1
rs1957106	14	34,943,520	*NFKBIA*	sSNP	5
rs1051861	14	57,908,453	*ARID4A*	sSNP	8
rs11549465	14	61,277,309	*HIF1A*	nsSNP	3
rs1131877	14	102,411,801	*TRAF3*	nsSNP	3
rs705	15	22,770,604	*SNRPN*	5'UTR	9
rs743581	15	72,115,193	*PML*	3'UTR	7
rs9479	15	72,115,628	*PML*	3'UTR	14
rs3129	15	77,001,269	*CTSH*	3'UTR	6
rs3826007	15	78,050,271	*BCL2A1*	nsSNP	3
rs859	15	79,388,376	*IL16*	3'UTR	9
rs1131445	15	79,388,836	*IL16*	3'UTR	11
rs17273563(rs2227933)	15	89,138,482	*BLM*	sSNP	6
rs17274095(rs2227934)	15	89,147,926	*BLM*	sSNP	6
rs1063147	15	89,155,508	*BLM*	sSNP	7
rs4777755	15	91,311,606	*CHD2*	sSNP	1
rs2272457	15	91,337,200	*CHD2*	sSNP	5
rs1048326	15	99,427,167	*LRRK1*	3'UTR	4
rs393521	16	277,678	*AXIN1*	3'UTR	3
rs1051771	16	2,078,584	*TSC2*	sSNP	1
rs1054028	16	22,834,714	*HS3ST2*	3'UTR	3
rs7593	16	24,490,906	*RBBP6*	sSNP	6
rs3135500	16	49,324,386	*CARD15*	3'UTR	6
rs2066852	16	49,385,018	*CYLD*	sSNP	1
rs10748	16	52,062,221	*RBL2*	sSNP	7
rs3929	16	52,081,808	*RBL2*	3'UTR	6
rs28216	16	63,579,614	*CDH11*	sSNP	5
rs1800566	16	68,302,645	*NQO1*	nsSNP	4
rs2239359	16	88,376,980	*FANCA*	nsSNP	6
rs12727	17	1,747,814	*RPA1*	3'UTR	4
rs2270121	17	9,757,395	*GAS7*	3'UTR	6
rs1047365	17	9,759,694	*GAS7*	3'UTR	7
rs4792219	17	11,986,629	*MAP2K4*	3'UTR	9
rs1801052	17	26,532,900	*NF1*	sSNP	4
rs2285892	17	26,577,610	*NF1*	sSNP	0
rs7505	17	26,668,977	*EVI2A*	3'UTR	8
rs13695	17	35,798,718	*TOP2A*	3'UTR	8
rs3198502	17	37,716,519	*STAT5A*	3'UTR	5
rs1053004	17	37,719,617	*STAT3*	3'UTR	8
rs1799966	17	38,476,619	*BRCA1*	nsSNP	5

Continuation of Table 1

rs16942	17	38,497,525	*BRCA1*	nsSNP	5
rs799917	17	38,498,461	*BRCA1*	nsSNP	5
rs16940	17	38,498,762	*BRCA1*	sSNP	5
rs1049620	17	44,836,512	*PHB*	3'UTR	4
rs1061237	17	45,617,773	*COL1A1*	3'UTR	5
rs4626	17	46,295,420	*TOB1*	sSNP	3
rs3088093	17	59,560,169	*ERN1*	3'UTR	2
rs8905	17	64,039,396	*PRKAR1A*	3'UTR	1
rs6958	17	64,040,372	*PRKAR1A*	3'UTR	6
rs9367	17	71,265,255	*ITGB4*	3'UTR	3
rs2239680	17	73,731,377	*BIRC5*	3'UTR	4
rs1042489	17	73,731,800	*BIRC5*	3'UTR	2
rs699517	18	663,015	*TYMS*	3'UTR	7
rs2790	18	663,085	*TYMS*	3'UTR	8
rs1061035	18	712,117	*YES1*	3'UTR	2
rs1060922	18	712,401	*YES1*	3'UTR	6
rs677688	18	20,274,540	*IMPACT*	nsSNP	1
rs1053474	18	20,287,258	*IMPACT*	3'UTR	8
rs10470	18	46,860,013	*SMAD4*	3'UTR	2
rs2229082(rs2229082)	18	49,190,992	*DCC*	sSNP	7
rs7614	18	49,935,241	*MBD2*	3'UTR	11
rs6567211(rs8766)	18	51,046,528	*TCF4*	sSNP	8
rs8766	18	51,046,528	*TCF4*	sSNP	9
rs1059442	18	54,567,672	*MALT1*	3'UTR	5
rs4987853	18	58,944,634	*BCL2*	3'UTR	6
rs4987852	18	58,944,900	*BCL2*	3'UTR	1
rs1564483	18	58,945,633	*BCL2*	3'UTR	3
rs6810	18	59,149,380	*FVT1*	3'UTR	10
rs2288288	18	59,705,945	*SERPINB2*	5'UTR	1
rs6104	18	59,721,508	*SERPINB2*	nsSNP	3
rs8259	19	533,926	*BSG*	3'UTR	3
rs3752174	19	2,465,564	*GNG7*	3'UTR	3
rs1053395	19	6,445,370	*TUBB4*	3'UTR	4
rs3093032	19	10,257,335	*ICAM1*	3'UTR	2
rs7935	19	10,966,607	*SMARCA4*	sSNP	7
rs2482	19	14,591,108	*EMR3*	3'UTR	1
rs16982300(rs9413)	19	18,251,867	*JUND*	3'UTR	2
rs6554	19	18,546,963	*UBA52*	3'UTR	10
rs1406	19	35,006,951	*CCNE1*	3'UTR	11
rs2070132	19	41,419,204	*ZNF146*	nsSNP	4
rs4806293	19	41,420,129	*ZNF146*	3'UTR	6
rs3547	19	48,739,389	*XRCC1*	sSNP	10
rs7026	19	50,016,355	*LU*	3'UTR	3
rs1052559(rs13181)	19	50,546,758	*ERCC2*	nsSNP	6
rs13181	19	50,546,758	*ERCC2*	nsSNP	6
rs1049698	19	50,668,597	*FOSB*	3'UTR	3
rs1049739	19	50,670,214	*FOSB*	3'UTR	5
rs4645900	19	54,156,174	*BAX*	3'UTR	2
rs704243	19	54,156,782	*BAX*	3'UTR	1
rs10413435	19	57,420,841	*PPP2R1A*	sSNP	4

Continuation of Table 1

rs16997057(rs3197744)	20	1,866,486	*PTPNS1*	3'UTR	5
rs8156	20	3,734,636	*CDC25B*	3'UTR	2
rs7828	20	10,567,013	*JAG1*	3'UTR	2
rs6061216	20	30,245,773	*PLAGL2*	3'UTR	5
rs2424932	20	30,860,196	*DNMT3B*	3'UTR	3
rs2076546	20	45,701,899	*NCOA3*	sSNP	2
rs7121	20	56,912,201	*GNAS*	3'UTR	2
rs3730171	20	56,917,635	*GNAS*	3'UTR	1
rs1059293	21	33,731,562	*IFNGR2*	3'UTR	6
rs3165	21	37,517,781	*DSCR3*	3'UTR	2
rs1051420	21	39,117,333	*ETS2*	3'UTR	3
rs9975285(rs2230688)	21	45,720,739	*COL18A1*	sSNP	0
rs1548410	22	19,634,462	*CRKL*	3'UTR	3
rs1043242	22	19,636,742	*CRKL*	3'UTR	5
rs180817	22	21,988,006	*BCR*	3'UTR	6
rs5030612(rs2229354)	22	22,497,512	*SMARCB1*	sSNP	4
rs1049583	22	30,682,790	*YWHAH*	3'UTR	4
rs763121	22	37,209,885	*DDX17*	3'UTR	3
rs86796	22	37,211,399	*DDX17*	3'UTR	4
rs5750609	22	37,214,720	*DDX17*	3'UTR	1
rs2272857	22	49,041,794	*MAPK12*	sSNP	4
rs1057403	X	100,491,336	*BTK*	3'UTR	5
rs700	X	100,491,412	*BTK*	3'UTR	6
rs5956583	X	122,862,191	*BIRC4*	nsSNP	4
rs9856	X	122,873,321	*BIRC4*	3'UTR	6
rs8371	X	122,873,595	*BIRC4*	3'UTR	2
rs5958343	X	122,873,952	*BIRC4*	3'UTR	4
rs1059701	X	152,937,676	*IRAK1*	sSNP	1
rs12877	X	153,283,448	*DNASE1L1*	3'UTR	4
rs11887	X	154,120,650	*VBP1*	3'UTR	4

* The number of heterozygotes for the allelic translation assay.

(2) The *Illumina* Golden Gate assay: For the RNA samples to be run on the *Illumina* assay, all the unfractionated and fractionated RNA samples were standardized based on the poly(A) RNA quantification, to their equivalent in total RNA. Exactly 250ng of 50ng/μl gDNA or equivalent amount of cDNA was activated for binding to paramagnetic particles, and immobilized on a solid support. Consequently, two allele-specific oligos (ASOs) and a locus-specific oligo (LSO) were hybridized to the template gDNA or cDNA. Following hybridization, several wash steps were performed to reduce noise and to remove excess and mis-hybridized probes. Because of the immobilization of the DNA templates, extra washing was permitted. The two ASOs and the LSO carried common complementary sequences for universal primers, which allowed universal PCR amplification. The two ASOs carried the complementary sequences for the universal primers P1 and P2 respectively. The LSO carried the complementary sequences for the universal primer P3. After hybridization and clean-up, the appropriate ASO was extended and ligated to the LSO, and formed the templates for the following universal PCR. During the PCR reaction, the universal primers P1 and P2 introduced two different dyes into the PCR products. The PCR products were hybridized to complementary beads according to the unique address sequences carried by the LSOs. The fluorescence signals introduced by the universal primers P1 or P2 were analyzed by the BeadArray Reader, and the allele ratio were called on the heterozygotes (http://www.*Illumina*.com/pages.ilmn?ID=11) (Fan, Chee et al. 2006).

4.6 Data analysis

As described in the resequencing method, we adapted two approaches to quantify the allelic ratio. To quantify the allelic ratio in the unfractionated mRNA, we used allelic proportions in heterozygous gDNA as 1:1 control; to quantify the allelic ratio in the polysome fractions, we used allelic proportions in heterozygous unfractionated mRNA as 1:1 control. Compared with allelic ratios in unfractionated mRNA, SNPs showing different AI pattern in the fractions were considered as an indication of *cis* translational effects of variants contained in the gene transcript.

To assess the translational effect, I used this scoring system: (1) Statistical significance between the two light fractions (pre-80S and light polysome) and the two heavy fractions (medium polysome and heavy polysome) at =0.05 by the Student's t test of the average of three replicates for each fraction; (2) Biologically meaningful difference between the two light fractions and the two heavy fractions, i.e. the difference of the allele proportion was more than 0.10, which is an empirical cutoff to judge AI generally used in our study; (3) Progressive change of allele proportions from pre-80S to heavy polysome, i.e. the correlation coefficient r between the allele ratio and the fraction points was more than 0.90; (4) Biologically meaningful difference of allele proportions from pre-80S to heavy polysome, i.e. the slope of the line was more than 0.05 by the regression analysis, which is set up according to the empirical AI threshold. Each of the four criteria was given one point with the total score of 4. If the accumulated score was

≥3, the CEPH individual would be taken as showing a possible translational effect at this genetic locus. Concordant results from different SNPs in the same gene were further evidence for TI.

5. Results

5.1 Polysome fractionation

Forty LCL lines from the samples of European individuals used for the HapMap project have been fractionated. The realtime OD monitoring profile is shown in Fig.4.

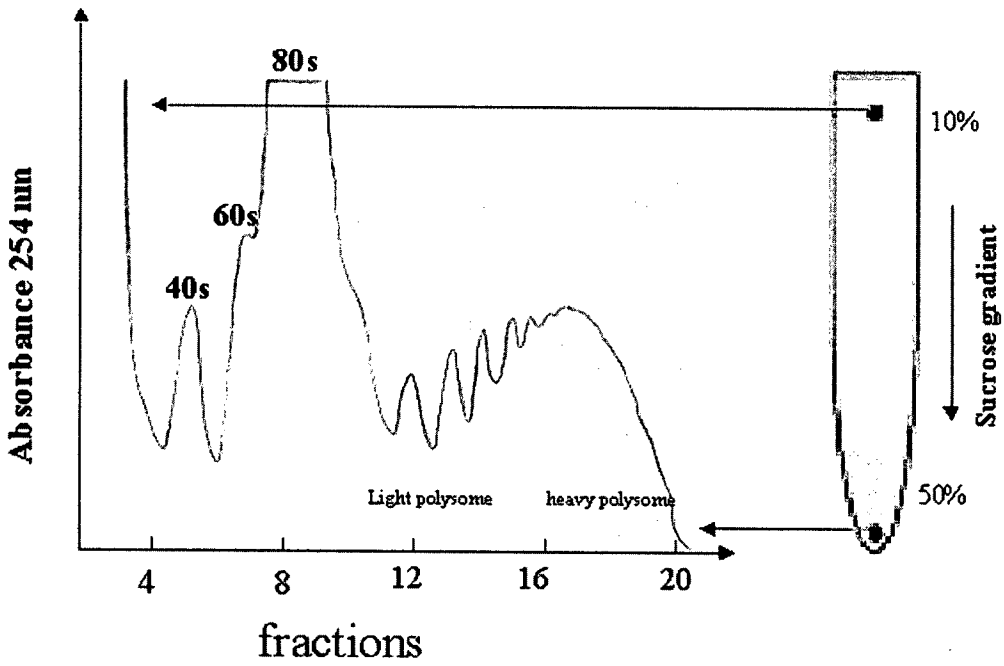

Fig.4. Realtime OD detection of the 24 polysome fractions. Fractions were monitored using an ISCO UA-6 UV detector. The positions of 80S ribosomes, light polysomes, and heavy polysomes, in the gradients are labeled. X-axis: the fraction number (from Fraction 1 to Fraction 24); Y-axis: the UV absorbance. Increases in polysome size by a single ribosome are indicated by secondary peaks in the up-slope of the broad polysome peak.

5.2 The Agilent assay of fraction RNA samples

The Agilent Lab-on-a-Chip assay verifies the amount of the 18S and 28S peaks by electrophoresis on a chip. Applied to my fractions (Fig.5) it shows that there is neither 18S nor 28S in Fraction1-6 (pre-40S RNA), which is exactly concordant with our realtime OD detection, which suggests these fractions contain mainly free proteins and RNA. From Fraction 7, we begin to see 18S RNA. From Fraction 7-10, there is only 18S, which is corresponding to the 40S realtime OD monitoring profile. From Fraction 11, we begin to see 28S peaks. The ratio of 28S/18S increases with fraction numbers. From Fraction 18, the ratio of 28S/18S begins to be >2, and increases until the last fraction, in a pattern perfectly concordant with the theoretically expected distribution of the subunits.

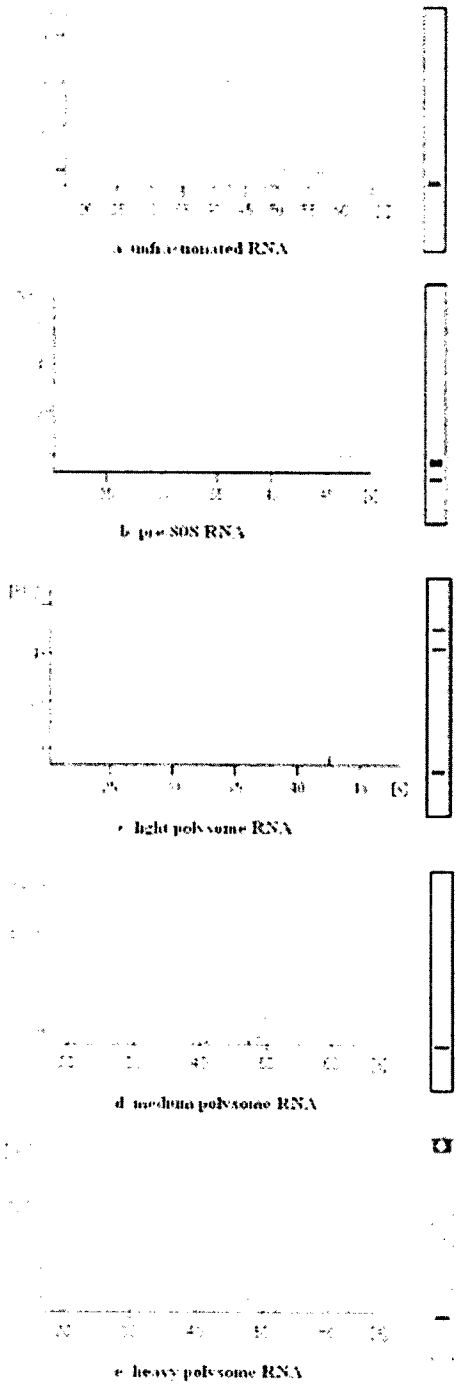

Fig.5. The Agilent assay of fraction RNA. From top to down, the five panels are the assays of unfractionated RNA, pre-40S RNA, light polysome RNA, medium polysome RNA, and heavy polysome RNA. From left to right, three sharp peaks seen in the panels correspond to 5S, 18S, and 28S respectively.

5.3 Expression profiles of six genes in the fractions

In order to determine the optimal fractions to sample for TI measurements, we studied six genes (*IFIH1*, *LTB4R*, *IL10RA*, *CD40*, *TAP2*, and the Cyclophilin housekeeping gene) with widely different expression levels in LCLs. Except Cyclophilin, these genes have important immune functions and may be involved in the pathogenesis of type 1 diabetes. Shown by the gene expression profiles, no gene transcript can be amplified in fraction1-2. In addition, different genes showed different expression profiles in the fractions, compatible with the expectation that different mRNAs have different translational property (Fig.6).

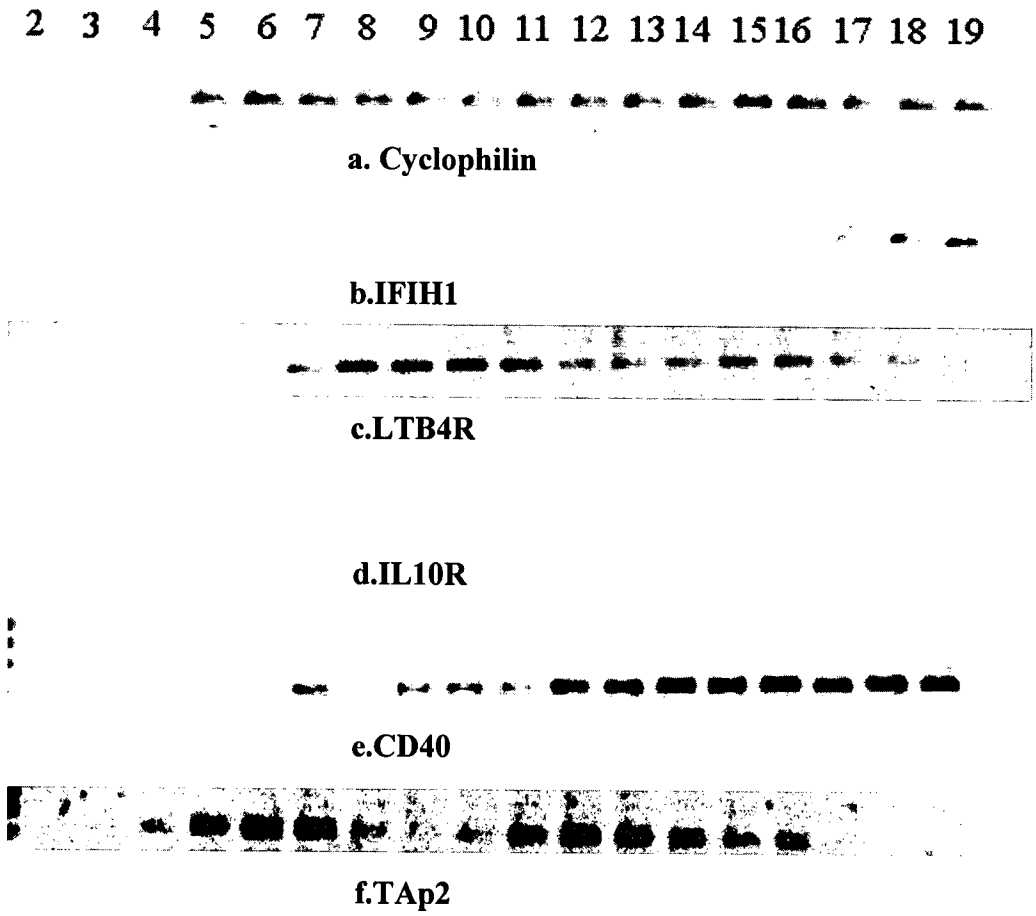

Fig.6. Gene expression profiles in the fractions. Each panel represents one gene. We can see that the mRNA of each gene has a different distribution in the polysome fractions. The Cyclophilin (a.) mRNAs are distributed in all the light, medium, and heavy fractions. The *IFIH1* (b.) mRNA can be seen mainly in heavy fractions. *LTB4R* (c.) mRNA has higher levels in light and medium fractions. The *IL10RA* (d.) can be seen mainly in medium and heavy fractions. The *CD40* (e.) gene is mainly seen in medium and heavy fractions. The *TAP2* (f.) gene has similar pattern with Cyclophilin.

5.4 Oligo-dT assay to quantify fraction poly(A) RNA

Fractionated RNA was diluted to 150ng/μl based on the OD assay. An 1μl sample from each fraction was used for the Poly(A) assay. The image of hybridization with DIG-labeled oligo(dT) is shown in Fig.7. We can see that mRNA is mainly seen in the middle and lower fractions, but none in the upper fractions. The highest peak is commonly seen in Fraction 7-11 (the original OD values right after RNA extraction are also the highest in these fractions), and the second peak is seen in Fraction 15-21. This pattern is concordant with the gene expression profiles. In my study, I established a robust standard curve to quantify fraction poly(A) RNA (Fig.8). According to the densitometry quantification, the poly(A) RNA normally accounts for 2-3% of total RNA in the middle fractions, and 1-2% of total RNA in the bottom fractions.

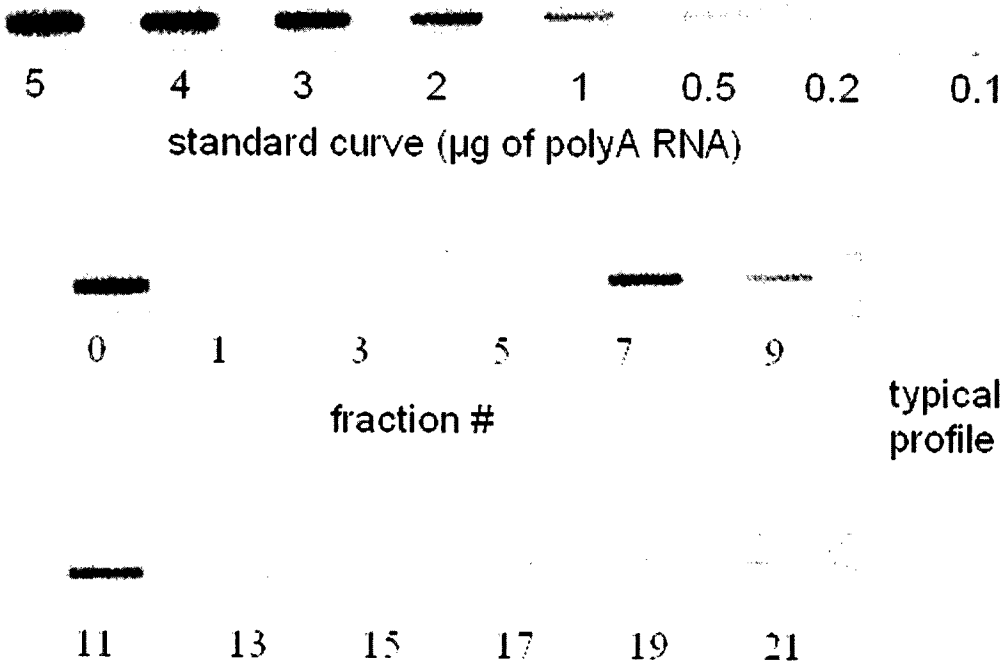

Fig.7. The DIG-labeled oligo-dT assay to quantify fraction poly(A) RNA. The first lane shows the commercial human pancreas Poly(A) RNA at the standard concentrations of 5ng/µl, 4ng/µl, 3ng/µl, 2ng/µl, 1ng/µl, 0.5ng/µl, 0.2ng/µl, and 0.1ng/µl respectively. The lower two lanes represent a typical profile of LCLs. The fraction numbers are shown under the bands. "0" represents unfractionated RNA.

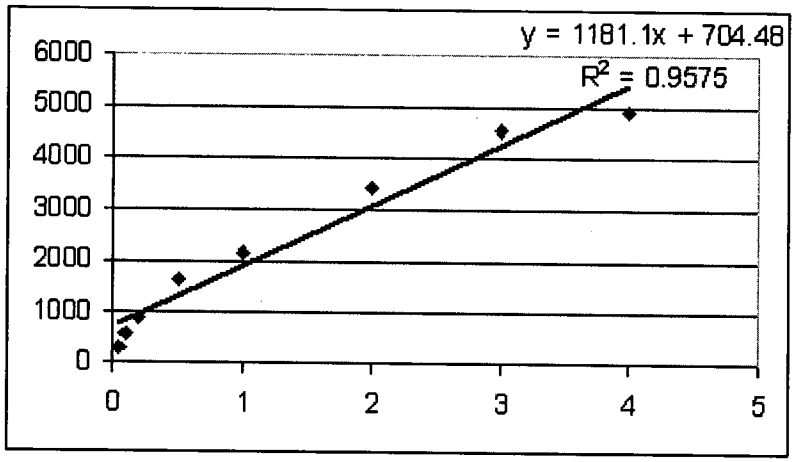

Fig. 8. The stand curve of the Poly(A) densitometry quantification. X-axis: Poly(A) concentration (ng/μl); Y-axis: OD value.

5.5 The fractionated RNA sample bank

I established a fractionated RNA sample bank, including 44 European CEU individuals used for the international HAPMAP project. Except one individual (NA10839), 43 individuals are the parents from the HAPMAP CEPH CEU families. The sample list is shown in Table 2. Each individual has a DNA sample, 1 unfractionated RNA sample, and 24 RNA samples from polysome fractions. Serving as a quality control, two cell lines (NA12872 and NA12146) were cultured and fractionated twice by myself. Four cell lines (NA06994, NA10839, NA12043, and NA12239) were cultured and fractionated twice by a different person (Chee LK Lam).

Table 2 The HAPMAP CEU individuals with fractionated RNA in my RNA bank

HAPMAP ID	Gender	Family ID
NA12145	female	EU1334
NA12146	male	EU1334
NA12239	female	EU1334
NA06994	male	EU1340
NA07000	female	EU1340
NA07056	female	EU1340
NA07034	male	EU1341
NA07055	female	EU1341
NA06993	male	EU1341
NA06985	female	EU1341
NA12056	male	EU1344
NA12057	female	EU1344
NA07357	male	EU1345
NA07345	female	EU1345
NA12043	male	EU1346
NA11881	male	EU1347
NA11840	female	EU1349
NA11829	male	EU1350
NA11830	female	EU1350
NA11831	male	EU1350
NA11832	female	EU1350
NA11992	male	EU1362
NA11993	female	EU1362
NA11994	male	EU1362
NA11995	female	EU1362
NA12234	female	EU1375
NA12154	male	EU1408
NA12155	male	EU1408
NA12156	female	EU1408
NA12248	male	EU1416
NA10839*	female	EU1420
NA12003	male	EU1420
NA12004	female	EU1420
NA12006	female	EU1420
NA12750	male	EU1444
NA12760	male	EU1447
NA12761	female	EU1447
NA12812	male	EU1454
NA12813	female	EU1454
NA12814	male	EU1454
NA12815	female	EU1454
NA12872	male	EU1459
NA12873	female	EU1459
NA12874	male	EU1459
NA12892	female	EU1463

* Offspring with the father NA12005 and the mother NA12006.

5.6 TI assay on the *Illumina* beadarray

TI was detected in 75 out of 1483 (5.06%) assays, and 63 out of 269 (23.4%) genes (Table 3). These results suggested that genetic effect on gene translation is a common mechanism of gene expression regulation.

Table 3 The interesting genes showing TI on the *Illumina* Golden Gate assay

SNP ID	chr	position	Gene	Sample	B allele ratio					P value[1]	Difference[2]	r[3]	slope[4]	score
					Unfracti-onated RNA	pre-80S	light polyso-me	medium polyso-me	heavy polyso-me					
rs17567	1	51,599,508	*EPS15*	NA07357	0.578	0.332	0.395	0.548	0.520	0.139	0.063	0.992	0.033	4
rs6573	1	112,056,911	*RAP1A*	NA12815	0.806	0.806	0.802	0.556	0.529	0.484	0.033	0.748	0.021	4
rs10593	2	9,463,480	*ITGB1BP1*	NA12043	0.583	0.227	0.422	0.503	0.537	0.186	0.195	0.941	0.101	3
rs17561	2	113,253,693	*IL1A*	NA12043	0.349	0.793	0.650	0.537	0.489	0.894	0.013	0.524	0.028	3
rs2070096	2	215,353,789	*BARD1*	NA12156	0.611	0.595	0.592	0.394	0.438	0.037	0.049	0.790	0.018	3
rs2953	3	41,256,391	*CTNNB1*	NA12815	0.382	0.465	0.356	0.293	0.255	0.144	0.140	0.981	0.072	3
rs1043261	3	53,874,315	*IL17RB*	NA12155	0.637	0.306	0.330	0.462	0.510	0.024	0.168	0.966	0.074	4
rs7297	3	101,950,365	*TFG*	NA12156	0.753	0.693	0.746	0.548	0.544	0.548	0.020	0.748	0.015	3
				NA12815	0.730	0.677	0.697	0.494	0.524	0.867	0.002	0.323	0.002	3
rs2681417	3	123,307,886	*CD86*	NA12155	0.495	0.198	0.282	0.400	0.420	0.059	0.170	0.970	0.078	3
rs2227928	3	143,764,301	*ATR*	NA12761	0.696	0.446	0.621	0.639	0.682	0.171	0.050	0.978	0.026	3
rs3772172	3	171,560,406	*SKIL*	NA12761	0.476	0.340	0.358	0.485	0.467	0.010	0.127	0.885	0.051	3
rs6141	3	185,572,959	*THPO*	NA11992	0.325	0.627	0.526	0.394	0.204	0.071	0.340	0.996	0.163	3
rs1047148	3	187,990,450	*RFC4*	NA12043	0.700	0.636	0.584	0.465	0.437	0.735	0.010	0.507	0.009	4
rs3733326	4	56,955,990	*PPAT*	NA12156	0.386	0.607	0.651	0.467	0.490	0.355	0.026	0.383	0.007	3
rs702689	5	56,213,199	*MAP3K1*	NA12043	0.552	0.433	0.449	0.572	0.627	0.031	0.158	0.964	0.071	4
rs2229992	5	112,190,752	*APC*	NA12156	0.638	0.637	0.662	0.461	0.509	0.728	0.019	0.133	0.004	3
rs3317	5	112,240,049	*C5orf18*	NA12156	0.434	0.695	0.767	0.521	0.549	0.861	0.007	0.464	0.011	3
rs1054204	5	151,022,742	*SPARC*	NA12815	0.380	0.510	0.459	0.281	0.341	0.305	0.126	0.465	0.038	3
rs299290	5	162,835,093	*HMMR*	NA12043	0.586	0.386	0.459	0.546	0.564	0.073	0.132	0.972	0.062	3
rs17759	5	177,570,487	*HNRPAB*	NA12156	0.395	0.048	0.111	0.175	0.259	0.197	0.105	0.981	0.057	3
rs9605	5	179,595,856	*MAPK9*	NA12155	0.793	0.759	0.703	0.522	0.502	0.851	0.012	0.562	0.020	4
rs2857713	6	31,648,534	*LTA*	NA12156	0.595	0.279	0.377	0.488	0.547	0.052	0.254	0.871	0.105	3
rs7194	6	32,520,457	*HLA-DRA*	NA07357	0.659	0.480	0.470	0.638	0.679	0.322	0.027	0.872	0.015	4
rs9273960	6	32,737,824	*HLA-DQB1*	NA12815	0.523	0.466	0.361	0.226	0.193	0.066	0.204	0.977	0.096	3
rs8807	6	33,140,828	*HLA-DPA1*	NA12156	0.791	0.750	0.753	0.535	0.561	0.817	0.010	0.485	0.012	3
rs7905	6	33,140,952	*HLA-DPA1*	NA12815	0.542	0.598	0.500	0.409	0.304	0.319	0.146	0.919	0.088	3

rs9277534	6	33,162,784	*HLA-DPB1*	NA12156	0.599	0.534	0.547	0.390	0.414	0.062	0.052	0.691	0.017	3
rs2066853	7	17,345,634	*AHR*	NA11993	0.418	0.249	0.334	0.380	0.410	0.148	0.104	0.974	0.053	3
rs10228436	7	55,205,761	*EGFR*	NA11992	0.787	0.861	0.844	0.773	0.675	0.124	0.129	0.959	0.063	3
rs13073	7	94,134,704	*PEG10*	NA07357	0.697	0.872	0.809	0.741	0.518	0.088	0.099	0.946	0.046	3
rs2896181	7	115,685,119	*TES*	NA12156	0.408	0.833	0.782	0.544	0.532	0.138	0.063	0.968	0.032	4
rs8713	7	115,987,032	*CAV1*	NA12815	0.534	0.680	0.593	0.488	0.476	0.071	0.155	0.963	0.072	3
rs10503929	8	32,733,524	*NRG1*	NA12156	0.812	0.780	0.773	0.556	0.562	0.537	0.012	0.626	0.007	3
				NA12815	0.777	0.744	0.739	0.536	0.525	0.017	0.022	0.928	0.009	4
rs1063045	8	91,064,194	*NBS1*	NA12043	0.671	0.258	0.324	0.480	0.488	0.029	0.193	0.952	0.085	4
rs3812536	9	71,059,292	*TJP2*	NA07357	0.417	0.571	0.636	0.402	0.449	0.472	0.036	0.096	0.003	3
rs1131510	10	35,339,090	*CUL2*	NA12145	0.580	0.417	0.545	0.564	0.620	0.202	0.076	0.983	0.042	3
rs14401	10	112,035,160	*MXI1*	NA12156	0.452	0.726	0.781	0.526	0.556	0.717	0.014	0.169	0.004	3
rs17658	10	112,036,139	*MXI1*	NA12145	0.684	0.504	0.535	0.578	0.678	0.172	0.109	0.962	0.057	3
rs10798	11	2,826,740	*KCNQ1*	NA12155	0.710	0.660	0.606	0.457	0.453	0.367	0.039	0.908	0.025	4
				NA12156	0.814	0.794	0.798	0.561	0.605	0.722	0.001	0.185	0.000	3
rs7357	11	33,686,836	*CD59*	NA12156	0.663	0.754	0.732	0.515	0.537	0.044	0.052	0.917	0.022	3
rs947894	11	67,109,264	*GSTP1*	NA12145	0.796	0.365	0.440	0.658	0.697	0.023	0.275	0.965	0.121	4
rs506504	11	125,030,404	*CHEK1*	NA12156	0.548	0.462	0.460	0.348	0.330	0.076	0.108	0.776	0.041	4
rs7310449	12	892,375	*RAD52*	NA11993	0.747	0.510	0.659	0.690	0.731	0.244	0.126	0.930	0.069	3
rs3217933	12	4,283,260	*CCND2*	NA12815	0.535	0.430	0.405	0.313	0.303	0.032	0.090	0.978	0.041	3
rs4703	12	14,986,824	*ARHGDIB*	NA11993	0.628	0.287	0.351	0.435	0.463	0.056	0.126	0.984	0.059	3
rs8664	12	15,664,687	*EPS8*	NA12155	0.485	0.198	0.215	0.341	0.383	0.005	0.187	0.888	0.075	4
rs13096	12	25,251,107	*KRAS*	NA12043	0.628	0.505	0.554	0.637	0.652	0.045	0.115	0.971	0.052	4
rs9266	12	25,253,483	*KRAS*	NA11993	0.634	0.486	0.477	0.657	0.602	0.651	0.027	0.224	0.008	3
rs1047775	13	42,579,426	*DNAJD1*	NA11993	0.449	0.683	0.748	0.481	0.501	0.949	0.002	0.303	0.006	3
				NA12043	0.916	0.922	0.916	0.609	0.637	0.736	0.003	0.631	0.003	3
rs17655	13	102,326,002	*ERCC5*	NA07357	0.699	0.745	0.767	0.470	0.524	0.363	0.050	0.270	0.009	3
rs1061386	13	110,171,368	*ING1*	NA12043	0.417	0.678	0.654	0.515	0.496	0.381	0.080	0.349	0.020	4
rs10748	16	52,062,221	*RBL2*	NA12043	0.475	0.292	0.394	0.428	0.462	0.199	0.102	0.956	0.054	3
rs2239359	16	88,376,980	*FANCA*	NA12145	0.266	0.124	0.132	0.387	0.362	0.003	0.247	0.876	0.097	3
rs1053004	17	37,719,617	*STAT3*	NA12815	0.632	0.344	0.345	0.223	0.215	0.044	0.223	0.982	0.103	4
rs16942	17	38,497,525	*BRCA1*	NA12043	0.466	0.531	0.465	0.451	0.315	0.223	0.069	0.971	0.039	3

rs1042489	17	73,731,800	*BIRC5*	NA12761	0.542	0.753	0.668	0.480	0.522	0.229	0.127	0.970	0.072	3
rs1053474	18	20,287,258	*IMPACT*	NA12815	0.487	0.191	0.327	0.465	0.442	0.107	0.194	0.915	0.089	3
rs6567211	18	51,046,528	*TCF4*	NA11993	0.401	0.204	0.385	0.465	0.493	0.181	0.184	0.940	0.095	3
				NA12155	0.667	0.336	0.410	0.625	0.619	0.051	0.219	0.943	0.097	4
rs7935	19	10,966,607	*SMARCA4*	NA12155	0.557	0.162	0.298	0.354	0.455	0.089	0.215	0.963	0.102	3
rs6554	19	18,546,963	*UBA52*	NA07357	0.825	0.359	0.538	0.767	0.768	0.758	0.014	0.526	0.014	3
rs1406	19	35,006,951	*CCNE1*	NA12145	0.503	0.461	0.482	0.293	0.342	0.170	0.145	0.934	0.073	3
				NA12155	0.527	0.674	0.456	0.273	0.312	0.179	0.228	0.952	0.118	3
				NA12761	0.528	0.513	0.547	0.383	0.399	0.101	0.238	0.929	0.110	3
rs16997057	20	1,866,486	*PTPNS1*	NA12003	0.122	0.119	0.158	0.305	0.306	0.048	0.185	0.938	0.082	4
rs8156	20	3,734,636	*CDC25B*	NA12145	0.398	0.471	0.547	0.625	0.627	0.530	0.034	0.146	0.005	3
rs7121	20	56,912,201	*GNAS*	NA12815	0.351	0.450	0.314	0.304	0.242	0.280	0.109	0.935	0.064	3
rs5030612	22	22,497,512	*SMARCB1*	NA12043	0.655	0.300	0.285	0.490	0.462	0.020	0.213	0.802	0.078	3
rs5958343	X	122,873,952	*BIRC4*	NA12145	0.313	0.244	0.286	0.369	0.390	0.039	0.115	0.977	0.052	4
				NA12156	0.772	0.890	0.873	0.516	0.512	0.004	0.179	0.932	0.075	4
				NA12815	0.751	0.872	0.836	0.578	0.564	0.005	0.283	0.929	0.118	4

[1] Statistical significance between the two light fractions (pre-80S and light polysome) and the two heavy fractions (medium polysome and heavy polysome); [2] The difference of the allele proportion between the two light fractions and the two heavy fractions; [3] The correlation coefficient r of allele proportions from pre-80S to heavy polysome; [4] the slope of the line of allele proportions from pre-80S to heavy polysome.

5.7 Confirmation of the *Illumina* results by resequencing

To confirm the *Illumina* beadarray results, I selected 6 genes with the TI score ≥3 (Table 4) and repeated the TI assay by resequencing. The resequencing results confirmed the TI results on the *Illumina* beadarray. Different TI patterns were seen in these genes, and reflected different types of effects on gene translation.

Table 4 The PCR primers for the confirmation of the *Illumina* beadarray results by resequencing

SNP	Gene	Primers for cDNA
rs10593	*ITGB1BP1*	L: 5'- TTTTAGCATGAGCGGTAATCTTT -3';
		R: 5'- tgaggaagactgaggaCATGG -3'.
rs2953	*CTNNB1*	L: 5'- GGTGGGCTGGTATCTCAGaa -3';
		R: 5'- agagctacttcaaagcaagcaaa -3'.
rs16997057	*PTPNS1*	L: 5'- CCATCTCTACGCGCTTTCTT -3';
(rs3197744)		R: 5'- aggggaggtgggattttct -3'.
rs1406	*CCNE1*	L: 5'- AGTACACCAGCCACCTCCAG -3';
		R: 5'- GCAGCACTTACaaaacagttcatc -3'.
rs947894	*GSTP1*	L: 5'- GGAGACCTCACCCTGTACCA -3';
(rs1695)		R: 5'- GGACAGCAGGGTCTCAAAAG -3'.
rs1053474	*IMPACT*	L: 5'- GATGggaattgcagaagctg -3';
		R: 5'- accatttttatggaaagaaaaataacc -3'.

5.7.1 Opposite AI direction in the light polysome fraction and heavy polysome fraction: the case of *ITGB1BP1* (Fig.9) and *CTNNB1* (Fig.10). The allele abundant in the light polysome fraction has lower translational efficiency, while the other allele abundant in the heavy fraction has higher translational efficiency.

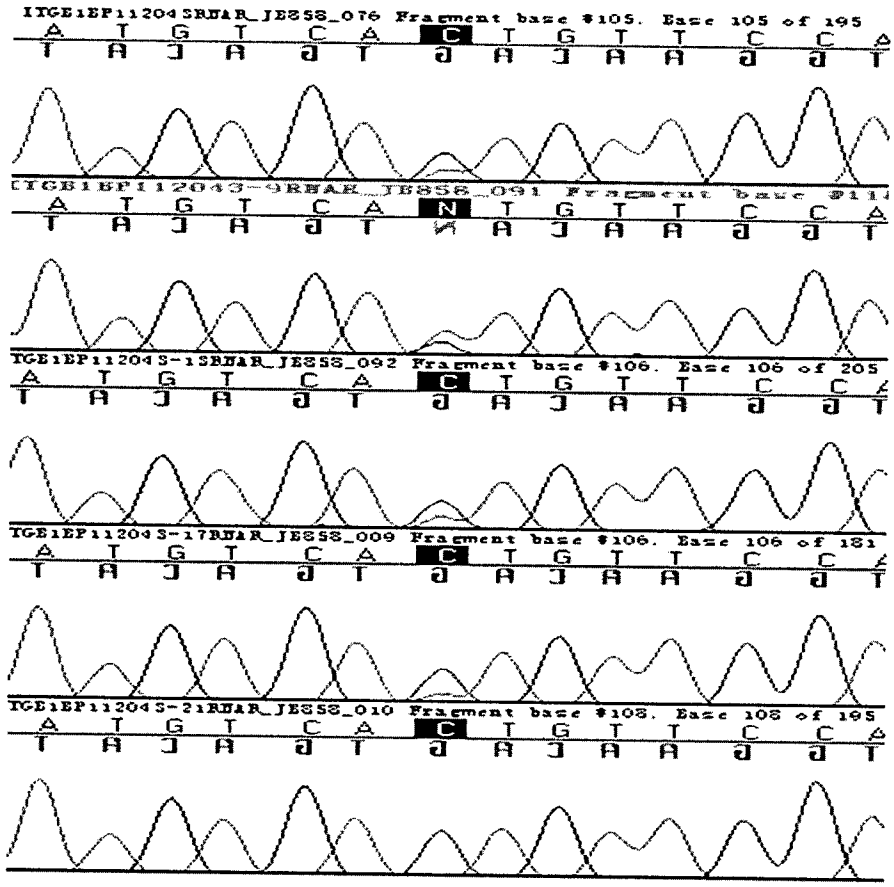

Fig.9. The TI of the two alleles of the *ITGB1BP1* gene. We can see that the T allele is higher in the light polysome fraction, but lower in the heavy polysome fraction, which suggests that the T allele has lower translational efficiency.

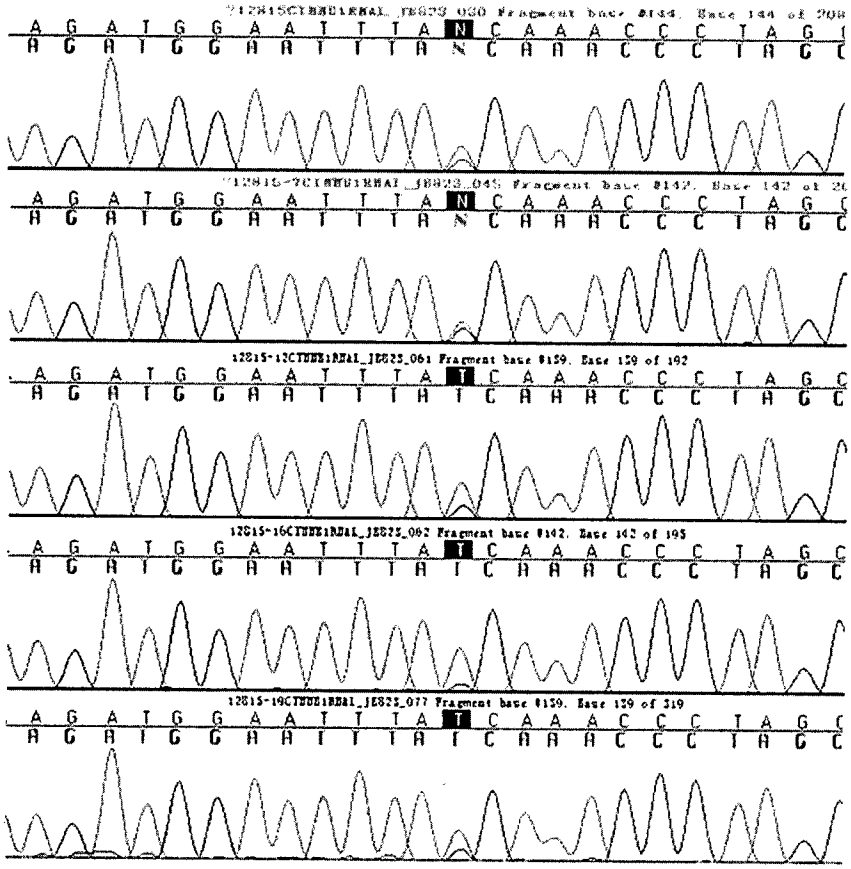

Fig.10. The TI of the two alleles of the *CTNNB1* gene. Referring to the unfractionated RNA (F0), we can see that the G allele proportion is higher in the light polysome fraction, but lower in the heavy polysome fraction, which suggests that the G allele has lower translational efficiency.

5.7.2 Abundant allele in the heavy fraction: the case of *PTPNS1*(Fig.11). The abundance of the allele in the heavy fraction suggests that this allele has higher translational efficiency. If only a small proportion of mRNA went to the heavy fractions, the allele ratio in the light polysome fraction may not have obvious difference.

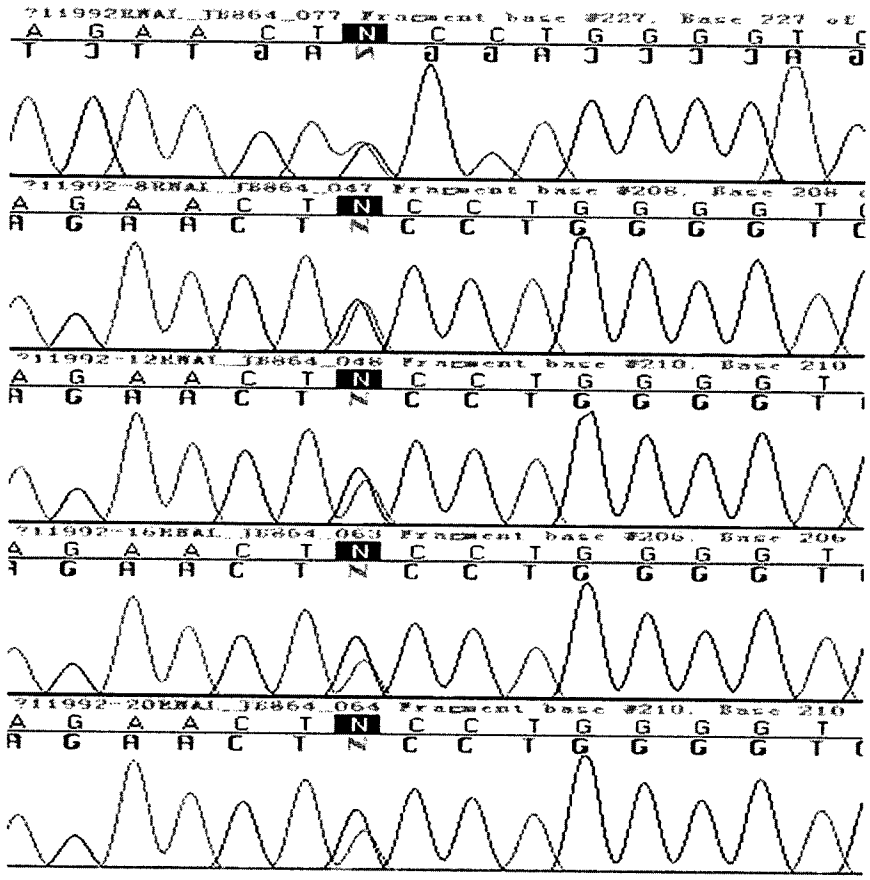

Fig.11. The TI of the two alleles of the *PTPNS1* gene. Referring to the unfractionated RNA (F0), we can see that the G allele has similar proportion with F0 in the light polysome fraction, but higher in the heavy polysome fraction, which suggests the G allele has higher translational efficiency.

5.7.3 Abundant allele in the light fraction: the cases of *CCNE1*(Fig.12) and *GSTP1* (Fig.13) The similar ratio of the two alleles in the heavy polysome fraction suggests that both alleles are actively translated with similar efficiency. The abundance of the allele in the light fraction suggests that this allele may have higher stability in the free mRNA status. If the major proportion of the gene mRNA goes to the heavy polysome fractions, the effect of free mRNA stability on the total mRNA level can be small.

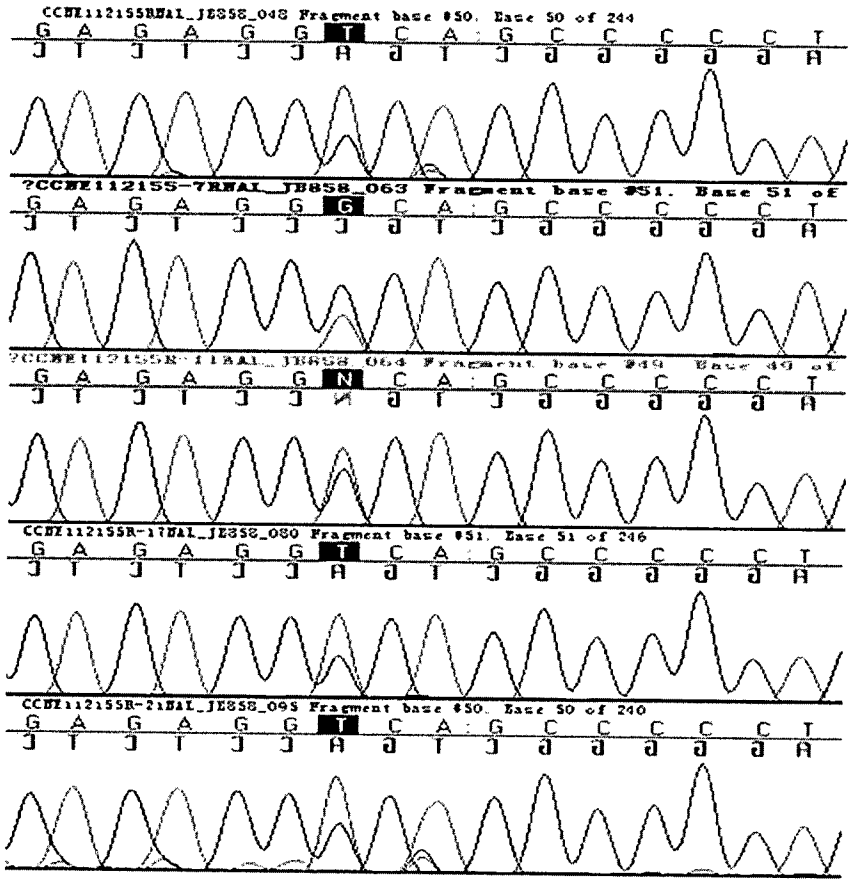

Fig.12. The TI of the two alleles of the *CCNE1* gene. Referring to the unfractionated RNA (F0), we can see that the G allele is much more abundant in the light polysome fraction than in the heavy polysome fraction.

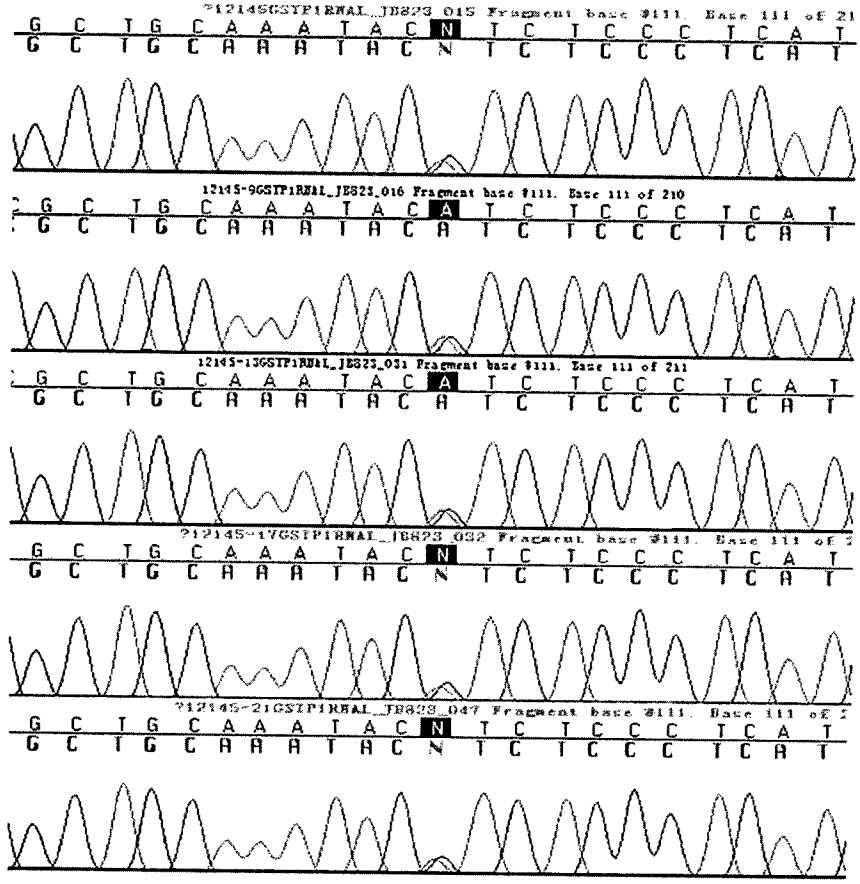

Fig.13. The TI of the two alleles of the *GSTP1* gene. Referring to the unfractionated RNA (F0), we can see that the A allele has similar proportion with F0 in the heavy polysome fraction, but higher in the light polysome fraction.

5.7.4 The detection of TI without referable unfractionated RNA: the cases of *IMPACT* (Fig.14). In this case, the unfractionated RNA has the mixture of un-translated isoforms. We cannot correct the polysome fractions using F0. Instead, I calculated the allele ratio using neighboring nucleotides around the SNP site.

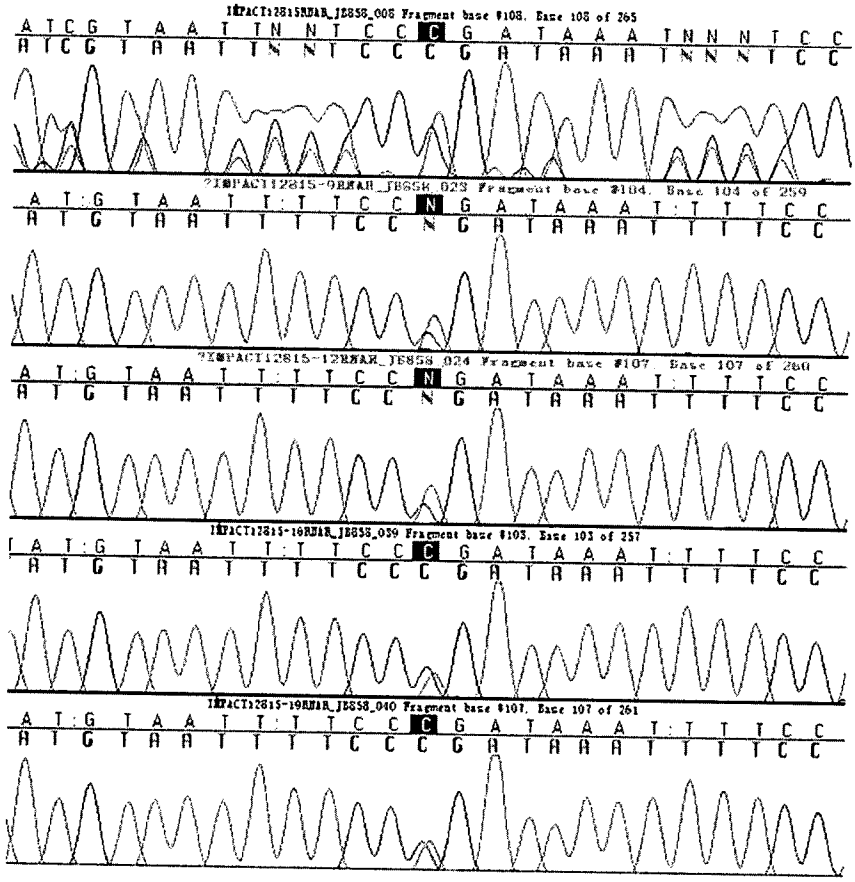

Fig.14. The TI of the two alleles of the *IMPACT* gene. We can see that the T allele proportion is higher in the light polysome fraction, but lower in the heavy polysome fraction, which suggests the T allele has lower translational efficiency.

6. DISCUSSION

6.1 A novel approach to cover an aspect of functional genomics

Because of the rapid development of genotyping technology and availability of high throughput genotyping microarray (www.affymetrix.com, Affymetrix, Santa Clara, CA; www.*Illumina*.com, *Illumina*, San Diego CA), a large number of disease-associated loci have been found in the past one or two years (Duerr, Taylor et al. 2006; Easton, Pooley et al. 2007; Saxena, Voight et al. 2007; Sladek, Rocheleau et al. 2007; Wellcome Trust Case Control Consortium 2007). However, many disease loci have extended linkage disequilibrium for more than a few hundred kb, and may include several genes, as shown in the discovery of a novel T1D locus by our group (Hakonarson, Qu et al. 2008). There may be multiple variants in multiple genes in the same LD block that can explain the association because of LD. It is difficult to clarify the disease causative gene in a locus. When the limits of genetic dissection have been reached, functional studies must be undertaken. The approach to identify genes with functional genetic variation(s) is an important supplement to locate a disease gene. For a gene without functional variation, it cannot be the cause of a genetic disease. On the other hand, if a gene is identified to have functional DNA variation, its potential role in a genetic disease will be highlighted.

Better understanding of the genetic regulation of gene function is likely to be valuable in devising effective disease prevention and risk prediction. To date, studies on the function of DNA variations have focused on two major directions: (1) how an amino acid substitution caused by a nsSNP changes the protein function (Smyth, Cooper et al.

2006; Burke, Worth et al. 2007); (2) how a regulatory SNP change the mRNA levels (Cheung, Spielman et al. 2005; Kiekens, Vercauteren et al. 2006). However, the DNA variations with translational effect will likely neither change amino acid nor affect transcription levels. Because no high throughput method was available to study the genetic effect on gene translation, the effect on gene translation of DNA variations has been poorly understood. The results of this study suggest that genetic effect on gene translation is a common mechanism of gene expression regulation, which needs further study. My study creates a novel additional approach to cover an aspect of functional genomics that has not received much attention so far. Therefore, this study adds an important tool in the evaluation of loci associated with complex disorders.

6.2 The exclusion of non-causative SNPs

Our study suggests that as many as 23.4% genes have the change of gene translational efficiency by genetic variations (more extensive confirmation pending), which suggests that the translational effect of DNA variations is a common phenomenon (Table 3). It is worth noting that the SNPs used for the TI assay on the *Illumina* Golden Gate beadarray, or the resequencing, are only genetic markers. A real causative SNP that affects gene translation change can be located anywhere in the gene transcription region. There is no simple way to test the translational effect of a SNP. However, a SNP can be excluded to have translational effect explicitly by our high throughput method. If a SNP is the cause of the change of gene translation, each of the two alleles should always correspond to a specific change in different samples, i.e. either increased or decreased in all the different

samples, but cannot be increased in some samples and decreased in some other samples. We will take the case of the *ITGB1BP1* gene to explain this approach.

The *ITGB1BP1* gene maps to Chr2p25.2, and encodes the integrin beta 1 binding protein 1, which may play important roles in integrin-dependent cell adhesion (Chang, Wong et al. 1997). This gene is highly expressed in lymphocytes (Fig.15). In our study, *ITGB1BP1* was found to have strong evidence of genetic effect on translation by a common polymorphism. Numerous SNPs have been found in the mRNA. However, most SNPs in the mRNA are found in the 3'UTR (Fig.16). Only one nsSNP and one sSNP are found in the coding region. Altogether, only two 3'UTR SNPs have common frequencies, i.e. rs10593 (minor allele G, frequency=0.175 in European) and rs4798 (minor allele A, frequency=0.125 in European). All other SNPs in the mRNA are rare (minor allele frequency <0.05). Therefore, one of these two SNPs, rs10593 and rs4798, may be TI causative. Both 3'UTR SNPs rs10593 and rs4798 have been used for TI marker in our study. The two SNPs are only 775bp from each other. However, the poor linkage disequilibrium (LD) between the two SNPs (Fig.17) enables us to identify which SNP is non-causative. The details of the methodology will be shown in the following paragraphs.

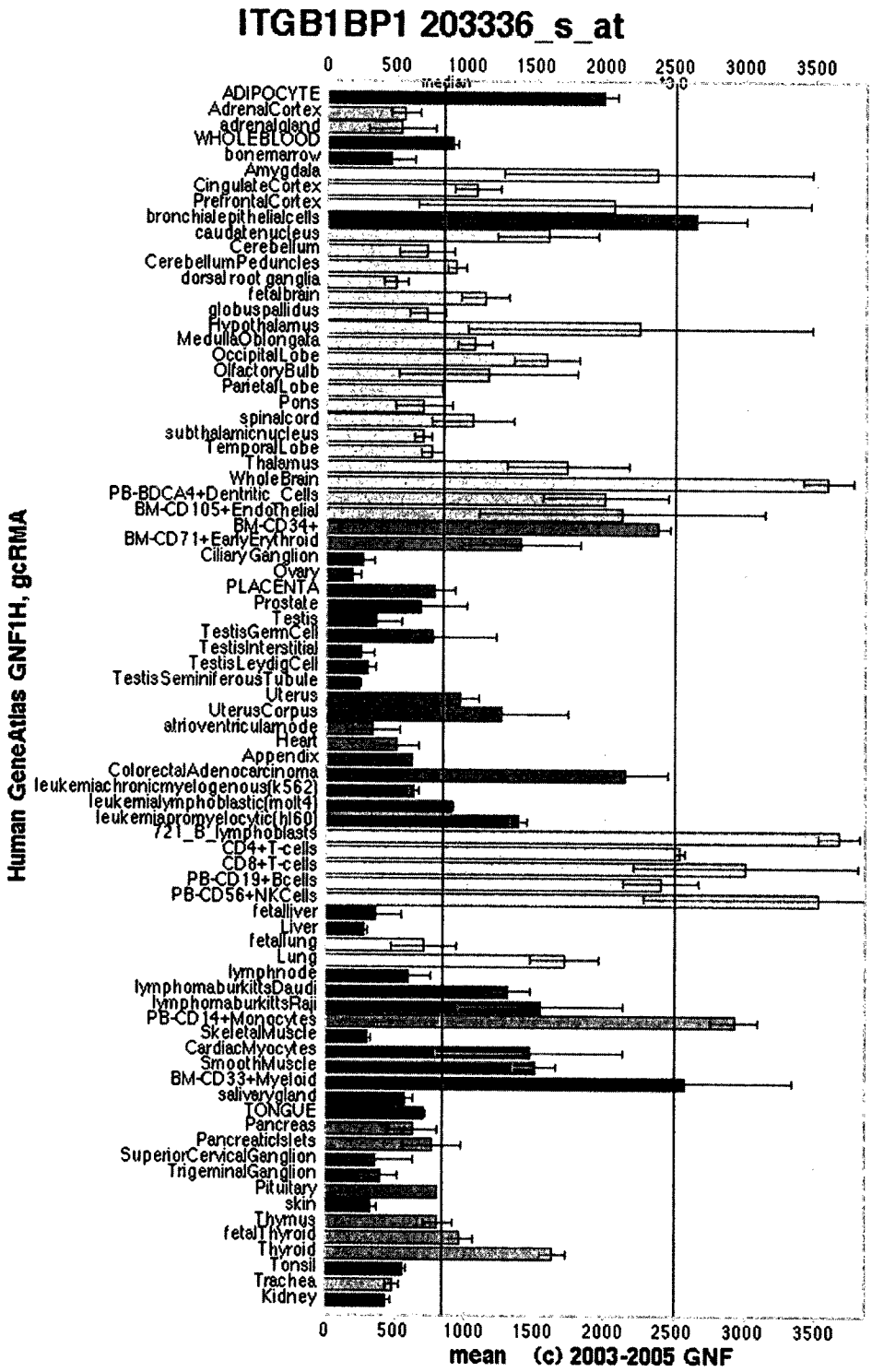

Fig. 15. The *ITGB1BP1* gene expression profile as shown by the SymAtlas database (http://symatlas.gnf.org/SymAtlas/).

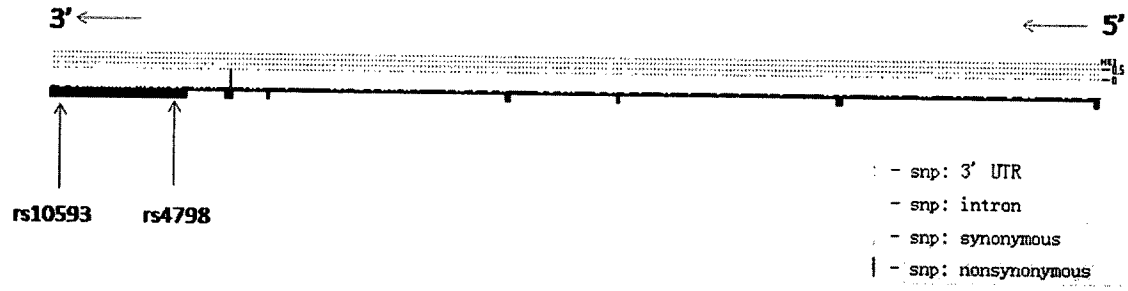

Fig. 16. The SNPs in the *ITGB1BP1* gene region (the NCBI dbSNP database, http://www.ncbi.nlm.nih.gov/SNP).

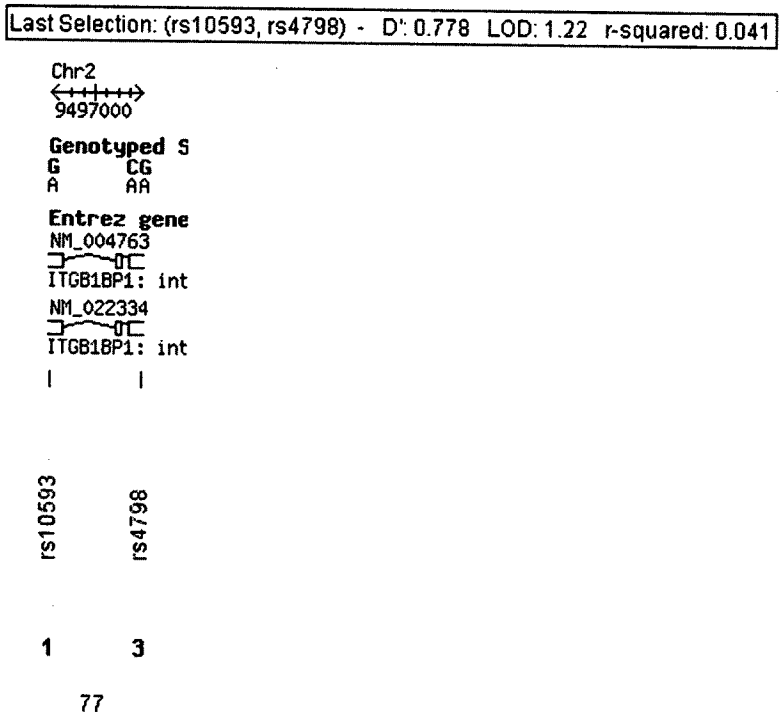

Fig. 17. The LD of the two *ITGB1BP1* SNPs based on the European HapMap data (http://www.hapmap.org) (2003; 2007), produced by Haploview v4.0 software(www.broad.mit.edu/personal/jcbarret/haploview) (Barrett, Fry et al. 2005). D' value (%) is shown in the box. D'<100% suggests the existence of recombination between the two SNPs.

Each of the two SNPs rs10593 (A/G) and rs4798 (A/G) have two alleles. Therefore, four haplotypes are expected in general population, i.e. rs10593 A- rs4798 A, rs10593 A- rs4798 G, rs10593 G- rs4798 A, and rs10593 G- rs4798 G. Each haplotype represents a type of chromosome, tagged by a rs10593 allele and a rs4798 allele. If there is no 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 (Boehnke 2000). A major factor of LD is the physical proximity of the two SNPs. There are two criteria to describe the LD, i.e. D' and r^2. When there is no recombination between two SNPs, we can only see three of the four haplotypes at most. In this condition, D'=1. In addition to no recombination (D'=1), when the frequencies of the two SNPs are exactly the same, we can only see two haplotypes (r^2=1), i.e. one allele of one SNP always co-occurs with a specific allele of the other SNP. If r^2=1, we cannot exclude one SNP as TI causative based on LD. In the case of rs10593 and rs4798, this approach is feasible because of r^2=0.041. The expected frequency and empirical frequency of each haplotype is shown in Fig.18.

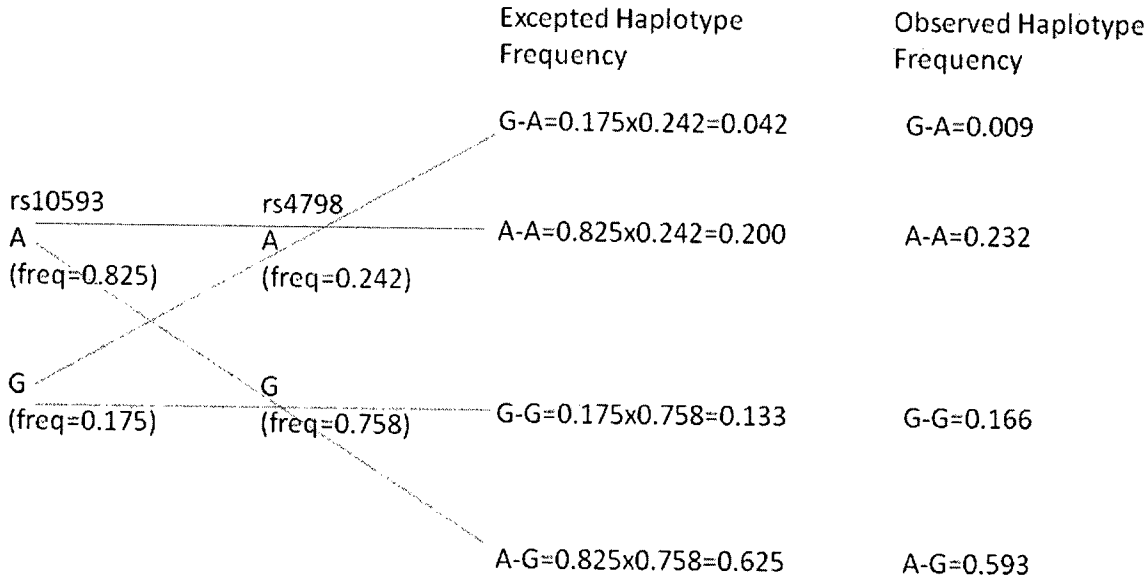

Fig. 18. The four haplotypes of the *ITGB1BP1* gene. Because of LD between the two SNPs, the two haplotypes A-G and G-G have lower frequencies than expected, and the other two haplotypes have higher frequencies than expected.

Of the 10 HapMap CEU LCL samples that were run in our *Illumina* Golden Gate TI assay, there are 3 heterozygotes of rs10593 (NA07357, NA12043, and NA12761), and 3 heterozygotes of rs4798 (NA11993, NA12043, and NA12815). For the 3 heterozygotes of rs10593, dramatic TI can be seen in NA12043, i.e. the G allele of NA12043 has higher translation. However, no TI was seen in either NA07357, or NA12761. On the other hand, all three heterozygotes of rs4798 showed TI in the same direction, i.e. the minor A allele has lower translational efficiency. These results excluded rs10593 as TI causative. Because there are only these two common SNPs in the gene region, rs4798 is the candidate causing the translational effect. When I looked back at the rs10593 results, both the two samples NA07357 and NA12761 without TI have GG genotype of rs4798. For

the sample NA12043, the TI of rs10593 is concordant with the TI of rs4798 (Table 5). Therefore, none of these data is against rs4798 of the translational effect and the effect is likely due to it, or to an unknown SNP in tight linkage disequilibrium with it.

Table 5 The genotypes and haplotypes of the *ITGB1BP1* SNPs

sample	rs10593	rs4798	**rs10593-rs4798**
NA07357	A/G	G/G	A-G/G-G
NA12043	A/G	A/G	A-A<G-G
NA12761	A/G	G/G	A-G/G-G
NA11993	A/A	A/G	A-A<A-G
NA12815	A/A	A/G	A-A<A-G

6.3 The analysis of the TI mechanism

As explained in the introduction part of my thesis, DNA variations may cause TI by many potential mechanisms. A 5'UTR SNP located at an uAUG, an uORF, or a conservative IRE sequence, can be expected to have a translational effect. For the 3'UTR SNP rs4798, the conserved RNA sequence motif UCPAS, with the consensus hexamer sequence AAUAAA, CPE with the consensus UUUUUAU, EDEN with the consensus sequence of a repetition of U(A/G) dinucleotides, are not involved.

6.3.1 The miRNA target site

Besides the UCPAS, the 3'UTR of genes contains miRNA target sites (Du and Zamore 2007). MicroRNAs (miRNAs) are highly conserved, small RNA molecules (around 21 nucleotides) that regulate the expression of genes by binding to the complementary mRNAs (Ambros 2004). Up to now, 3518 miRNAs have been identified in different organisms, and 326 miRNAs have been found in human (miRBase release 7, http://microrna.sanger.ac.uk/sequences/help/summary.shtml). One miRNA can control the expression of thousands of target mRNAs (Lim, Lau et al. 2005). On the other hand, one mRNA moleculer may be targeted by multiple microRNAs (Pillai 2005). In recent years, miRNAs have received extensive attention and the list of reported miRNA functions is growing rapidly. Present studies have shown the roles of miRNAs on development (Alvarez-Garcia and Miska 2005), viral infection (Sullivan and Ganem 2005), and oncogenesis (Hammond 2006).

To target at a mRNA molecular, the seed region of the miRNA (nucleotides 2–8) needs to bind with a complementary conserved sequence at the mRNA 3'UTR (Lewis, Burge et al. 2005). A number of bioinformatics tools can be used to predict the miRNA target sites, such as PicTar (http://pictar.bio.nyu.edu/)(Krek, Grun et al. 2005), TargetScan (http://www.targetscan.org/) (Lewis, Burge et al. 2005), and microInspector (http://www.imbb.forth.gr/microinspector) (Rusinov, Baev et al. 2005). Using these tools, we did not find any miRNA target sites involving rs4798. For the other genes found to have evidence of TI, further mapping of candidate SNPs with translational effect is needed for the purpose of the prediction of miRNA target sites.

6.3.2 The mRNA secondary structure change

The change of mRNA secondary structure caused by a SNP may influence the gene translation by multiple mechanisms. A SNP in the 5'UTR may change the mRNA secondary structure, and hamper the 40S subunit scanning for the AUG start codon(Lee, Guertin et al. 1983; de Smit and van Duin 1994). The change of mRNA secondary structure by a SNP in the gene coding region may block the sliding of the 80S ribosome along the mRNA molecular for peptide elongation (Shen, Basilion et al. 1999). The change of mRNA secondary structure by a SNP in the 3'UTR region may influence the binding of miRNAs and PABP, thus impair the translational efficiency. We will look at the case of the *ITGB1BP1* gene as an example.

The 3'UTR SNP rs4798 is located 158 bp downstream of the coding region of the *ITGB1BP1* mRNA NM_004763. This SNP does not map to any known regulatory motif. The functional effect of this SNP may be from its effect on mRNA secondary structure. Changes in the minimum free energy (MFE, is the goal of mRNA secondary structure prediction) of the mRNA can be used to determine the effect of rs4798 on the mRNA secondary structure *in silico*. MFE of full-length mRNA is computed on the basis of an energy minimization algorithm (Zuker and Stiegler 1981), by the Vienna RNA Package (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (Hofacker 2003). As illustrated in Fig.19, we can see that the 3'UTR SNP rs4798 may have an obvious effect on the mRNA MFE and the secondary structure by the *in silico* prediction. Therefore, in the absence of any other known mechanism, it is likely that the TI observed in the *ITGB1BP1* gene is due to effects on the mRNA secondary structure, a possibility that needs to be explored further.

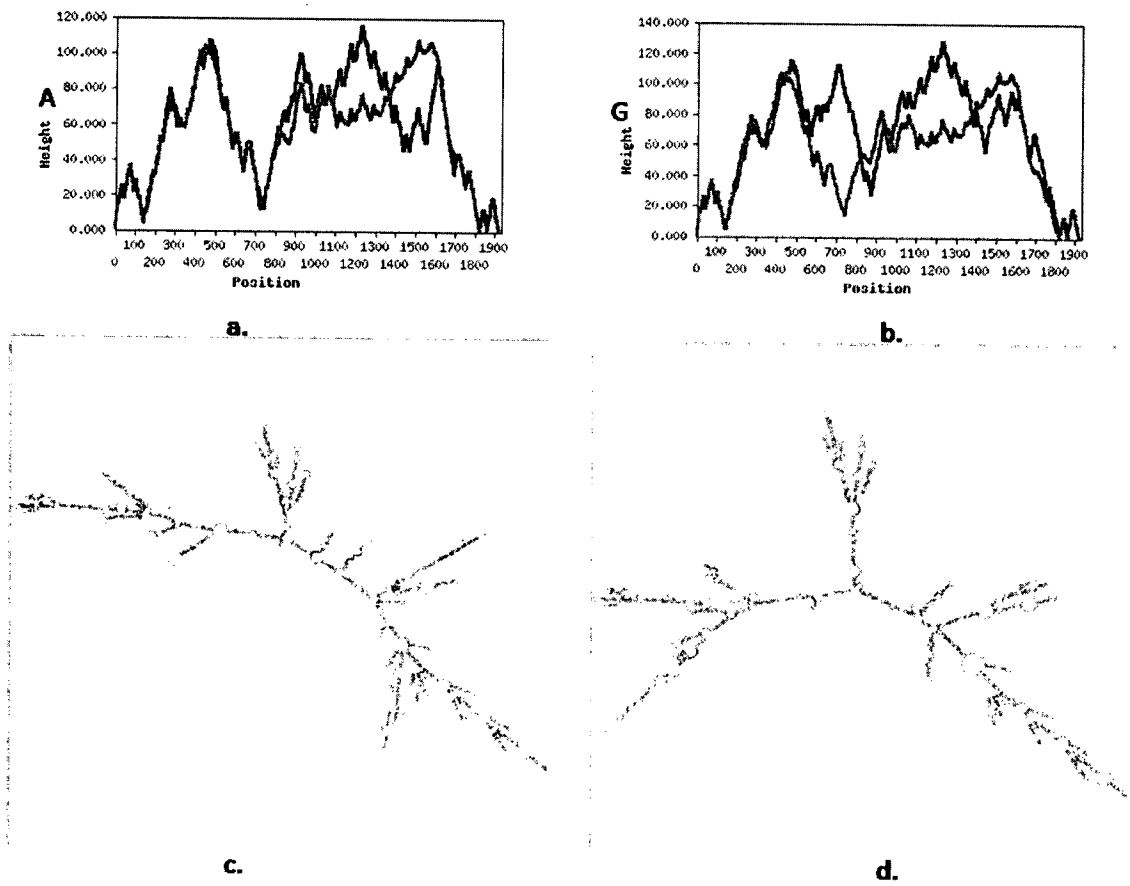

Fig. 19. The mRNA secondary structure change of the *ITGB1BP1* transcript NM_004763 by the 3'UTR SNP rs4798. The two upper panels are the plot of the minimum free energy and pair probabilities. The red line is the minimum free energy, and the green line is the pair probabilities. The two lower panels are the predicted mRNA secondary structure. The two left panels correspond to the A allele of the *ITGB1BP1* mRNA, and the two right panels correspond to the G allele of the *ITGB1BP1* mRNA.

7. Conclusion

1. This study creates a novel high throughput method to discover the genetic effect on gene translation. This method covers an aspect of functional genomics that has not received much attention so far, and adds an important tool in the evaluation of genetic loci associated with complex disorders.

2. By the study of 269 genes, we found 23.4% genes with evidence of genetic effect on gene translation, which suggest that genetic effect on gene translation may be a common mechanism of gene expression regulation.

3. Shown by our best hit found in the *ITGB1BP1* gene, the role of the mRNA 3'UTR secondary structure in the genetic control of gene translational may be important.

4. As a fundamental study, this study has established a high-quality RNA bank, including 1,100 RNA fraction samples, for further research work. To broaden the discovery study of genome-wide translational effect, the assay on a high throughput platform, such as the 454 re-sequencing technology, can be performed on these RNA samples, following the mature methodology created in this study. To fine map the causative DNA variations of translation effect, this study will provide the methodological reference for the experimental assay of computationally predicted candidate SNPs, and a sample bank.

5. A biological validation of these findings, by transfecting epitope-tagged constructs of the two cDNA alleles will be needed in future studies.

8. Acknowledgements

I would like to thank all the people for all their great contributions to my project.

My sincere thanks to my supervisor, Dr. Constantin Polychronakos, for mentoring me with his great intelligence and knowledge in human genetics, with his great patience and encouragement.

My gratitude to Dr. Francis Robert, for his tremendous help in establishing the method of polysome RNA fractionation. Thanks to Dr. Tomi Pastinen and Dr. Eef Harmsen for their helpful advices on the high throughput allelic imbalance study. Thanks to Kevin Lam for his technical support.

My gratitude to our super technician, Luc Marchand, for huge tons of technical help and advices from him. My thanks to my teachers and colleagues, Rosemary Grabs for her always available help, and Marylene Rousseau for her help in developing the poly(A) RNA quantification assay. Thanks to Dr. Cindy Goodyer for her help as my academic advisor.

My special thanks to Dr. Huiqi Qu, for his technical advices as a scientist, and his care as my husband. Thanks for my son, Jingchun Qu. Thanks for my family and my husband's family far away in China. They gave me so much support spiritually and financially.

Thanks for the salary support from the Huntley Macdonald Sinclair Fund Studentship of the Faculty of Medicine, McGill University, and the Maria Raiche Graduate Student Scholarship of the Montreal Children's Hospital Research Institute.

Thanks Dr. Hugues Beauchemin for the help of the French translation of the Résumé.

9. Reference

(2003). "The International HapMap Project." Nature **426**(6968): 789-796.

(2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-861.

Allerson, C. R., M. Cazzola and T. A. Rouault (1999). "Clinical Severity and Thermodynamic Effects of Iron-responsive Element Mutations in Hereditary Hyperferritinemia-Cataract Syndrome." J. Biol. Chem. **274**(37): 26439-26447.

Alvarez-Garcia, I. and E. A. Miska (2005). "MicroRNA functions in animal development and human disease." Development **132**(21): 4653-62.

Ambros, V. (2004). "The functions of animal microRNAs." Nature **431**(7006): 350.

Andreson, R., T. Puurand and M. Remm (2006). "SNPmasker: automatic masking of SNPs and repeats across eukaryotic genomes." Nucleic Acids Res **34**(Web Server issue): W651-5.

Anjos, S. M., W. Shao, L. Marchand and C. Polychronakos (2005). "Allelic effects on gene regulation at the autoimmunity-predisposing CTLA4 locus: a re-evaluation of the 3' +6230G>A polymorphism." Genes Immun **6**(4): 305-11.

Anjos, S. M., M. C. Tessier and C. Polychronakos (2004). "Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block." J Clin Endocrinol Metab **89**(12): 6257-65.

Barratt, B. J., F. Payne, C. E. Lowe, R. Hermann, B. C. Healy, D. Harold, P. Concannon, N. Gharani, M. I. McCarthy, M. G. Olavesen, R. McCormack, C. Guja, C. Ionescu-Tirgoviste, D. E. Undlien, K. S. Ronningen, K. M. Gillespie, E. Tuomilehto-Wolf, J. Tuomilehto, S. T. Bennett, D. G. Clayton, H. J. Cordell and J. A. Todd (2004). "Remapping the insulin gene/IDDM2 locus in type 1 diabetes." Diabetes **53**(7): 1884-9.

Barrett, J. C., B. Fry, J. Maller and M. J. Daly (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-5.

Boehnke, M. (2000). "A look at linkage disequilibrium." Nat Genet **25**(3): 246-247.

Burke, D. F., C. L. Worth, E. M. Priego, T. Cheng, L. J. Smink, J. A. Todd and T. L. Blundell (2007). "Genome bioinformatic analysis of nonsynonymous SNPs." BMC Bioinformatics **8**: 301.

Chang, D. D., C. Wong, H. Smith and J. Liu (1997). "ICAP-1, a novel beta1 integrin cytoplasmic domain-associated protein, binds to a conserved and functionally important NPXY sequence motif of beta1 integrin." J Cell Biol **138**(5): 1149-57.

Chen, J. M., C. Ferec and D. N. Cooper (2006). "A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview." Hum Genet **120**(1): 1-21.

Chen, J. M., C. Ferec and D. N. Cooper (2006). "A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants." Hum Genet **120**(3): 301-33.

Cheung, V. G., R. S. Spielman, K. G. Ewens, T. M. Weber, M. Morley and J. T. Burdick (2005). "Mapping determinants of human gene expression by regional and genome-wide association." Nature **437**(7063): 1365-9.

Couzin, J. and J. Kaiser (2007). "GENOME-WIDE ASSOCIATION: Closing the Net on Common Disease Genes." Science **316**(5826): 820-822.

De Angioletti, M., G. Lacerra, V. Sabato and C. Carestia (2004). "Beta+45 G --> C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence." Br J Haematol **124**(2): 224-31.

de Moor, C. H., H. Meijer and S. Lissenden (2005). "Mechanisms of translational control by the 3' UTR in development and differentiation." Semin Cell Dev Biol **16**(1): 49-58.

de Smit, M. H. and J. van Duin (1994). "Control of Translation by mRNA Secondary Structure in Escherichia coli: A Quantitative Analysis of Literature Data." Journal of Molecular Biology **244**(2): 144-150.

Diatchenko, L., G. D. Slade, A. G. Nackley, K. Bhalang, A. Sigurdsson, I. Belfer, D. Goldman, K. Xu, S. A. Shabalina, D. Shagin, M. B. Max, S. S. Makarov and W. Maixner (2005). "Genetic basis for individual variations in pain perception and the development of a chronic pain condition." Hum. Mol. Genet. **14**(1): 135-143.

Dixon, A. L., L. Liang, M. F. Moffatt, W. Chen, S. Heath, K. C. Wong, J. Taylor, E. Burnett, I. Gut, M. Farrall, G. M. Lathrop, G. R. Abecasis and W. O. Cookson (2007). "A genome-wide association study of global gene expression." Nat Genet **39**(10): 1202-7.

Du, T. and P. D. Zamore (2007). "Beginning to understand microRNA function." Cell Res **17**(8): 661-3.

Duerr, R. H., K. D. Taylor, S. R. Brant, J. D. Rioux, M. S. Silverberg, M. J. Daly, A. H. Steinhart, C. Abraham, M. Regueiro, A. Griffiths, T. Dassopoulos, A. Bitton, H. Yang, S. Targan, L. W. Datta, E. O. Kistner, L. P. Schumm, A. T. Lee, P. K. Gregersen, M. M. Barmada, J. I. Rotter, D. L. Nicolae and J. H. Cho (2006). "A genome-wide association

study identifies IL23R as an inflammatory bowel disease gene." Science **314**(5804): 1461-3.

Easton, D. F., K. A. Pooley, A. M. Dunning, P. D. Pharoah, D. Thompson, D. G. Ballinger, J. P. Struewing, J. Morrison, H. Field, R. Luben, N. Wareham, S. Ahmed, C. S. Healey, R. Bowman, K. B. Meyer, C. A. Haiman, L. K. Kolonel, B. E. Henderson, L. Le Marchand, P. Brennan, S. Sangrajrang, V. Gaborieau, F. Odefrey, C. Y. Shen, P. E. Wu, H. C. Wang, D. Eccles, D. G. Evans, J. Peto, O. Fletcher, N. Johnson, S. Seal, M. R. Stratton, N. Rahman, G. Chenevix-Trench, S. E. Bojesen, B. G. Nordestgaard, C. K. Axelsson, M. Garcia-Closas, L. Brinton, S. Chanock, J. Lissowska, B. Peplonska, H. Nevanlinna, R. Fagerholm, H. Eerola, D. Kang, K. Y. Yoo, D. Y. Noh, S. H. Ahn, D. J. Hunter, S. E. Hankinson, D. G. Cox, P. Hall, S. Wedren, J. Liu, Y. L. Low, N. Bogdanova, P. Schurmann, T. Dork, R. A. Tollenaar, C. E. Jacobi, P. Devilee, J. G. Klijn, A. J. Sigurdson, M. M. Doody, B. H. Alexander, J. Zhang, A. Cox, I. W. Brock, G. MacPherson, M. W. Reed, F. J. Couch, E. L. Goode, J. E. Olson, H. Meijers-Heijboer, A. van den Ouweland, A. Uitterlinden, F. Rivadeneira, R. L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, J. L. Hopper, M. McCredie, M. Southey, G. G. Giles, C. Schroen, C. Justenhoven, H. Brauch, U. Hamann, Y. D. Ko, A. B. Spurdle, J. Beesley, X. Chen, A. Mannermaa, V. M. Kosma, V. Kataja, J. Hartikainen, N. E. Day, D. R. Cox and B. A. Ponder (2007). "Genome-wide association study identifies novel breast cancer susceptibility loci." Nature **447**(7148): 1087-93.

Fan, J.-B., M. S. Chee and K. L. Gunderson (2006). "Highly parallel genomic assays." Nat Rev Genet **7**(8): 632-644.

Ge, B., S. Gurd, T. Gaudin, C. Dore, P. Lepage, E. Harmsen, T. J. Hudson and T. Pastinen (2005). "Survey of allelic expression using EST mining." Genome Res **15**(11): 1584-91.

Gebauer, F. and M. W. Hentze (2004). "Molecular mechanisms of translational control." Nat Rev Mol Cell Biol **5**(10): 827-835.

Ghilardi, N., A. Wiestner, M. Kikuchi, A. Ohsaka and R. C. Skoda (1999). "Hereditary thrombocythaemia in a Japanese family is caused by a novel point mutation in the thrombopoietin gene." British Journal of Haematology **107**(2): 310-316.

Gunderson, K. L., F. J. Steemers, G. Lee, L. G. Mendoza and M. S. Chee (2005). "A genome-wide scalable SNP genotyping assay using microarray technology." Nat Genet **37**(5): 549-54.

Hakonarson, H., S. F. Grant, J. P. Bradfield, L. Marchand, C. E. Kim, J. T. Glessner, R. Grabs, T. Casalunovo, S. P. Taback, E. C. Frackelton, M. L. Lawson, L. J. Robinson, R. Skraban, Y. Lu, R. M. Chiavacci, C. A. Stanley, S. E. Kirsch, E. F. Rappaport, J. S. Orange, D. S. Monos, M. Devoto, H. Q. Qu and C. Polychronakos (2007). "A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene." Nature **448**(7153): 591-4.

Hakonarson, H., H. Q. Qu, J. P. Bradfield, L. Marchand, C. E. Kim, J. T. Glessner, R. Grabs, T. Casalunovo, S. P. Taback, E. C. Frackelton, A. W. Eckert, K. Annaiah, M. L. Lawson, F. G. Otieno, E. Santa, J. L. Shaner, R. M. Smith, C. C. Onyiah, R. Skraban, R. M. Chiavacci, L. J. Robinson, C. A. Stanley, S. E. Kirsch, M. Devoto, D. S. Monos, S. F. Grant and C. Polychronakos (2008). "A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study." Diabetes **57**(4): 1143-6.

Hammond, S. M. (2006). "MicroRNAs as oncogenes." Curr Opin Genet Dev **16**(1): 4-9.

Hentze, M. W. and L. C. Kuhn (1996). "Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress." Proceedings of the National Academy of Sciences **93**(16): 8175-8182.

Heward, J. M., M. J. Simmonds, J. Carr-Smith, H. Foxall, J. A. Franklyn and S. C. Gough (2004). "A single nucleotide polymorphism in the CD40 gene on chromosome 20q (GD-2) provides no evidence for susceptibility to Graves' disease in UK Caucasians." Clin Endocrinol (Oxf) **61**(2): 269-72.

Hofacker, I. L. (2003). "Vienna RNA secondary structure server." Nucleic Acids Res **31**(13): 3429-31.

Hsiao, J. Y., K. J. Tien, C. T. Hsiao and M. C. Hsieh (2008). "A C/T Polymorphism in CD40 Gene is Not Associated with Susceptibility and Phenotype of Graves' Disease in Taiwanese." Endocr J.

Hudson, T. J. (2003). "Wanted: regulatory SNPs." Nat Genet **33**(4): 439-40.

Human Genome Sequencing, C. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-945.

Jaffrey, S. R., D. J. Haile, R. D. Klausner and J. B. Harford (1993). "The interaction between the iron-responsive element binding protein and its cognate RNA is highly dependent upon both RNA sequence and structure." Nucleic Acids Research **21**(19): 4627-4631.

Kahvejian, A., Y. V. Svitkin, R. Sukarieh, M. N. M'Boutchou and N. Sonenberg (2005). "Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms." Genes Dev **19**(1): 104-13.

Kamb, A., N. A. Gruis, J. Weaver-Feldhaus, Q. Liu, K. Harshman, S. V. Tavtigian, E. Stockert, R. S. Day, 3rd, B. E. Johnson and M. H. Skolnick (1994). "A cell cycle regulator potentially involved in genesis of many tumor types." Science **264**(5157): 436-40.

Kidd, J. M., G. M. Cooper, W. F. Donahue, H. S. Hayden, N. Sampas, T. Graves, N. Hansen, B. Teague, C. Alkan, F. Antonacci, E. Haugen, T. Zerr, N. A. Yamada, P. Tsang, T. L. Newman, E. Tuzun, Z. Cheng, H. M. Ebling, N. Tusneem, R. David, W. Gillett, K. A. Phelps, M. Weaver, D. Saranga, A. Brand, W. Tao, E. Gustafson, K. McKernan, L.

Chen, M. Malig, J. D. Smith, J. M. Korn, S. A. McCarroll, D. A. Altshuler, D. A. Peiffer, M. Dorschner, J. Stamatoyannopoulos, D. Schwartz, D. A. Nickerson, J. C. Mullikin, R. K. Wilson, L. Bruhn, M. V. Olson, R. Kaul, D. R. Smith and E. E. Eichler (2008). "Mapping and sequencing of structural variation from eight human genomes." Nature **453**(7191): 56-64.

Kiekens, R., A. Vercauteren, B. Moerkerke, E. Goetghebeur, H. Van Den Daele, R. Sterken, M. Kuiper, F. van Eeuwijk and M. Vuylsteke (2006). "Genome-wide screening for cis-regulatory variation using a classical diallel crossing scheme." Nucleic Acids Res **34**(13): 3677-86.

Kikuchi, M., T. Tayama, H. Hayakawa, I. Takahashi, H. Hoshino and A. Ohsaka (1995). "Familial thrombocytosis." Br J Haematol **89**(4): 900-2.

Kozak, M. (1986). "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes." Cell **44**(2): 283-292.

Kozak, M. (1991). "An analysis of vertebrate mRNA sequences: intimations of translational control." J Cell Biol **115**(4): 887-903.

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 Journal **16**(9): 2482-2492.

Kozak, M. (2003). "Alternative ways to think about mRNA sequences and proteins that appear to promote internal initiation of translation." Gene **318**(1-2): 1-23.

Krek, A., D. Grun, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel and N. Rajewsky (2005). "Combinatorial microRNA target predictions." Nat Genet **37**(5): 495-500.

Kuersten, S. and E. B. Goodwin (2003). "The power of the 3' UTR: translational control and development." Nat Rev Genet **4**(8): 626-37.

Kwan, T., D. Benovoy, C. Dias, S. Gurd, C. Provencher, P. Beaulieu, T. J. Hudson, R. Sladek and J. Majewski (2008). "Genome-wide analysis of transcript isoform variation in humans." Nat Genet **40**(2): 225-31.

Kwan, T., D. Benovoy, C. Dias, S. Gurd, D. Serre, H. Zuzan, T. A. Clark, A. Schweitzer, M. K. Staples, H. Wang, J. E. Blume, T. J. Hudson, R. Sladek and J. Majewski (2007). "Heritability of alternative splicing in the human genome." Genome Res **17**(8): 1210-8.

Lee, K. A., D. Guertin and N. Sonenberg (1983). "mRNA secondary structure as a determinant in cap recognition and initiation complex formation. ATP-Mg2+ independent cross-linking of cap binding proteins to m7I-capped inosine-substituted reovirus mRNA." J Biol Chem **258**(2): 707-10.

Leibold, E. A. and B. Guo (1992). "Iron-dependent regulation of ferritin and transferrin receptor expression by the iron-responsive element binding protein." Annu Rev Nutr **12**: 345-68.

Lewis, B. P., C. B. Burge and D. P. Bartel (2005). "Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets." Cell **120**(1): 15-20.

Lim, L. P., N. C. Lau, P. Garrett-Engele, A. Grimson, J. M. Schelter, J. Castle, D. P. Bartel, P. S. Linsley and J. M. Johnson (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature **433**(7027): 769-73.

Liu, L., D. Dilworth, L. Gao, J. Monzon, A. Summers, N. Lassam and D. Hogg (1999). "Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma." Nat Genet **21**(1): 128-132.

Mannisto, P. T. and S. Kaakkola (1999). "Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology, and Clinical Efficacy of the New Selective COMT Inhibitors." Pharmacol Rev **51**(4): 593-628.

McCallum, S. A. and A. Pardi (2003). "Refined Solution Structure of the Iron-responsive Element RNA Using Residual Dipolar Couplings." Journal of Molecular Biology **326**(4): 1037-1050.

Menzaghi, C., G. Paroni, C. De Bonis, T. Soccio, A. Marucci, S. Bacci and V. Trischitta (2006). "The -318 C>G single-nucleotide polymorphism in GNAI2 gene promoter region impairs transcriptional activity through specific binding of Sp1 transcription factor and is associated with high blood pressure in Caucasians from Italy." J Am Soc Nephrol **17**(4 Suppl 2): S115-9.

Merrick, W. C. (1992). "Mechanism and regulation of eukaryotic protein synthesis." Microbiol. Mol. Biol. Rev. **56**(2): 291-315.

Muckenthaler, M., N. K. Gray and M. W. Hentze (1998). "IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F." Mol Cell **2**(3): 383-8.

Nackley, A. G., S. A. Shabalina, I. E. Tchivileva, K. Satterfield, O. Korchynskyi, S. S. Makarov, W. Maixner and L. Diatchenko (2006). "Human Catechol-O-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure." Science **314**(5807): 1930-1933.

Orkin, S. H., T. C. Cheng, S. E. Antonarakis and H. H. Kazazian, Jr. (1985). "Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene." Embo J **4**(2): 453-6.

Paillard, L., F. Omilli, V. Legagneux, T. Bassez, D. Maniey and H. B. Osborne (1998). "EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos." Embo J **17**(1): 278-87.

Pastinen, T., B. Ge, S. Gurd, T. Gaudin, C. Dore, M. Lemire, P. Lepage, E. Harmsen and T. J. Hudson (2005). "Mapping common regulatory variants to human haplotypes." Hum Mol Genet **14**(24): 3963-71.

Pastinen, T. and T. J. Hudson (2004). "Cis-acting regulatory variation in the human genome." Science **306**(5696): 647-50.

Pattengale, P. K., R. W. Smith and P. Gerber (1973). "Selective transformation of B lymphocytes by E.B. virus." Lancet **2**(7820): 93-4.

Paulin, F. E. M., S. A. Chappell and A. E. Willis (1998). "A single nucleotide change in the c-myc internal ribosome entry segment leads to enhanced binding of a group of protein factors." Nucl. Acids Res. **26**(13): 3097-3103.

Pelletier, J. and N. Sonenberg (1988). "Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA." Nature **334**(6180): 320-5.

Pestova, T. V. and C. U. T. Hellen (2006). "Translation, interrupted." Nat Struct Mol Biol **13**(2): 98-99.

Pestova, T. V. and V. G. Kolupaeva (2002). "The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection." Genes Dev. **16**(22): 2906-2922.

Pillai, R. S. (2005). "MicroRNA function: Multiple mechanisms for a tiny RNA?" RNA **11**(12): 1753-1761.

Pirastu, M., R. Galanello, M. A. Doherty, T. Tuveri, A. Cao and Y. W. Kan (1987). "The Same {beta} -globin Gene Mutation is Present on Nine Different {beta} -thalassemia Chromosomes in a Sardinian Population." Proceedings of the National Academy of Sciences **84**(9): 2882-2885.

Prévôt, D., J.-L. Darlix and T. Ohlmann (2003). "Conducting the initiation of protein synthesis: the role of eIF4G." Biology of the Cell **95**(3-4): 141-156.

Qu, H. Q., S. G. Lawrence, F. Guo, J. Majewski and C. Polychronakos (2006). "Strand bias in complementary single-nucleotide polymorphisms of transcribed human sequences: evidence for functional effects of synonymous polymorphisms." BMC Genomics **7**: 213.

Qu, H. Q., Y. Lu, L. Marchand, F. Bacot, R. Frechette, M. C. Tessier, A. Montpetit and C. Polychronakos (2007). "Genetic control of alternative splicing in the TAP2 gene: possible implication in the genetics of type 1 diabetes." Diabetes **56**(1): 270-5.

Qu, H. Q., L. Marchand, R. Frechette, F. Bacot, Y. Lu and C. Polychronakos (2007). "No association of type 1 diabetes with a functional polymorphism of the LRAP gene." Mol Immunol **44**(8): 2135-8.

Richards, A. J., M. Laidlaw, S. P. Meredith, P. Shankar, A. V. Poulson, J. D. Scott and M. P. Snead (2007). "Missense and silent mutations in COL2A1 result in Stickler syndrome but via different molecular mechanisms." Hum Mutat **28**(6): 639.

Rozen, S. and H. J. Skaletsky (2000). Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols: Methods in Molecular Biology. S. Krawetz and S. Misener, Humana Press, Totowa, NJ, pp 365-386.

Rund, D., C. Dowling, K. Najjar, E. A. Rachmilewitz, H. H. Kazazian, Jr. and A. Oppenheim (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**(10): 4324-8.

Rusinov, V., V. Baev, I. N. Minkov and M. Tabler (2005). "MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence." Nucl. Acids Res. **33**(suppl_2): W696-700.

Saxena, R., B. F. Voight, V. Lyssenko, N. P. Burtt, P. I. de Bakker, H. Chen, J. J. Roix, S. Kathiresan, J. N. Hirschhorn, M. J. Daly, T. E. Hughes, L. Groop, D. Altshuler, P. Almgren, J. C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H. N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E. K. Speliotes, M. R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S. B. Gabriel, G. W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke and S. Purcell (2007). "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels." Science **316**(5829): 1331-6.

Scheper, G. C., M. S. van der Knaap and C. G. Proud (2007). "Translation matters: protein synthesis defects in inherited disease." Nat Rev Genet **8**(9): 711-723.

Serrano, M., G. J. Hannon and D. Beach (1993). "A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4." Nature **366**(6456): 704-707.

Serre, D., S. Gurd, B. Ge, R. Sladek, D. Sinnett, E. Harmsen, M. Bibikova, E. Chudin, D. L. Barker, T. Dickinson, J.-B. Fan and T. J. Hudson (2008). "Differential Allelic Expression in the Human Genome: A Robust Approach To Identify Genetic and Epigenetic Cis-Acting Mechanisms Regulating Gene Expression." PLoS Genetics **4**(2): e1000006.

Shen, L. X., J. P. Basilion and V. P. Stanton, Jr. (1999). "Single-nucleotide polymorphisms can cause different structural folds of mRNA." Proc Natl Acad Sci U S A **96**(14): 7871-6.

Simon, R., J. P. Tassan and J. D. Richter (1992). "Translational control by poly(A) elongation during Xenopus development: differential repression and enhancement by a novel cytoplasmic polyadenylation element." Genes Dev **6**(12B): 2580-91.

Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T. J. Hudson, A. Montpetit, A. V. Pshezhetsky, M. Prentki, B. I. Posner, D. J. Balding, D. Meyre, C. Polychronakos and P. Froguel (2007). "A genome-wide association study identifies novel risk loci for type 2 diabetes." Nature **445**(7130): 881-5.

Smyth, D. J., J. D. Cooper, R. Bailey, S. Field, O. Burren, L. J. Smink, C. Guja, C. Ionescu-Tirgoviste, B. Widmer, D. B. Dunger, D. A. Savage, N. M. Walker, D. G. Clayton and J. A. Todd (2006). "A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region." Nat Genet **38**(6): 617-9.

Sonenberg, N. (2008). "eIF4E, the mRNA cap-binding protein: from basic discovery to translational research." Biochem Cell Biol **86**(2): 178-83.

Sonenberg, N. and A. G. Hinnebusch (2007). "New modes of translational control in development, behavior, and disease." Mol Cell **28**(5): 721-9.

Spahn, C. M., M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jorgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. Ballesta and J. Frank (2004). "Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation." Embo J **23**(5): 1008-19.

Steemers, F. J., W. Chang, G. Lee, D. L. Barker, R. Shen and K. L. Gunderson (2006). "Whole-genome genotyping with the single-base extension assay." Nat Methods **3**(1): 31-3.

Sullivan, C. S. and D. Ganem (2005). "MicroRNAs and viral infection." Mol Cell **20**(1): 3-7.

Takagaki, Y. and J. L. Manley (1997). "RNA recognition by the human polyadenylation factor CstF." Mol. Cell. Biol. **17**(7): 3907-3914.

Tomer, Y., E. Concepcion and D. A. Greenberg (2002). "A C/T single-nucleotide polymorphism in the region of the CD40 gene is associated with Graves' disease." Thyroid **12**(12): 1129-35.

Ueda, H., J. M. Howson, L. Esposito, J. Heward, H. Snook, G. Chamberlain, D. B. Rainbow, K. M. Hunter, A. N. Smith, G. Di Genova, M. H. Herr, I. Dahlman, F. Payne, D. Smyth, C. Lowe, R. C. Twells, S. Howlett, B. Healy, S. Nutland, H. E. Rance, V.

Everett, L. J. Smink, A. C. Lam, H. J. Cordell, N. M. Walker, C. Bordin, J. Hulme, C. Motzo, F. Cucca, J. F. Hess, M. L. Metzker, J. Rogers, S. Gregory, A. Allahabadia, R. Nithiyananthan, E. Tuomilehto-Wolf, J. Tuomilehto, P. Bingley, K. M. Gillespie, D. E. Undlien, K. S. Ronningen, C. Guja, C. Ionescu-Tirgoviste, D. A. Savage, A. P. Maxwell, D. J. Carson, C. C. Patterson, J. A. Franklyn, D. G. Clayton, L. B. Peterson, L. S. Wicker, J. A. Todd and S. C. Gough (2003). "Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease." Nature **423**(6939): 506-11.

Vafiadis, P., S. T. Bennett, J. A. Todd, J. Nadeau, R. Grabs, C. G. Goodyer, S. Wickramasinghe, E. Colle and C. Polychronakos (1997). "Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus." Nat Genet **15**(3): 289-92.

Vattem, K. M. and R. C. Wek (2004). "Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells." Proceedings of the National Academy of Sciences **101**(31): 11269-11274.

Wang, D., A. D. Johnson, A. C. Papp, D. L. Kroetz and W. Sadee (2005). "Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C>T affects mRNA stability." Pharmacogenet Genomics **15**(10): 693-704.

Wang, X.-Q. and J. A. Rothnagel (2004). "5'-Untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation." Nucl. Acids Res. **32**(4): 1382-1391.

Wellcome Trust Case Control Consortium (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-78.

Wells, S. E., P. E. Hillner, R. D. Vale and A. B. Sachs (1998). "Circularization of mRNA by Eukaryotic Translation Initiation Factors." Molecular Cell **2**(1): 135-140.

Wiestner, A., R. J. Schlemper, A. P. van der Maas and R. C. Skoda (1998). "An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia." Nat Genet **18**(1): 49-52.

Yoon, A., G. Peng, Y. Brandenburg, O. Zollo, W. Xu, E. Rego and D. Ruggero (2006). "Impaired Control of IRES-Mediated Translation in X-Linked Dyskeratosis Congenita." Science **312**(5775): 902-906.

Zhang, W., S. Duan, E. O. Kistner, W. K. Bleibel, R. S. Huang, T. A. Clark, T. X. Chen, A. C. Schweitzer, J. E. Blume, N. J. Cox and M. E. Dolan (2008). "Evaluation of genetic variation contributing to differences in gene expression between populations." Am J Hum Genet **82**(3): 631-40.

Zhao, J., L. Hyman and C. Moore (1999). "Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis." Microbiol Mol Biol Rev **63**(2): 405-45.

Zuker, M. and P. Stiegler (1981). "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information." Nucleic Acids Res **9**(1): 133-48.