Quantification of Protein Levels in Single Living Cells

Chiu-An Lo
Integrated Program in Neuroscience,
McGill University, Montreal
November, 2015

A thesis submitted to McGill University in partial
fulfilment of the requirements of the degree of Doctor of Philosophy
© Chiu-An Lo, 2015
Table of Contents

Abstract .................................................................................................................................8
Resume ..................................................................................................................................10
Acknowledgements ...........................................................................................................13
Author Contributions ........................................................................................................14
List of Abbreviation ...........................................................................................................15
List of Figures ....................................................................................................................17
Introduction .......................................................................................................................19
Outline of the Thesis ..........................................................................................................21
Chapter 1: Current state of protein quantification ..............................................................22

1-1 Unbalanced protein levels shift cellular homeostasis and cause cells to malfunction.......22
1-2 Current assays for protein quantification: Bulk protein quantification assays...............23
   1-2-1 Copper chelation-based assays ...............................................................................23
      1-2-1a The bicinchoninic acid (BCA) based assay .......................................................23
      1-2-1b The Lowry assay .............................................................................................25
   1-2-2 Dye-labeled protein quantification assays .............................................................26
      1-2-2a Bradford or Coomassie brilliant blue protein assay ........................................26
   1-2-3 Antibody-Antigen based assays ............................................................................27
1-2-3a Enzyme linked immunosorbent assay (ELISA) .................................................. 27
1-2-3b Quantitative western blotting ........................................................................ 29

1-3 Common issues and limitations of current assays ................................................. 37
  1-3-1 Lack of sensitivity and cellular resolution ....................................................... 37
  1-3-2 Destructive procedures ................................................................................. 38
  1-3-3 Inconsistent results ....................................................................................... 38

1-4 Single-cell protein quantification ...................................................................... 40

1-5 A genetically encoded co-expression system can ensure noninvasive single-cell protein measurement .............................................................. 42
  1-5-1 Fusion protein ............................................................................................... 43
  1-5-2 Internal Ribosome Entry Site (IRES) ............................................................ 44
  1-5-3 Cis-acting Hydrolase Element (CHYSEL) .................................................... 45

1-6 The origin of CHYSEL ....................................................................................... 45

1-7 Ribosome skipping is non-enzymatic and peptide autonomous ...................... 47

1-8 Factors affecting the 2A ribosome skipping activity ......................................... 48

1-9 Different 2A and 2A like sequences and their optimal activity in different organisms .... 50

1-10 Issues related to correct targeting and functional maintenance of multiple proteins .... 50

1-11 Strategies to increase 2A separation efficiency or minimize the inhibitory effect from the upstream sequence ................................................................. 52

1-12 The rationale of CHYSEL-based protein quantification .................................... 53
1-13 Issues associated with CHYSEL-based protein quantification……………………………..55

Figures........................................................................................................................................59

Chapter 2: The quantification of protein levels in single living cells…………61

Introduction......................................................................................................................................61

Materials and Methods....................................................................................................................64

Results...............................................................................................................................................71

- Creation and verification of Protein Quantitation Reporters (PQRs).........................71
- RFP and GFP fluorescence is directly proportional to its protein concentration over several orders of magnitude.................................................................72
- PQR can produce discrete proteins in a stoichiometric ratio.............................................72
- PQR can determine gene to phenotype relationship in single HEK293T cells..........74
- PQR can be used with different microscopy methods....................................................75
- PQR can measure protein concentration in single neurons in animals.....................75
- PQR functionality was verified in single clock neurons.............................................75
- PQR can determine quantitative relationships between protein amount and cellular phenotype in single living neurons.......................................................76

Discussion......................................................................................................................................78

- Codon deoptimization improves the separation efficiency...............................80
- PQR can relate cellular phenotype as a function of protein concentration.............82
- The study of gene to phenotype relationship in transgenic animals......................82
Chapter 3: The quantification of endogenous protein levels in single living cells

Introduction

- Why quantify endogenous proteins?
- Targeting PQR into genomic locus through the CRISPR-Cas9 gene editing technique
- Ribosomal Protein L13A (RPL13A)
- Creation of RPL13A-PQR-XFP knock-in Drosophila

Materials and Methods

Results

- PQRs can be inserted into any genomic locus to quantitate endogenous RPL13A protein levels
- Generation and evaluation of RPL13A-PQR-RFPnols knock-in Drosophila
- Red fluorescence intensities vary between different tissues but stay constant among the same type of cells

Discussion
Chapter 4: Potential application of PQR–Knock-in stable cell line production

Introduction .................................................................................................................. 127

- Generation of knock-in stable cell lines using multicistronic vector and gene editing technology ........................................................................................................... 129

Materials and Methods .............................................................................................. 130

Results .......................................................................................................................... 132

- Identifying a genomic locus for integration in human (HEK293T) cells ............. 132
- Construction of a PQR multicistronic knock-in vector ..................................... 132
- Generation of a stable and pure HEK293T cell line expressing a fluorescent reporter and a drug resistant protein ................................................................. 133

Discussion ................................................................................................................... 134

- Expression of recombinant proteins at a physiological and quantifiable level .... 135

Future perspectives .................................................................................................... 136

Figures ....................................................................................................................... 139
Chapter 5: Conclusion and future directions………………………………………141

References……………………………………………………………………….144
Abstract

Accurate protein level measurement is important for biologists since many cellular phenotypes and molecular processes are sensitive to protein levels. However, the current assays for protein quantification have remained unchanged for several decades, and suffer from several limitations such as being destructive to samples, inconsistent, and time consuming, and having poor sensitivity and poor cellular resolution. These methods can include protein assays, enzyme-linked immunosorbent assays, quantitative western blotting, and mass spectrometry.

Considering the limitations of classical tools, I created an in vivo protein quantification technique that allows protein concentration measurement at the level of single living cells. This technique, called protein quantitation ratioing (PQR), is based on the co-expression of a fluorescent protein and a protein of interest in which one molecule of the fluorescent protein production occurs with one molecule of protein of interest generation. The fluorescence measurement is a quantitative method for protein level determination because the fluorescence intensity is directly proportional to the concentration of fluorescent protein, which are expressed at the same concentration as the protein of interest. Moreover, fluorescent proteins are an innocuous reporter to the host cells, and its intensity from a single cell can be detected noninvasively through a standard fluorescence microscope. Thus the invasive and time consuming procedures of protein quantification can be replaced by a simple analysis of fluorescence intensity using any standard fluorescence microscope.

I conducted single-cell imaging, electrophysiology, and cellular phenotype assessment in animals to verify the functionality of PQR in single-cell protein quantification. Single-cell image analysis verified that the levels of protein of interest could be reliably
determined by the fluorescence output; while electrophysiology and phenotype analyses further demonstrated that PQR can relate cellular phenotype as a function of protein concentration.

Using the CRISPR-Cas9 gene editing technique, I inserted Protein Quantitation Reporters (PQRs) into different genomic loci to measure endogenous protein expression. I successfully inserted PQRs into three different genomes: human, mouse, and Drosophila to measure their Ribosomal Protein L13A (RPL13A) protein expression levels both in vitro and in vivo. This assay opens up a new avenue for biologists to quantify endogenous protein in single living cells.

PQR is a simple and quick assay that can be used by any biologist with access to standard fluorescence microscopy. It provides single-cell resolution and sensitivity for measuring even low expression proteins. Its fluorescence-based non-invasive quantification procedure allows relating cellular phenotype to protein concentration in real time. This will greatly benefit cell biologists in understanding gene to phenotype relationships.
Résumé

La mesure précise du niveau de protéines est importante pour les biologistes parce que beaucoup de phénotypes cellulaires et de processus moléculaires sont sensibles aux niveaux de protéines. Pourtant, toutes les méthodes actuelles pour la quantification de protéines sont restées inchangées depuis plusieurs décennies et souffrent de plusieurs limitations telles que leur faible sensibilité et leur manque de résolution cellulaire, leur effet destructeur sur les échantillons, leur incohérence et leur aspect chronophage. Ces méthodes comprennent des dosages de protéine, l’ELISA (enzyme-linked immunosorbent assays), le western blot quantitatif et plus récemment la spectrophotométrie de masse.

En raison de toutes les limitations inhérentes aux méthodes traditionnelles, j'ai créé une technique de quantification de protéines in vivo qui nous permet de mesurer la concentration de protéines au niveau de cellules individuelles vivantes. Cette technique, appelée protein quantitation ratioing (PQR), est basée sur la co-expression d'une protéine fluorescente et d'une protéine cible dans laquelle une molécule de la production de la protéine fluorescente a lieu avec la génération d'une molécule de protéine cible. La mesure de la fluorescence est une méthode quantitative pour déterminer le niveau de protéine parce que l'intensité de fluorescence est directement proportionnelle à la concentration de protéines fluorescentes, qui sont exprimées à la même concentration que la protéine cible. De plus, les protéines fluorescentes sont un rapporteur inoffensif pour les cellules hôtes et leur intensité dans une cellule individuelle pourrait être déterminée tout simplement par le biais d’un microscope à fluorescence standard. Par conséquent, les procédures traditionnelles de quantification de protéines, invasives et longues, pourront être
remplacées par une analyse de l’intensité de fluorescence en utilisant n'importe quel microscope à fluorescence standard.

J'ai mené des expériences d'imagerie de cellules individuelles, d'électrophysiologie et d'évaluation de phénotype cellulaire dans des animaux afin de vérifier la fonctionnalité de le PQR pour quantifier des protéines dans des s individuelles. L'analyse d’images de cellules individuelles a confirmé que les niveaux de protéine cible pourraient être déterminés de manière fiable par l’intensité de fluorescence; de plus, l'électrophysiologie et l’analyse de phénotype confirment que le PQR pourrait associer des changements de phénotypes cellulaires en fonction de la concentration de protéine.

En utilisant le technique de manipulation de gène CRISPR-Cas9, j'ai inséré le protein quantitation ratioing (PQRs) dans différents loci génomiques pour mesurer l'expression de protéine endogène. J'ai déjà inséré le PQRs avec succès dans trois génomes différents: l'humain, la souris et la Drosophile pour mesurer les niveaux d'expression de la protéine Ribosomal Protein L13A (RPL13A) in vitro et in vivo. Cet essai offre de nouvelles possibilités pour les biologistes dans le domaine de la quantification de protéines endogènes dans les cellules vivantes individuelles.

PQR est un essai simple et rapide qui pourrait être utilisé par n'importe quel biologiste ayant accès à un microscope à fluorescence standard. Il fournit une résolution au niveau de cellules individuelles et une sensibilité supérieure pour quantifier les protéines à faible niveau d’expression. Sa procédure de quantification non-invasive basée sur la fluorescence nous permet d’associer le phénotype cellulaire à la concentration de la protéine en temps réel. Cela profitera
beaucoup aux biologistes cellulaires pour la compréhension des relations entre gènes et phénotypes.
Acknowledgements

Five years of intensive but enjoyable teamwork culminated in the accomplishment of this thesis. First, I want to express my sincere gratitude to my supervisor and mentor, Dr. Brian E. Chen for his tremendous advice and guidance along this academic adventure. Alongside of him are members of the lab who supported and raised me to become a good scientist; especially Ibrahim Kays who began working on this thesis project with me as part of his undergraduate work and subsequently as a graduate student in the lab. Other members of lab such as Dr. Vedrana Cvetkovska, Tsung-Jung Lin, and Dr. Farida Emran, who also contributed substantially to the published and unpublished work presented in this thesis.

Many thanks also due to the members of my advisory committee, Dr. Keith Murai and Dr. Ed Ruthazer, their valuable advice and encouragement over the past five years helped bring this thesis project into completion.

The past five years spent at the Centre for Research in Neuroscience and McGill University had been extremely enjoyable. All the academic and non-academic activities have not only broadened my horizon in science but also enriched my general soft skills such as critical and analytical thinking, time management, and other interpersonal and communication skills.

Finally, much love and gratitude to my parents and brother in Taiwan who continuously support my academic and life journey here in Canada.
Author Contributions

Chapter 1 is composed of the background information of protein measurement assays and rationale to develop a new method for protein quantification, written by Chiu-An Lo.

Chapter 2 and part of 3 are a modified version of a research article which has been published in Cell Reports.


Brian E. Chen designed the experiments and supervised the project. Chiu-An Lo, Ibrahim Kays, Farida Emran, Tsung-Jung Lin, Vedrana Cvetkovska and Brian E. Chen performed experiments and analyzed the data. Chiu-An Lo, Ibrahim Kays, and Brian E. Chen wrote the manuscript.

Part of Chapter 3 and the entire chapter 4 presented unpublished work relevant to this thesis, written by Chiu-An Lo.
## List of abbreviation

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>alkaline phosphatase</td>
</tr>
<tr>
<td>BCA</td>
<td>bicinchoninic acid</td>
</tr>
<tr>
<td>BFP</td>
<td>blue fluorescent protein</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CFP</td>
<td>cyan fluorescent protein</td>
</tr>
<tr>
<td>CHYSEL</td>
<td>Cis-acting HYdrolaSe ELement</td>
</tr>
<tr>
<td>CRISPR-Cas9</td>
<td>Regularly Interspaced Short Palindromic Repeats-Cas9</td>
</tr>
<tr>
<td>Da neuron</td>
<td>dendritic arborization neuron</td>
</tr>
<tr>
<td>ECL</td>
<td>enhanced chemiluminescence</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme linked immunosorbent assay</td>
</tr>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>gRNA</td>
<td>guide RNA</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent Protein</td>
</tr>
<tr>
<td>HRP</td>
<td>horse radish peroxidase</td>
</tr>
<tr>
<td>IRES</td>
<td>internal ribosome entry site</td>
</tr>
<tr>
<td>NLS</td>
<td>nucleus localization signal</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>Nols</td>
<td>nucleolus localization signal</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PER</td>
<td>period</td>
</tr>
<tr>
<td>PQR</td>
<td>Protein Quantitation Ratioing (the technique)</td>
</tr>
<tr>
<td>PQR</td>
<td>Protein Quantitation Reporter (the DNA sequence)</td>
</tr>
<tr>
<td>RFP</td>
<td>red fluorescent protein</td>
</tr>
<tr>
<td>ROI</td>
<td>regions of interest</td>
</tr>
<tr>
<td>RPL13A</td>
<td>Ribosomal Protein L13A</td>
</tr>
<tr>
<td>YFP</td>
<td>yellow fluorescent protein</td>
</tr>
<tr>
<td>TALEN</td>
<td>transcription activator-like effector nuclease</td>
</tr>
</tbody>
</table>
List of Figures

**Figure 1.1** The organization of picornavirus genome and the processing of polyprotein……...59

**Figure 1.2** Different viruses adopt different strategies to separate each component from others………………………………………………………………………………………………60

**Figure 2.1** Protein Quantitation Ratioing (PQR) can determine relative protein concentration in single living cells……………………………………………………………………………………………86

**Figure 2.2** Wildtype viral CHYSEL sequences produce un-separated fusion proteins; PQR sequences produce reliable separation of proteins. ………………………………………………………………91

**Figure 2.3** Protein Quantitation Reporters can correlate fluorescence intensity with protein amount…………………………………………………………………………………………………95

**Figure 2.4** Protein Quantitation Reporters can produce discrete and functional fluorescent proteins with correlated fluorescence intensity………………………………………………………97

**Figure 2.5** PQR fluorescence intensity is linearly correlated with potassium channel current density…………………………………………………………………………………………………98

**Figure 2.6** Different fluorescence microscopy methods can be used with PQR………………100

**Figure 2.7** No fusion products produced from all PQR constructs……………………………101

**Figure 2.8** PQR can relate cellular phenotype as a function of protein concentration………..103

**Figure 2.9** PQR can relate cellular phenotype as a function of protein concentration………..105

**Figure 2.10** Consistent in phase fluctuation of red and yellow fluorescence intensities ……..107
Table 2.1 CHYSEL sequences used in the study

Figure 3.1 PQRs can be inserted into any genomic locus to quantitate endogenous protein levels

Figure 3.2 Knock-in of PQR into endogenous loci does not produce fusion proteins nor significantly alter the mRNA expression

Figure 3.3 RPL13A-PQR-RFPnols knock-in Drosophila

Figure 4.1 Stable cell lines created through CRISPR-Cas9 mediated insertion

Figure 4.2 No fusion proteins were produced from the maintained stable cell lines
Introduction

Proteins play an essential role in cellular mechanisms, as they are the effectors executing the function of genetic information. Many diseases are caused by either too low or too high protein expression. For cell biologists, accurate measurement of protein concentration ensures the accurate interpretation of gene function in certain phenotypes. The ability to quantify proteins in single cells at any given spatial and temporal moment and then associate the measured protein concentration to the corresponding cellular processes would provide the ideal quantitative determination of a gene-to-phenotype relationship. However, most of the commonly used protein quantification techniques rely on destructive assays such as protein assays (Walker, 2002), enzyme-linked immunosorbent assay (ELISA) (Engvall & Perlmann, 1971), or quantitative western blotting (Taylor & Posch, 2014). These assays are not suitable for time-lapse protein level measurement in living cells and can only measure protein abundance from cell populations or tissue sections but not from individual cells because of their limited detection capabilities. Dynamic cellular mechanisms, such as transcriptional and translational states of cells, cell signaling, cell cycle staging, apoptosis, and other molecular processes, produce cellular heterogeneity, even in immortalized clonal cell lines (Li & Xie, 2011). Due to this heterogeneous nature of cells, experimental results from single cells are not expected to be in agreement between cell populations and individual cells. This discrepancy can preclude accurate diagnosis of gene-to-phenotype relationships based on average readouts from heterogeneous cell populations. Precise protein concentration to phenotype determination is not possible in cell populations unless the processes of interest are highly synchronized such as cell cycle, circadian clock, and neuron depolarization. Nevertheless, synchronization is uncommon and undesirable as it may change the normal cell and protein functions.
In contrast to the limited cellular resolution in protein measurement, the development of single-cell polymerase chain reaction (PCR) and next-generation deoxyribonucleic acid (DNA) sequencing technology has enabled biologists to measure the concentrations of nucleic acids at the single cell level. This technological advancement provides the opportunity to relate cellular phenotypes as a function of nucleic acid concentration at the single-cell scale. Nonetheless, although proteins are synthesized from messenger ribonucleic acid (mRNA), it has been shown that cellular mRNA levels and protein abundance are poorly correlated (Schwanhäusser et al., 2011; Vogel & Marcotte, 2012). That is to say, measured mRNA levels cannot be considered a proxy for protein concentration, and hence, gene expression. Since proteins directly determine phenotypes, we need an approach to directly quantify protein amounts in single cells.

In this thesis, I developed a protein quantification technique, called protein quantitation ratioing (PQR), with the following features that we believe would be of great benefit for biologists.

1. A stoichiometric production of a fluorescent protein and a protein of interest that allows protein quantification through fluorescence intensity measurement.
2. Relative protein levels can be quantified in single living cells for accurate interpretation of gene to phenotype relationships.
3. Protein quantitation reporters (PQR) can be inserted into genomic loci using gene editing techniques to measure endogenous protein production.
4. Broad application of protein quantitation ratioing (PQR) in other areas of biology.
Outline of the Thesis

Chapter 1. Current state of protein quantification

This chapter highlights the importance of protein quantification. First, the basic principles of the current technologies are summarized and compared. Then, examples are given of different analyses of single cells, such as molecular processes and cellular phenotypes that are protein-level dependent, the poor correlation between mRNA and protein levels is discussed, and how the nature of cell heterogeneity complicates data interpretation from population samples. Finally, a discussion of the advantages of genetically encoded bicistronic expression systems in protein quantification in single cells concludes this section.

Chapter 2. Quantification of protein levels in single living cells

This chapter is part of a published article in Cell Reports. The rationale, proof of concept experiments, and potential application of protein quantitation ratioing (PQR) in biology are demonstrated.

Chapter 3. Quantification of endogenous protein levels in cultured cells and animals

This chapter consists of content from the published article in Cell Reports as well as unpublished work.

Chapter 4. Potential application of PQR–Knock-in stable cell line production

PQR offers a new approach to create stable cell lines with consistent and quantifiable levels of recombinant protein production.

Chapter 5. Conclusion and future directions
Chapter 1.

Current state of protein quantification

1-1 Unbalanced protein levels shift cellular homeostasis and cause cells to malfunction

Proteins are the final executors of gene function. Cellular phenotypes and molecular processes are dependent on protein concentration. For example, overexpression of transcription factors have been shown to increase or decrease dendritic arbor complexity in *Drosophila melanogaster* dendritic arborization neurons (Grueber et al., 2003). The excessive production of Down Syndrome Cell Adhesion Molecule (Dscam), a molecule implicated in Down Syndrome, has been shown to be associated with many abnormal neuronal phenotypes such as a decreased complexity in mouse hippocampal neurons (Alves-Sampaio, Troca-Marín, & Montesinos, 2010), enlarged presynaptic terminals of *Drosophila* dendritic arborization neurons (Jung Hwan Kim, Wang, Coolon, & Ye, 2013; Sterne, Kim, & Ye, 2015), and increased axonal targeting errors of *Drosophila* mechanosensory neurons (Cvetkovska, Hibbert, Emran, & Chen, 2013). Abnormally increased concentrations of Ephxin 5, a molecule involved in excitatory synapse development, has been shown to reduce excitatory synapse density (Margolis et al., 2010). Elevated expression of survival of motor neuron 2 (SMN2) proteins can ameliorate the symptoms of spinal muscular atrophy patients (Mailman et al., 2002). These findings demonstrate the importance of regulated protein concentration, and how it influences a wide range of cellular phenotypes and molecular processes.

To clarify complex gene-to-phenotype relationships, scientists ideally need to measure protein amounts at a single-cell scale and relate them to their corresponding cellular phenotypes. In the following sections of chapter 1, I review common protein quantification techniques used
in research labs, highlighting their working principles, utility, and limitations. Finally, I discuss why a new assay for protein quantification is essential to study gene to phenotype relationships.

1-2 Current assays for protein quantification: Bulk protein quantification assays

Most of the current protein quantification approaches are carried out at a cell population scale because of their limited detection capabilities. The three most commonly used methods are copper chelation based assays, dye-labeled assays, and antibody-antigen detection assays.

1-2-1 Copper chelation-based assay:

1-2-1a BCA (Detection range of protein concentration: 20-2000 µg/mL)

The bicinchoninic acid (BCA) based assay is one of the most widely used colorimetric methods to quantify total protein concentration in a protein sample (Smith et al., 1985). Colorimetry is a technique used to determine the concentration of a given colored compound in a solution. By passing an incident light through a cuvette that contains light absorbing compounds, the intensity of the light leaving the sample (transmitted light) is weaker than the light entering the sample (incident light) due to the absorption of light by the compounds. The degree of absorption (absorbance) is defined by the Beer-Lambert law:

\[ A = \log_{10} \left( \frac{P_0}{P} \right) = \varepsilon bc \]

\( P_0 \): radiant power of incident light

\( P \): radiant power of transmitted light
$\varepsilon$ : the molar absorptivity (L mol$^{-1}$ cm$^{-1}$)

$b$ : the light path length of the sample (i.e., the path length of the cuvette), often in centimeters.

c : the concentration of the compound in solution, expressed in mol L$^{-1}$

The measured light absorbance of the sample, also known as optical density (OD), is a function of the concentration of the light absorbing compound. Many protein assays like BCA use colorimetry as the detection method to determine the sample protein concentration.

After mixing a protein sample with the BCA reagent, the proteins in the solution will reduce cupric cation $\text{Cu}^{2+}$ (a component of BCA reagent) to cuprous cation $\text{Cu}^{+}$ in an alkaline condition, a reaction known as the Biuret reaction (Step 1). Subsequently, one molecule of the resulting $\text{Cu}^{+}$ will then react with two molecules of BCA to form a purple complex with absorbance maximum at 562 nm (Step 2). The absorbance is directly proportional to the $\text{Cu}^{+}$ concentration which reflects the abundance of total protein content. The amount of total protein can be estimated through a pre-established standard curve of absorbance and compared against a standard known protein concentration, such as bovine serum albumin (BSA).

**Step 1.**

$\text{Protein} + \text{Cu}^{2+} \xrightarrow{\text{OH}^{-}} \text{Cu}^{+}$ (Biuret Reaction)

**Step 2.**

$\text{Cu}^{+} + 2\text{BCA} \xrightarrow{} \text{Cu(BCA)$_2$ Complex}$

Absorbance : 562nm
The advantage of BCA assay is that it can be carried out in a one-step protocol as opposed to the Lowry assay, which requires a two-step reaction.

1-2-1b Lowry (Detection Range: 10-1000 µg/ml)

The Lowry assay is based on a two-step reaction: the Biuret and the Folin–Ciocalteu reaction (Lowry, Rosebrough, Farr, & Randall, 1951). Similar to BCA, the first step is the Biuret reaction, in which a divalent cupric cation Cu$^{2+}$ is reduced to a monovalent cuprous cation Cu$^{+}$ by the peptide bonds of proteins. The Cu$^{+}$ ions then react with the Folin reagent (phosphomolybdate and phosphotungstate) through a process called the Folin–Ciocalteu reaction. In the Folin–Ciocalteu reaction, the Cu$^{+}$ reduces the aromatic amino acid of the phosphomolybdate and phosphotungstate to form a blue complex with absorbance maxima at 750 nm (Peterson, 1979). Like the BCA, the amount of protein content can then be estimated by comparing it to a known protein standard.

\[
\text{Step 1. } \text{Protein} + \text{Cu}^2+ \xrightarrow{\text{OH}^-} \text{Protein-Cu}^+ \\
\text{Step 2. } \text{Folin-Ciocalteu} \text{Absorbance : 750 nm}
\]
1-2-2 Dye-labeled protein quantification assay

1-2-2a Bradford or Coomassie brilliant blue protein assay (Detection range: 20-2000 µg/ml)

The Bradford method is the most widely used dye-labeled protein quantification assay, and is based on the formation of a blue complex between proteins and Coomassie Blue G250 dyes. The quantity of the blue compounds with absorbance maximum around 595 nm is proportional to the amount of protein in the sample.

Free Coomassie Blue dye exists in four different ionic states depending on the pH of the environment: the red protonated cationic state (pH<0.39), the green neutral state (pH ~1.3), the blue unprotonated anionic state (pH>1.3), and the pink further anionic state (pH=12.4) (Chial, Thompson, & Splittgerber, 1993). The former three species are the predominant forms in an acidic solution with the maximal absorbance at 470 nm, 590 nm, and 650 nm, respectively. Only the neutral and anionic species can interact with protein and form protein-dye complexes that shift the absorbance maxima from 650 nm to 615 nm (neutral) or from 590 nm to 615 nm (anionic). This change in absorption spectra forms the foundation of the quantification of protein-dye complexes which in turn determines the concentration of proteins in the sample. In an acidic environment, the dominant red protonated cationic form (absorbance maximum at 470 nm) along with traces levels of green neutral and blue anionic species maintain a reddish color solution. Added proteins will complex with the neutral and anionic species and form the blue compounds with absorbance maximum at 615 nm (Georgiou, Grintzalis, Zervoudakis, & Papapostolou, 2008). The electrostatic attraction and hydrophobic effect between proteins and dyes in the protein-dye complex stabilize the anionic form of dye in the acidic solution. To prevent spectral crossover or contamination of measurement from free neutral dye (absorbance
maximum at 650 nm), the measurement of absorbance is usually taken at 595 nm (Georgiou et al., 2008).

On the other hand, under a very acidic environment (pH=0.23~0.64), the neutral species has been suggested to be the main source of dye to form complexes with protein components as the anionic form can exist only in solution with a pH value over 1.3 (Georgiou et al., 2008).

1-2-3 Antibody-Antigen based assays

1-2-3a Enzyme linked immunosorbent assay (ELISA)

The enzyme linked immunosorbent assay (ELISA), also known as enzyme immunoassay (EIA), provides a way to detect and quantify a specific protein of interest in a sample consisting of a mixture of proteins (Engvall & Perlmann, 1971). Although there are various formats of ELISA depending on the capture and detection of target antigens (i.e., protein of interest), the basic principles can be broken down into the following elements:

- **Coating or Capture of antigens**

  The coating/capture of antigens is the immobilization of antigens. Antigens need to be immobilized in a microplate in order to be detected and measured. This can be accomplished either by direct absorption onto the bottom of plate (coating) or through the capture of a pre-coated capture antibody. In the direct absorption (coating) approach, the whole mixtures of protein samples are attached onto the plate through passive absorption. The target antigen is then detected through a detection antibody. Because the coating is nonspecific and competitive between the target antigen and non-relevant proteins, this method has more limited detection capability as fewer antigens can be accommodated by the limited space. In contrast, protein
antigens can be captured by a pre-coated antigen specific antibody. This configuration ensures only the target antigen to be captured by the pre-coated antibody, offering a greater sensitivity and a lower background. For a sample containing low expression levels of the protein of interest, the indirect antibody-antigen capture approach can provide a better signal-to-noise ratio.

- **Blocking**

  After antigen immobilization, the remaining unoccupied sites on the plate need to be blocked to prevent nonspecific binding with detection antibody. Otherwise, any unoccupied region on the plate will be bound by antibody or any other detection reagents because of its high affinity to proteins. Any protein that does not interact with detection antibodies and target proteins and can consistently cover the remaining region can be used as blocker. Some blockers such as BSA are preferred by most investigators as they can also stabilize the antibody used. An ideal blocking step should reduce the background signal and intensify the signal-to-noise ratio.

- **Probing or detection**

  The detection of the immobilized antigen can also be carried out through a direct or indirect way. In the direct assay, the detection antibody, conjugated to a signal producing enzyme, such as horse radish peroxidase (HRP) or alkaline phosphatase (AP), or labelled with a fluorophore, binds directly to the captured antigens. On the other hand, in indirect detection, a primary antibody first recognizes the immobilized antigen and then a secondary antibody, conjugated with an enzyme or fluorophore, will bind to the primary antibody (a method sometimes called a sandwich assay). Generally speaking, a primary antibody contains more than one epitope for several secondary antibodies; multiple secondary antibodies conjugated with signal producing enzymes can bind to one primary antibody, thus amplifying the signal. Therefore, an indirect detection method usually provides higher sensitivity than a direct detection one.
• **Signal quantification**

Once the antigen is bound with the antibody, the next step is to measure the quantity of the captured target protein. In an enzymatic reaction-based detection assay, the detection is generally accomplished through the addition of appropriate substrates that can be catalyzed by the conjugated enzyme into quantitative levels of colorimetric, fluorescent, or luminescent compounds. The signal intensity is thus determined by the light absorbance, the emitted fluorescence, or luminescence intensity, respectively. In general, fluorescence and luminescence based detection methods have broader dynamic ranges than colorimetric assays.

In addition to the enzymatic reaction-based method, the primary or secondary antibodies can be labeled with a fluorophore. The fluorescence can be directly read out through a fluorimeter without substrate addition. This provides a greater linear range than enzymatic reaction-based assays in which the range of linearity is often restricted by the dynamic nature of enzymatic reaction. No matter which the detection method is used, the measured signal is used to determine the relative quantity of target antigens existing in the sample.

**1-2-3b Quantitative western blotting**

Immuno-blotting, or more commonly known as western blotting, is another widely used antibody-based approach to detect a specific protein of interest from purified or crude lysates. Since it was invented over 3 decades ago (Towbin, Staehelin, & Gordon, 1979), western blotting has become one of the standard protein analysis assays in research labs. As opposed to ELISAs, western blotting can detect the specific protein of interest from a protein complex mixture more precisely, as an electrophoresis step separates out different protein components according to their molecular weights. That is, even if the primary or secondary antibodies cross react to non-target antigens, the signal (on target) can still be differentiated from the noise (non-specific binding) on
the basis of molecular weights. In the past, western blotting was used mainly to simply detect certain antigens (protein of interest) in a given sample; but thanks to the advancement of detection media, reagents, and antibody quality, the results from western blotting have become more and more reliably quantitative. Nowadays, a standard western blotting result is often accompanied with quantitative interpretation such as fold changes of target protein levels between control and experimental conditions (Taylor & Posch, 2014). Below, I identify and discuss common caveats associated with each step of the western blotting procedure that need to be considered when it comes to performing quantitative or, sometimes referred to as, semi-quantitative western blotting.

**Sample preparation**

Well-prepared fresh samples are undoubtedly the most important part for producing high quality western blot data. Protein samples often come from cellular or tissue lysates, which can be extracted through physical (grinding, sonication, freeze-thaw, etc.) or non-physical (lysis buffer) approaches. No matter what methods are employed for the extraction, ensuring an effective extraction and preserving the integrity of target proteins are crucial. Several factors can affect the extraction efficiency that leads to underestimated protein quantification, including the presence of active proteases, incomplete cell lysis, ineffective reagents, and the used of pure non-ionic detergents (Triton X-100 or NP-40) which can cause proteins to partition in soluble and insoluble phases after centrifugation. Moreover, for studies focusing on cell signaling, in which a quantification of the ratio of phosphorylated and nonphosphorylated enzymes is often required, the addition of effective phosphatase inhibitors is also required.
**The linear dynamic range of sample load**

The total lysates or the purified protein samples loaded into each well can range from 10 to 90 µg depending on the well size and target protein expression level. The amount of loading should lie within the linear dynamic range of the detection method, i.e., the range in which the densitometric or fluorescence signal is proportional to the amount of protein loaded. Many investigators often overlook the importance of the linear dynamic range when loading the samples. Sample overloading can lead to signal saturation that results in a non-linear relationship between the signal intensity and the protein abundance. In other words, the observed differences in signal intensities between different samples cannot relate to the proportional differences in real protein amount. One cause is when protein samples are overloaded, they may form layers of proteins in the gel and subsequently in the blot membrane after transfer. The detection antibody (primary antibody) has access to and can detect only the upper layer of antigens. Proteins located below are not effectively probed. Consequently, the result is a saturated image caused by overloading which manifests itself as a plateau in the protein load versus signal intensity plot. To avoid the issue of membrane saturation, establishing a standard curve of a protein amount versus signal intensity is often recommended. Any point falling outside of the linear range of the standard curve cannot be used for quantification.

**The appropriate loading control for normalization**

Many “housekeeping” proteins, including β-actin, tubulin, or glyceraldehyde 3-phosphate dehydrogenase (GAPDH), are often used as loading controls to normalize the signal intensities from sample to sample. Nevertheless, the perception that housekeeping proteins are a reliable
reference for quantitative analyses in quantitative PCR (qPCR) and western blotting has been heavily challenged.

- The expression levels of many housekeeping proteins are not always uniform within the same type of cells across experimental conditions (cell cycle and differentiation phase, cell density, drug treatment, age, etc.), not to mention between different tissues. This differential expression level of housekeeping proteins can preclude an accurate protein quantitation (Dittmer & Dittmer, 2006; Ferguson et al., 2005; Greer, Honeywell, Geletu, Arulanandam, & Raptis, 2010; Rocha-Martins, Njaine, & Silveira, 2012).

- The highly expressed housekeeping proteins often have very narrow linear dynamic ranges as a result of membrane saturation. This issue is notably troublesome when it comes to normalizing a protein of interest with low expression levels, as the minimal amount of sample loading might still saturate the housekeeping protein signal (Gorr & Vogel, 2015; Taylor, Berkelman, Yadav, & Hammond, 2013; Taylor & Posch, 2014).

- A recommended alternative loading control for quantitative western blotting is the total protein quantification. That is, instead of searching for the ideal housekeeping protein for a specific experiment, which is time consuming and often non-feasible, quantifying the total protein loaded at each well by the Ponceau S, the SyproRuby Protein blot stain, or the Coomassie Blue staining either directly on the gel (before transfer) or on the membrane (after transfer) has been demonstrated with greater signal consistency and broader dynamic range (linearity of 1-40µg protein load versus 1-30µg for β-actin or 1-8 µg for tubulin using LI-COR Odyssey Fluorescence quantitative system and less than 1-2 and 1-5 µg when using ECL detection for β-actin and tubulin, respectively) (Eaton et al., 2013; Welinder & Ekblad, 2011). Among them, the Coomassie Blue staining is one of
the most widely used thanks to the simplicity, non-covalent interaction with protein, compatibility with downstream mass spectrometric analysis, and time and cost effectiveness. Recently, a stain free total protein quantification assay has been demonstrated with equivalent or even better sensitivity and dynamic range than other staining methods (Colella et al., 2012; Gilda & Gomes, 2013; Gürtler et al., 2013). This method is based on the use of a gel incorporated with a trihalo compound, which will covalently react with tryptophan residues of the proteins in the gel upon UV irradiation. The trihalo-tryptophan complex is fluorescent under UV exposure, providing the opportunity to pre-evaluate the quality of electrophoresis and protein transfer before proceeding to the subsequent time- and cost-intensive immuno-detection.

**Detection of signal**

Target protein abundance can be translated into either chemiluminescence or fluorescence intensity depending on which type of secondary antibody is used. For chemiluminescent western blotting, a secondary antibody conjugated with an enzyme such as horseradish peroxidase (HRP) or alkaline phosphatase (AP) is used to convert substrate into chemiluminescence (light) emitting product. Since the conjugated enzyme (HRP or AP) is in complex with target proteins, the light intensity is correlated to the target protein amount. In the case of HRP, a detection method called enhanced chemiluminescence (ECL) is often used. Specifically, in the presence of hydrogen peroxide and a phenolic enhancer commonly included in the ECL reagent, HRP catalyzes the oxidization of luminol into an excited state product which then quickly decays back to ground state and emits blue light (428nm) (Patton, 2000). The phenolic enhancer can increase light emission up to 1,000 fold and extends the time of emission to 60 minutes. This allows chemiluminescent western blotting to detect target protein levels as
low as 1-10 pg (Patton, 2000). The light signal can either be detected directly by an imager or captured by an X-ray film through the light conversion of silver halides into metallic silver on the film. The optical density (darkness) of the band on the film is used to measure the abundance of protein on the membrane. Although more quantitative digital imaging has recently gained in popularity, X-ray film is still the most widespread signal detection method for most research labs (Janes, 2015).

A potential caveat inherent to chemiluminescent method is the limited signal linearity due in part to the dynamics of the enzymatic reaction. For a quantitative measurement, a linear dynamic range is often established and then the concentration of an unknown protein can be calculated through the linear regression model. In some cases, the correlation between the captured chemiluminescent signal intensity and the real protein abundance might be better fitted by a hyperbolic line with narrow linear range. The lack of linearity caused by membrane saturation or runaway artifact of the enzymatic reaction can compromise the accurate quantification and even affect qualitative interpretation. As opposed to saturation which often leads to underestimated protein levels, in a typical runaway artifact, a two-fold change in protein abundance might result in 5 fold or greater differences in band intensity owing to the non-linear nature of enzymatic reaction, causing an overestimation of the target protein abundance (Janes, 2015). This overestimation of protein levels can potentially lead to an exaggerated claim that the effect in gene expression induction or reduction is significant whereas the true difference in protein quantity might be subtle.

Another disadvantage of chemiluminescence is associated with the use of photographic film for capturing chemiluminescence. Film imaging has a much narrower dynamic and linear range than digital imaging because as the film gets more exposed, fewer silver halide crystals are
left for absorbing more photons. This results in a logarithmic value curve of optical density (OD). The ability to quantify low expression target proteins can be compromised because the signals of reference proteins (housekeeping proteins) often saturate the membrane before the level of protein of interest reaches the detection limit. Although the use of a charge-coupled device (CCD) camera detection method offers a wider dynamic range and increased sensitivity for chemiluminescent western blotting, the intrinsic non-linear enzymatic reaction renders it a qualitative or semi-quantitative assay.

On the other hand, fluorescent western blotting employs a secondary antibody appended to a fluorophore. The fluorophore will fluoresce when illuminated with light of the appropriate wavelength of excitation. The fluorescence output, captured by a CCD camera or a photomultiplier tube (PMT) in a laser scanner, is mainly determined by the amount of the fluorophore. Therefore, as opposed to chemiluminescence, in which the light intensity depends on several factors (enzymatic reaction time, quantities of enzymes and substrates, pH of the reagent, etc.), the fluorescence intensity is a static value directly proportional to the abundance of target proteins in complex with the fluorophore-conjugated antibodies (Furtado & Henry, 2002). Indeed, fluorescence western blotting has wider linear and dynamic ranges than its chemiluminescence counterpart (Zellner et al., 2008), thus it is a more accurate approach for quantitative western blotting. The only drawback is the relatively lower sensitivity than chemiluminescence since the CCD cameras that are used for signal detection cannot amplify weak signals to detect low abundance proteins (Zellner et al., 2008). In contrast, in PMT based laser scanning imaging systems (e.g., LI-COR Odyssey Infrared Fluorescent Western), the emitted light signal can be multiplied by the PMT as much as $10^8$, in which even a single photon can be readily detected (Zellner et al., 2008; Hensley, 2009). A PMT-based scanning system has
sensitivity comparable to a chemiluminescence method and can be a good alternative when both sensitivity and wide dynamic range are required.

**Analysis of signal**

The signal of chemiluminescence is represented either as optical density (film based detection) or pixel intensity (CCD camera) when the fluorescence is acquired through CCD camera or laser scanner. The OD value on the film is analyzed through densitometry, which has been demonstrated as an unreliable way of quantification. In addition to the aforementioned issues such as membrane saturation caused by overloading, non-linearity, and small dynamic range, two pitfalls associated to densitometry have also been raised (Gassmann, Grenacher, Rohde, & Vogel, 2009). The first issue is inherent to the digitization of film images into computer. In many cases, a standard office image scanner is used to acquire images. Gassmann et al. demonstrated that the automatic gain control and the lack of shading correction feature render the acquired images through image scanners misrepresentative of the original data. The resultant gray values (OD) of the bands are a function of the neighboring values resulting from the automatic gain control. The nonhomogeneous illumination results in a non-uniform brightness of the field of view, which can be corrected by shading correction in a standard CCD camera but not in an office scanner. This uncorrected illumination can lead to a 1.5 to 105-fold OD variation in the acquired image. The recommended procedure is digitizing the film image through a CCD camera with shading correction or completely replacing the use of film by using a phosphorimager for photostimulated luminescence.

The second pitfall is related to the image measurement, especially the algorithm used for background subtraction. Gassmann et al. also showed that background correction needs to be
performed at each lane independently, instead of picking the average and applying it to the whole image like the rolling ball algorithm used in ImageJ.

**Detection of Non-target bands**

Detection of non-target signals is one of the most commonly encountered issues of western blotting and other antibody antigen based assays including ELISA. The appearing of unexpected bands can be the result of several factors, such as cross reactivity of antibody, phosphorylation states of the protein antigens, the existence of protein isoforms sharing the same epitopes (antigen recognition sites), insufficient cleavage of protein domain by protease, or incomplete separation of discrete proteins from polycistronic mRNA ([Figure 2 in Chapter 2](#)). Often, in the attempt to show favorable results, the non-target bands and background are cropped out and only the framed expected bands are analyzed and presented in figures. If the unexpected bands are coming from different phosphorylated forms or uncleaved products, this will lead to inaccurate quantification of protein levels. Indeed, when the sources of non-target band are unaccountable, including them in analysis is recommended (Gassmann et al., 2009). To avoid image manipulation and ensure accurate data interpretation, entire western blot images are requested by scientific journals.

1-3 Common issues and limitations of current assays

1-3-1 Lack of sensitivity and cellular resolution

To prepare samples for each assay, the first step is collecting a population of cells or section of tissues in order to provide enough sample for the detection processes. Tissue dissection is often not as precise as desired for preparing a homogeneous specimen, as the dissected samples inevitably contain non-relevant tissue. Even an isolated cell population derived
from an immortalized cell line cannot be considered a homogeneous pool, as cell-to-cell
variation in transcriptional and translational profiles has been demonstrated even in stably
maintained cell lines (G.-W. Li & Xie, 2011). Given the considerable cell-to-cell heterogeneity
in cellular processes such as gene expression, cell cycle, and development, the results obtained
from a population sample cannot be translated into the single-cell scale. Single-cell western
blotting has been developed but it is not a common method (Hughes et al., 2014).

1-3-2 Destructive procedures

To measure protein quantity, protein samples need to be extracted from cells or tissues
for subsequent analysis. This invasive procedure precludes experiments such as time-lapse
measurement of protein levels, which is often an essential task for studying synaptic plasticity,
neuronal development, or circadian rhythms. The information acquired from these assays is
simply a static picture of existing protein components and concentration from a population of
cells or tissues at the time of collection.

1-3-3 Inconsistent results

Several factors render the results from standard protein assays inconsistent and result in
wide variation from batch to batch.

- The presence of interfering substances

Protein assays are often subject to the influence of interfering substances existing in the
sample. These substances can affect the accuracy of protein measurement through the following
three effects:

a. Suppressing the response of the assay

b. Enhancing the response of the assay
c. Increasing the background signal

For example, the Lowry-based assay is very sensitive to reducing and chelating agents such as DTT, beta-mercaptoethanol, cysteine, and EDTA; while the Bradford method can be impaired by many detergents. In contrast, the BCA assay is more tolerant to many compounds that affect the Lowry assay but it is more sensitive to reducing sugars. Any reagents that might chelate copper can underestimate both the Lowry and the BCA protein level measurement. Although the interfering substances can be removed through protein purification, such as filtration and dialysis, the procedure is time-consuming and usually results in a diluted sample.

- **Issues with antibodies**

For the antibody-based methods (i.e., ELISA and western blotting), the quality of antibody often determines the accuracy and consistency of the results. Problems inherent to antibody, such as avidity and affinity of the antibody, accessibility to the epitopes (target antigens), phosphorylation state of the antigen, and the cross-reactivity of antibodies, often lead to low reproducibility and prevent a reliable quantification.

- **Time consuming procedure**

To quantify the concentration of proteins in each sample, a standard curve is needed to convert the assay reading into protein amount. This is usually created by making a serial dilution (1:2 or 1:10) of a known concentration of pure protein standard, like bovine serum albumin (BSA), and plotted against their corresponding absorbance reading in the assay. The target protein sample concentration can then be extrapolated from the linear regression line generated by the protein standard. For quantitative western blotting, the amount of protein sample loaded into each well needs to be controlled as well because a saturated membrane cannot produce an
accurate readout reflecting the true sample quantity. To avoid such membrane saturation issues, it is desirable to establish a standard curve of loaded protein quantity versus signal intensity prior to the real experiment. This linear standard curve can ensure the amount of sample loaded into each well falls within the linear dynamic range for quantification. These prerequisite procedures can potentially lengthen the protocol of protein quantification.

1-4 Single-cell protein quantification

Most of the commonly used protein quantification techniques in research labs lack single-cell resolution because the data collected is just an average of thousands of individual cells. Such a bulk measurement on a population of cells is not representative of the biology of single cells because of the tremendous heterogeneities at the level of individual cells (Hoppe et al., 2014; Larson et al., 2010; Tsioris et al., 2014). The variations of biology at single cell levels can manifest in many ways, from the genome to transcriptome, to how the cell reacts to external signals and pressures. Different forms of single-cell variations are often masked or averaged out in the bulk measurement of cell population. Therefore, the data collected from cell populations (bulk measurement) often fail to quantify the responses of single cells to stimuli and identify events at the level of single cells. Recently, the accessibility towards single cell biology has increased tremendously thanks to advances in technology that improve the manipulation and measurements of sparse samples at unprecedented precision. For example, the advance of single cell reverse transcription quantitative polymerase chain reaction (RT-qPCR) technology has enabled researchers to measure the transcript (mRNA) levels at different conditions. However, according to transcriptomic and proteomic studies, mRNA and protein abundances are poorly correlated even at the population scale (Schwanhäusser et al., 2011; Vogel & Marcotte, 2012). This suggests that the use of transcript levels as a proxy for protein abundance is likely to be
erroneous. Given that proteins are the real actors in executing most cellular functions, developing an approach for single-cell protein quantification is highly desirable for biologists. As the technologies advance, accurate measurement of protein levels at the single cell level is no longer unattainable. Essentially, numerous mature and emerging techniques such as single cell capillary electrophoresis (Harwood et al., 2006), proximity extension assay (Darmanis et al., 2015; Greenwood et al., 2015; Lundberg et al., 2011), and mass spectrometry (Fujii et al., 2015) have proved to be invaluable to quantify protein abundance in single cells. Nonetheless, they remain destructive and lengthy assays as cells must be lysed to extract the protein content for manipulation and measurement.

Ideally, a reliable assay for single-cell protein measurement should meet the following criteria:

- **Fast and easy**
  
  Instead of going through the time-consuming processes from cell collection and lysis, protein extraction and purification, standard curve establishment, to protein concentration measurement, the ideal method would provide an instant readout of protein levels without a complicated procedure.

- **Non-invasive**
  
  To study gene-to-phenotype relationships, a time-lapse protein measurement is required to relate cellular phenotype as a function of protein concentration.

- **Single-cell resolution and sensitivity**
  
  The method should allow quantification of low abundance protein expression levels at single-cell scale.

- **Robust and universal**
The assay should be autonomous and insensitive to factors such as cell status, experimental condition, and presence of interfering substances, which affect all the other currently used protein quantification methods. It should also be applicable in any cell type.

1-5 A genetically encoded co-expression system can ensure noninvasive single-cell protein measurement

To meet these four requirements, we searched for a genetically-encoded multiple protein expression system that allows for co-expression of a fluorescent reporter and a protein of interest, so that the fluorescence output can be used to determine the expression levels of the protein of interest. Such co-expression vectors, also known as bicistronic vectors, have been widely used to express multiple components in biomedical applications, such as subunits of multimeric proteins like antibodies or hormones, multiple pluripotent genes for iPS cell reprogramming, or the co-expression of a selectable marker of drug resistance gene or fluorescent reporter to track transfected cells (Szymczak et al., 2004; Walseng et al., 2015; Polypeptide et al., 1999; Sommer et al., 2009; Daniels, et al., 2014).

Co-expressing a fluorescent protein along with the protein of interest through bicistronic vectors has enormous advantages. First, the output of the co-expressed fluorescent reporter can be used to read out the production of the protein of interest through time-lapse single-cell imaging using standard fluorescence microscopy methods (Figure 1 in Chapter 2). Secondly, because fluorescence intensity is directly proportional to its protein level over a wide dynamic range (Furtado & Henry, 2002) (Figure 1c in Chapter 2), the fluorescence output (brightness) from each cell indicates the expression level of the target protein (Figure 1b in Chapter 2). Finally, the fluorescence intensity can be enhanced by choosing brighter fluorophores or
sequestering fluorescent proteins into small subcellular compartments, such as nuclei or nucleoli, to allow for quantifying low expression proteins (Figure 9 in Chapter 2). Thus far, several strategies have been employed by biologists to express multiple proteins, each of which exhibit advantages and disadvantages.

1-5-1 Fusion protein

A fused fluorescent protein tag can be used to track the expression and distribution of a protein of interest (Figure 5a shakerGFP in Chapter 2). With the fluorescent reporter, the production and trafficking of the target protein can be observed through fluorescence microscopy. Nonetheless, the fused fluorophore might cause undesirable modifications to the protein of interest. Indeed, to most biologists, any modification, regardless of whether it is a fusion reporter or just a short peptide tag, should be minimized to prevent any influences on the protein of interest, such as creating steric hindrance, changing the half-lives of proteins (Figure 8b in Chapter 2), or interfering with the functions of proteins. Additionally, for easy and reliable quantification, the protein of interest needs to be expressed at sufficiently high levels and be localized to a geometrically uniform place. For proteins with punctate distribution (e.g., in mitochondria, dendritic spines, or axonal terminals) or low expression levels (many transcription factors), taking an accurate quantification can be challenging, while completely obviating usage for secreted proteins such as hormones or antibodies. A simple solution to this issue is to incorporate a protease recognition site at the linker between the two proteins (Bowler et al., 2002; Fang et al., 2007a; Verrier et al., 2011), so that they can be cleaved off by endogenous cellular proteases. The cleaved fluorophore can be sequestered into other subcellular compartments, such as the nucleus, for a more accurate measurement. But this approach requires an effective contact between endogenous enzymes and substrates; often a co-localization at the same subcellular
compartment is needed. The different types of enzymes and recognition sequences can also change the cleavage efficiency (Fang et al., 2007a). Failure to do so can lead to partial or ineffective cleavage.

1-5-2 Internal Ribosome Entry Site (IRES)

Another option of multiple-gene co-expression is the incorporation of the internal ribosome entry site (IRES) sequence between the gene of interest and a selectable marker. The upstream and downstream genes separated by an IRES sequence are transcribed by a common promoter into a single multi-cistronic mRNA. The intervening IRES, serving as an internal ribosome entry site, recruits eukaryotic ribosomes to initiate translation at the downstream gene in a 5’cap independent manner; whereas the upstream gene is translated in a standard 5’cap dependent way. The two different mechanisms of translation initiation are independently regulated as IRES-driven translation is insensitive to many translational inhibitors that shut off the upstream gene translation (Sugimoto et al., 1995; Zhou et al., 1998). The lack of coordination in translation between the two cistrons manifests through the fact that the translation of the downstream genes is generally attenuated (Sugimoto et al., 1995; Zhou et al., 1998). Furthermore, different downstream gene sequences (Rijnbrand et al., 2001) or even different arrangements of coding genes in the construct can lead to various efficiencies in IRES-directed translation (Hennecke, 2001). The IRES element was first identified in RNA viruses and subsequently in DNA viruses and even eukaryotic cells (Balvay, Rifo, Ricci, Decimo, & Ohlmann, 2009). Different IRES sequences with different activities in different organisms and various lengths (100 DNA bases to 1 kilobase) have been isolated. The widely used versions for co-expression in biotechnological application are around 500 bases. Their bulky sizes can potentially limit their utility because it would compromise the efficiency of homologous recombination in knock-in
cell or organism creation, and many commonly used virus-based vectors have limited coding accommodation capacity (~5 kilobases for adeno-associated vectors and ~7-8 kilobases for retroviral vectors) (Minskaia & Ryan, 2013).

1-5-3 Cis-acting Hydrolase Element (CHYSEL)

The Cis-acting Hydrolase Element (CHYSEL), also called 2A or 2A-like sequences, is an alternative bicistronic expression system for multiple protein production. As opposed to IRES, in which the upstream and downstream proteins are translated independently through a cap-dependent or independent manners, CHYSEL is a 20-30 amino acid self-processing peptide which can direct co-translation of two separate proteins from a single polycistronic transcript through the manipulation of the ribosome.

The equimolar production of upstream and downstream proteins thus suggests a potential application in protein quantification as the production of one molecule of fluorescent reporter protein will come with one molecule of the protein of interest (Figure 1a in Chapter 2). The fluorescence output can be used to quantify the expression of protein of interest.

1-6 The origin of CHYSEL

The CHYSEL sequences were originally identified from picornavirus to express multiple proteins from its single polycistronic genome. Picornavirus is a family of single strand RNA viruses including enterovirus, rhinovirus, cardiovirus, and aphthovirus (also called foot and mouth disease virus). A single picornavirus genome contains a single long open reading frame that encodes a polycistronic protein chain (Figure 1).
Nevertheless, a long single precursor protein never exists, as each cistron will be “cleaved” into discrete components, P1, P2, and P3, through either intramolecular proteolytic cleavage or translational interruption (Figure 2).

P1 is a group of capsid proteins used to assemble materials into new viral particles. The nonstructural portion, consisting of P2 and P3, form the replicative domain that is responsible for viral genome replication. To separate each component from the others, different types of viruses adopt distinct strategies. In the case of enteroviruses and rhinoviruses, the encoded proteinases from the 2A sequence at region 2 (2A\textsuperscript{pro}) and 3C sequence at region 3 (3C\textsuperscript{pro}) will cleave the nascent proteins at their own N termini, delineating the junctions of each region. In cardio- and aphthovirus, a nonezymatic cleavage event will separate region 2 apart from region 1 at the C-terminus of 2A together with the analogous proteolytic cleavage at region 3. The 2A sequence in this case does not encode any protease for enzymatic cleavage but rather translationally interrupts the peptide bond formation at the C terminus of 2A, which ends up creating discrete products translationally. The 2A and other similar 2A like sequences have been discovered in many other different viruses. They all share the same conserved C terminal consensus sequence, DxExNPG^P (x can be any other amino acid; ^ is the point of separation), with various N terminal sequences. The synthesis of the upstream peptide ends at the amino acid glycine; the translation reinitiates from the following proline to synthesize the downstream peptide. A proposed ribosome skipping mechanism was proposed to describe this non-proteolytic phenomenon.

When the translational machinery reaches the end of the 2A sequence, the presence of the newly synthesized nascent 2A peptide in the ribosome exit tunnel will conformationally restrict the peptide bond formation between the last two amino acids, glycine and proline. Specifically,
certain residues in the nascent α-helical 2A peptide are proposed to interact with the ribosome exit tunnel. This interaction creates a tight reverse turn at the C-terminus of the 2A peptide in the peptidyl transferase center (PTC), in which a new incoming amino acid is added to the growing peptide chain through the activity of peptidyl transferase. This turn will shift the ester linkage between the peptide and tRNA\textsuperscript{ Gly} (P site) away from the incoming prolyl–tRNA (A site), inhibiting the peptide bond formation with prolyl–tRNA. Proline at prolyl–tRNA is a poor nucleophile because its ring structure functional group sterically and conformationally constrains the degrees of freedom of the nucleophilic attack. The fact that the tRNA\textsuperscript{ Gly} is shifted away further disfavors the ester linkage formation. Therefore, translation will stall and then the ester bond between the nascent peptide and tRNA\textsuperscript{ Gly} will be hydrolyzed within the ribosome, thereby releasing the first peptide chain through a still inconclusive mechanism (V. A. Doronina et al., 2008; Machida et al., 2014). The prolyl-tRNA will translocate into the P site of the ribosome and the next aminoacyl tRNA with complementary anti-codon comes in, occupying the A site where the translation of the downstream protein can be reinitiated.

1-7 Ribosome skipping is non-enzymatic and peptide autonomous

Several lines of evidence exclude the involvement of any cellular proteases in the separation process. In an in vitro translation assay using rabbit reticulocyte lysates, constructs with FMDV 2A produced a significant level (20%) of uncleaved products among the total proteins (Ryan & Drew, 1994). The uncleaved fusion proteins were not subsequently cleaved after a prolonged incubation with the addition of protein synthesis inhibitors RNase A and cycloheximide, indicating the “cleavage” is not enzymatic but solely a cotranslational process. Also, when FMDV 2A is placed after a protein containing an endoplasmic reticulum (ER) targeting signal peptide, the newly translated nascent signal peptide will trigger the engagement
of ribosomes with ER and form a translocon complex to translocate the synthesized peptide into the ER lumen. The growing peptide chain is then co-translationally translocated into ER through the pore of the translocon complex. The docking of ribosome into the ER translocon can isolate the nascent peptide from any cytosolic proteases. This confirms that the F2A separation mechanism is a translational effect rather than an enzymatic cleavage (de Felipe et al., 2003). The ribosome skipping effect has been observed in all the eukaryotic ribosomes (mammal, insect, fungus, and plant) both in vitro and in vivo (de Felipe et al., 2006). On the contrary, when 2A was expressed in prokaryotic cells, no ribosome skipping effect has ever been observed. (Donnelly et al., 1997; Machida et al., 2014).

1-8 Factors affecting the 2A ribosome skipping activity

Ribosome skipping is the outcome of the inhibition on the peptidyl transferase activity that interrupts the peptide bond formation between glycine and proline. Any factors that inhibit the peptidyl transferase activity can result in various separation efficiencies. The interaction between the nascent 2A peptide with the ribosome exit tunnel has been proposed as the key factor to ensure the proper production of discrete proteins, and employing different lengths or sequences of 2A peptides has been shown to result in different activities in a heterologous context (Sharma et al., 2012). For F2A, the most widely used version for heterologous expression is 18 amino acids (aa) long. The N terminal end of the 18 aa FMDV 2A is delineated by the cleavage of viral protease as part of the posttranslational process in order to remove the residual 2A from the upstream 1D protein. The sequence upstream of the 18aa is thus considered nonessential for 2A activity. However, in heterologous systems an increase in efficiency of separation and reliability of equimolar production of N terminal and C terminal proteins has been shown as the length of 2A peptide increases from 18 aa (85% separation), to 24 aa (96%
separation), and to 30 aa (>99% separation) (Donnelly et al., 1997; Donnelly et al., 2001b). That is to say, the C terminal sequence of the 1D protein, upstream of the 18 aa 2A, can improve the 2A activity. A mutagenesis study also has confirmed that the amino acid residues in the 2A sequence upstream of the consensus DxExNPGP can considerably affect the activity (Sharma et al., 2012). Given that the ribosome exit tunnel can accommodate a nascent peptide chain of up to 40 aa (Malkin & Rich, 1967; Nissen, 2000; Picking, Picking, Odom, & Hardesty, 1992; Voss, Gerstein, Steitz, & Moore, 2006), the separation efficiency of FMDV 2A is thus proposed to be determined mainly by the sequence within the 30 to 40 aa context. Similar improvements in activity have been also found in other 2A like sequences when a longer version of 2A was used. For example, the separation efficiency of infectious flacherie virus (IFV) increases from ~63% to ~99%. In cricket paralysis virus (CrPV), the improvement is relatively marginal (from ~88% to ~90%) (Luke et al., 2008). In addition, when the short version (18 aa) was utilized for co-expression, the peptide sequence immediately upstream of 2A has been shown to affect the separation if the N-terminal protein was targeted to the exocytic pathway, confirming that an authentic 30 aa 2A nascent peptide is essential for an optimal activity presumably resulting from the interaction with the residues in the exit tunnel (de Felipe et al., 2010). In my thesis project, we also find that deoptimized codon usage with 18 aa length of P2A and T2A can also enhance their separation activities (Figure 2 in Chapter 2).

Stoichiometric production of discrete proteins

Theoretically, 2A can ensure a stoichiometric production of upstream and downstream proteins. This feature was first verified using Fluorescence Activated Cell Sorting (FACS) analysis of cells expressing different fluorescent proteins (Szymczak et al., 2004) and single-cell fluorescence microscopy (Goedhart et al., 2011) (Figure 3 in Chapter 2).
1-9 Different 2A and 2A like sequences and their optimal activity in different organisms

Although the separation of the two peptides is proposed to be the outcome of the interaction between the nascent 2A peptide and the exit tunnel of eukaryotic ribosomes, accumulated evidence suggests that the separation efficiency also depends on the type of 2A used and the host organisms. For the most widely used F2A, its activity can vary from 40% (Jin Hee Kim et al., 2011), to 90% (M. Donnelly et al., 2001a), and to even 100% (Szymczak et al., 2004), depending on the experimental conditions such as the model organisms or cell lines used. Kim and colleagues systematically evaluated the activities of different 2A like peptides in different host cells or organisms: two different human cell lines, zebrafish embryos, and mouse liver cells (Jin Hee Kim et al., 2011). Their data show that porcine teschovirus-1 2A (P2A) outperformed other 2A peptides used: F2A, equine rhinitis A virus 2A (E2A), and Thosea asigna virus 2A (T2A).

In Drosophila, T2A has been shown to have optimal activity both in cell lines and animals (Daniels et al., 2014; González et al., 2011).

1-10 Issues related to correct targeting and functional maintenance of multiple proteins

In addition to the concern of CHYSEL separation efficiency, another issue is the potential mis-targeting or disruption of protein function. In particular, when the protein upstream of 2A is targeted through the exocytic pathway, the protein downstream of 2A (with or without any target signals) might slipstream through the ribosome::ER translocon pore co-translationally, and then accumulate inside the ER lumen. This mis-targeting of the downstream protein has been proposed to be caused by signal shielding, as the docking tunnel of the ribosome-translocon complex could shield the downstream protein from recognition by the appropriate signal
recognition particles (SRP). To this test issue, in 2004 de Felipe & Ryan conducted a systematic study to analyze the targeting of cyan and yellow fluorescent proteins bearing different signal peptides in mammalian cells. Their data showed that a targeting signal sequence in the second protein but not the first (e.g., YFP-2A-signal-CFP), allows the first protein (yellow fluorescent protein, YFP) to be located in the cytoplasm and nucleus, while the second one (cyan fluorescent protein, CFP) could be located in the ER, Golgi, and plasma membrane depending on the signal used (de Felipe & Ryan, 2004). On the other hand, when the upstream protein (YFP) bears a Golgi targeting peptide (GT), the downstream protein (CFP) with a nuclear or mitochondrial signal or no signal at all still accumulated inside the ER. This was suggested to be the result of slipstream translocation of the downstream protein rather than a disruption of targeting signal, as removal of the GT from the upstream protein restored the correct distribution of downstream proteins to their intended subcellular compartment. Interestingly, the plasma membrane targeting of the type II transmembrane protein paramyxovirus simian virus 5 hemagglutinin/neuraminidase (SV5HN) was unaffected by the upstream GT containing protein (de Felipe & Ryan, 2004). Their data suggest that the subcellular targeting of proteins can be affected depending on their position in the 2A constructs and the type of targeting signal peptides used.

However, this proposed slipstream mechanism has now been challenged by their new data. In a follow up study (deFelipe et al., 2010), they demonstrated that the previously considered mis-targeting of fluorescent proteins downstream of 2A into ER lumen or in some cases, mitochondria, was in fact due to inactive 2A separation. Their results in HeLa cells suggest that the EYFP sequence immediately upstream of 2A interacts with part of the ribosome exit tunnel and mainly with the ribosome-translocon pore on the ER. This interaction with the translocon complex was proposed to impair the ribosome skipping in the peptidyl transferase
center (PTC), because separation efficiency was preserved in an in vitro translation assay (De Felipe et al., 2010). These conflicting data show that the sequences of upstream proteins and targeting signal peptides used should be taken into consideration when designing polyprotein expressing vectors.

The residual 2A C-terminal extension to the upstream protein and the single amino acid proline addition to the downstream protein also may be a concern to cell biologists due to any unexpected effects posed on the proteins of interest. Indeed, any extra tags are considered undesirable when an intact terminus is required for subcellular targeting or normal protein activities. Several studies have suggested that the residual 2A peptides are tolerable, as it does not affect T-cell receptor and antibody assemblies (Szymczak et al., 2004; Walseng et al., 2015; Fang et al., 2007b), interleukin-12 production (comprising P35 and P40 subunits) (Polypeptide et al., 1999), and shaker potassium channel assembly and function (Figure 5 in Chapter 2).

Nevertheless, the effects of the C terminal 2A extension need to be verified for different proteins used. The addition of a single proline amino acid on the protein downstream of 2A should also not be overlooked, because the N-terminal proline could potentially confer a longer half-life (>20h) to the protein; a feature known as the N-end rule (Bachmair, Finley, & Varshavsky, 1986). The 2A peptide itself might also stimulate an immune response, although no reports have yet demonstrated this effect.

1-11 Strategies to increase 2A separation efficiency or minimize the inhibitory effect from the upstream sequence

Two common solutions have thus far been widely implemented to increase 2A activity. The first strategy is to include an extra 40 aa from the 1D protein into the 30 aa FMDV 2A, as
the 40 aa viral peptide extension does not interact with the translocon pore that affects the 2A activity. Other studies have also demonstrated that the addition of a flexible linker sequence, Gly–Ser–Gly (GSG) or Ser–Gly–Ser–Gly (SGSG), separating 2A from the upstream sequence can improve separation efficiency. Both approaches are suggested to insulate 2A by moving the upstream inhibitory sequence away from the exit tunnel or even translocon pore. Nonetheless, these strategies would undesirably lengthen the residual C-terminal extension on the upstream protein. One proposed solution is the incorporation of a 4 aa furin protease cleavage site, -R-X-R/K-R-, separating the protein coding sequence and the residual 2A peptide, so as to remove the undesirable C-terminal tail proteolytically. Furin is a protease enriched in Golgi. For exocytic proteins it has been shown that their extended 2A-linkers could be trimmed off by the endogenous Furin protease (Fang et al., 2005; Ho et al., 2013). But for non-exocytic proteins, the use of a shorter 2A might be a compromise, as they do not go through ER-Golgi secretory pathway. An alternative approach is reversing the order of encoded proteins in the construct so that the protein bearing sequences inhibitory to 2A separation would be placed after 2A to prevent its influence on 2A activity. In this thesis, we propose another strategy of improving 2A activity by codon-deoptimizing the 2A peptides used. Disfavored codon usage can slow down translation in the host cells. By retarding the progression of ribosome, we reason that the ribosome skipping might function more effectively than in a highly processive manner. Our data show that codon deoptimization improved the separation efficiency in Porcine teschovirus-1 2A (P2A), and to a lesser extent, in Thosea asigna virus 2A (T2A) (Figure 2 of Chapter 2).

1-12 The rationale of CHYSEL-based protein quantification

There are several advantages to using CHYSELS for protein quantification.
• **Small size**

  Its reasonably small size (20-30 amino acids) renders it much easier for molecular cloning procedures than IRES (450-500 bases), especially when the coding capacity of the carrier vector is limited. Using CHYSEL rather than the bulky IRES as the linker can free up more space to accommodate more genes for expression.

• **Stoichiometry**

  The equimolar expression feature ensures a reliable protein quantification, as the intensity of fluorescence can be used to determine the expression level of the protein of interest (Figure 1 in Chapter 2). In an IRES-based vector, the downstream genes are normally expressed in an inconsistent and attenuated level (Sugimoto et al., 1995; Zhou et al., 1998).

• **Autonomous**

  The self-autonomous processing mechanism is insensitive to the cellular status of host cells. CHYSEL activity (separation efficiency) is solely dependent on its interaction with the exit tunnel of the eukaryotic ribosome; whereas in IRES, cellular stresses such as hypoxia and serum-starvation can differentially decrease the translational efficiency of upstream cap-dependent translation but not the downstream IRES-mediated translation (Lang, Kappel, & Goodall, 2002; Stein et al., 1998). This precludes a correct protein quantification using IRES as the two proteins are not translated in a consistent ratio across experimental conditions.

• **Spatial and temporal resolution**

  The fact that the protein of interest and corresponding fluorescent reporter are co-translated allows for subcellular local protein expression measurement in large cells such as neurons in response to stimulation or stresses (e.g., glutamate receptors at dendritic spines, neurotransmitters in axonal terminals). Conducting time series imaging with the use of
photoswitchable fluorescent proteins can read out the production rate of proteins as a function of time and experimental manipulation.

1-13 Issues associated with CHYSEL-based protein quantification

As with all the other assays, there are important caveats when using the CHYSEL technique for protein level measurements. The two most common issues are:

Distinct maturation and degradation rates of different proteins

The distinct differences in protein maturation and turnover rates between the fluorescent reporter and the protein of interest might potentially preclude accurate quantification. An ideal CHYSEL peptide can ensure an equimolar translation of the upstream and downstream proteins, but it cannot guarantee an absolutely equivalent level of the two proteins as the subsequent posttranslational processes such as folding, maturation and turnover rate are independent and not coordinated between them. In other words, we are only certain that any production of a new fluorophore at any given moment will be accompanied by the translation of exactly one molecule of the protein of interest. The fluorescence intensity is the output of mature and functional fluorophores, which may not necessarily reflect the level of functional protein of interest because of their different dynamics in maturation and degradation. The variability in post-translational processing might not be an issue as the differences between the two proteins may simply change the slope of linear correlation (Figure 1b in Chapter 2). On the other hand, distinct processes and degradation rates could complicate quantification as some proteins have fast and tightly regulated degradation time windows, such as transcription factors and circadian genes. However, it has already been shown by Schwanhausser and colleagues using pulse-labelling mass spectrometry that protein degradation accounts for only 5% of the variation in protein levels.
Intracellular protein concentrations are mainly dictated by mRNA levels and translation rate, which explain 40 and 50% of protein variances, respectively (Schwanhäusser et al., 2011). As the only difference between the fluorescent reporter and protein of interest is the posttranslational modification and degradation rate, for the majority of cellular proteins their concentration can be determined by fluorescence intensity with limited error.

**Separation efficiency of CHYSEL peptide**

In biology nothing is 100% efficient, and CHYSEL-mediated separation is likely no exception. Several lines of evidence have shown various levels of incomplete upstream and downstream proteins separation in different organisms with different CHYSEL sequences (Daniels et al., 2014; González et al., 2011; Jin Hee Kim et al., 2011). Studies have shown different CHYSEL peptides exhibiting varying activities when they are expressed in different organisms. For example, T2A from *Thosea asigna* virus has optimal separation efficiency in insect cells (Daniels et al., 2014; González et al., 2011), while P2A from *porcine teschovirus-1* has been shown to have superior activity to other CHYSELS in mammalian cells (Jin Hee Kim et al., 2011). The search for an efficient and universal CHYSEL continues, and is part of the objective of this thesis. Common strategies to improve CHYSEL separation efficiency will be explained in more detailed in Chapter 2.

**Imbalanced production of the two separate proteins**

Although the upstream and downstream proteins from a single polycistronic RNA should theoretically be translated at the same level, in reality, an imbalanced production of upstream and downstream products is often encountered, particularly a molar excess of upstream product over downstream proteins (Michelle L. L. Donnelly et al., 2001). This is believed to be attributed to
the failure to reinitiate translation of the downstream peptide after translation stalling at the end of a CHYSEL, and not caused by differential protein degradation rates between the two translated proteins. In these cases, the ribosome is suspected to fall off after the release of the nascent peptide containing the upstream protein and residual CHESEL sequence (Michelle L. L. Donnelly et al., 2001).

**Different isoforms and posttranslational modifications of proteins**

Different proteins isoforms can be produced through alternative splicing, whereby different mRNA transcripts can be generated from the same genomic sequence through alternative exon selection during posttranscriptional processes. A famous example is the *Drosophila melanogaster* gene *Down syndrome cell adhesion molecule (Dscam)*. Dscam is an essential cell surface recognition molecule involved in axon guidance and dendritic branching (Schmucker & Chen, 2009). In *Drosophila*, Dscam has the potential to generate over 38,000 different protein isoforms through alternative splicing and each cell can express up to 50 different isoforms (Schmucker et al., 2000; Neves et al., 2004). In the CHYSEL dependent quantification approach, the CHYSEL sequences can be placed either at the beginning or the end of the coding region of the protein of interest. Thus, it is unable to differentiate between different isoforms of proteins. Measuring levels of specific isoforms can be accomplished using western blotting with isoform specific antibodies. A limitation similar to isoform detection is the inability of CHYSEL dependent quantification to detect posttranslational modifications (e.g., protein phosphorylation). Many proteins, especially enzymes, exist in either active or inactive form through phosphorylation or dephosphorylation. Studies on these proteins under different cellular statuses often necessitate the quantification of the ratio of active (phosphorylated) versus inactive
(dephosphorylated) states of enzymes rather than the level of total proteins. As the CHYSEL method occurs at the protein synthesis step, downstream modifications cannot be accounted for.
The genome of a typical picornavirus consists of one polycistronic RNA flanked by 5’ and 3’ untranslated regions. During and after translation, the single polyprotein will be processed into two groups of proteins. One of them comprises 4 structural proteins constituting the viral capsid while the other is an assembly of non-structural proteins used for virus replication.
Figure 2, Different viruses adopt different strategies to separate each component from others.

In the case of enteroviruses and rhinoviruses, the encoded proteinases from 2A sequence at region 2 (2A\textsuperscript{pro}) and 3C sequence at region 3 (3C\textsuperscript{pro}) will cleave the nascent proteins at their own N-terminus post-translationally, delineating the junctions of each region. On the other hand, both cardio- and aphthovirus use a nonezymatic cleavage mechanism to separate region 2 apart from region 1 at the C-terminus of 2A. The cleavage at region 3 is proteolytic and is carried out posttranslationally. As opposed from Entero-/Rhinovirus, the 2A sequences of Cardio-/Aphthovirus do not encode any protease for enzymatic cleavage but rather translationally interrupt the peptide bond formation at the C terminus of 2A, which ends up creating discrete products translationally.
Chapter 2.

The quantification of protein levels in single living cells

Introduction

The use of a 2A peptide as a genetically-encoded protein translation reporter can offer unprecedented accuracy in single-cell protein quantification. However, the pitfalls inherent to CHYSELs listed in the previous chapter, such as incomplete separation, imbalanced production of upstream and downstream proteins, and different degradation rates of different proteins, can potentially preclude a reliable protein quantification. Although the stoichiometric production and linear correlation between upstream and downstream fluorescent proteins from a P2A construct has been demonstrated (Goedhart et al., 2011), many more questions remain unanswered. For example, first, all CHYSEL versions used in studies thus far do not separate the N terminal and C terminal proteins consistently across experiments. Any uncleaved fusion products can obscure validation of protein quantification because they are equimolar. Second, although the stoichiometric production of different fluorescent proteins was shown to be consistent in a homogeneous cell line, it is unpredictable whether the linearity of protein levels can still hold between proteins with distinct features such as sizes, number of subunits, maturation time, and turnover rate. Third, the validation of associating cellular phenotypes (sensitive to protein levels) to protein concentration (determined by fluorescence output) would be a considerable advance for cell biology, but the feasibility of this concept has yet to be determined. And finally, a demonstration of its applicability in animal models, for example associating dendritic arbor complexity (phenotype) to the concentration of a transcription factor (protein levels) in single neurons, would further expand the utility of such a protein quantification technique. None of these issues have yet been systematically addressed and the concept of CHYSEL-mediated
protein quantification in living cells needs to be tested. We aim to answer these important questions in this chapter.

The first and foremost step is to optimize the currently used CHYSEL sequences to ensure a complete separation and stoichiometric production of a protein of interest and a fluorescent reporter. Over the past years, several strategies have been proposed and demonstrated with various improvements. Among them, the employment of longer versions of 2A (i.e., the incorporation of partial C terminal sequence of 1D protein into the 2A peptide used) and the addition of a flexible linker GSG or SGSG in front of the 2A sequence are the two most robust solutions. However, the two approaches end up increasing the C-terminal extension on the upstream protein, which is undesirable for functional analyses when a functional C-terminal is required. To address this issue, we sought to develop another strategy that would not only increase the separation efficiency but also keep the CHYSEL peptide short.

The protein separation (ribosome skipping) is preceded by the ribosome stalling at the C terminus of CHYSEL. Although the delay of stalling and resuming is indistinguishable from normal translation (V. A. Doronina et al., 2008a), several factors and reactions are suggested to be involved in the process (V. A. Doronina et al., 2008b). The steric hindrance in the peptidyl transferase center (PTC) prevents ester linkage formation between the nascent peptide and the last amino acid proline. This constraint also triggers the release of the pro-tRNA and then recruits the eukaryotic release factor 1 (eRF1) to hydrolyze and then release the nascent peptide bond in a stop codon independent translation termination (V. A. Doronina et al., 2008; Machida et al., 2014). Our hypothesis here is that if we manipulate the ribosome decoding speed on the mRNA (either speed up or slow down the ribosome), we might in turn change the ribosome
skipping efficiency resulting in a shortened or prolonged stalling. To test this idea, we need a way to adjust the progressivity of ribosome.

Ribosome decoding speed (progressivity) during the elongation phase of translation has been generally accepted to be dependent on the abundance of corresponding cellular tRNA that matches the codons on the mRNA. During peptide synthesis, the sequence of every 3 ribonucleotides of mRNA represents one specific codon that will be recognized by a complementary cognate tRNA charged with the corresponding amino acid. Owing to the degeneracy of the genetic code, 2 to 6 different codons are translated into the same amino acids through different cognate tRNA (a term called synonymous codon). As different organisms contain different pools of cognate tRNA, certain codons are decoded faster than others in the same organisms (codon usage bias) (Novoa & Ribas de Pouplana, 2012). For example, in recombinant protein production in a heterologous context, the expression efficiency of a cloned gene can vary from organism to organism (Gustafsson et al., 2004). By adjusting the codon composition of the cDNA, the expression level of the recombinant protein has been shown to increase 1000-fold (Gustafsson et al., 2004). In other words, an optimized codon usage can increase the ribosome decoding efficiency while an unfavorable codon composition can retard the progression of ribosomes. Here we codon optimized and deoptimized the CHYSEL sequences we used for both mammalian and insect systems and assessed their separation efficiency in their respective cell culture lines.
Materials and Methods

Protein Quantitation Reporter Constructs

Sequences for CHYSEL peptides were tested from Group IV, positive-sense ssRNA viruses, including the Picornaviridae family for 2A peptides, or the Permutotetraviridae family for 2A-like peptides (Jin Hee Kim et al., 2011). For our initial screens, we mostly focused on four broad CHYSEL peptide sequences from the following viruses: Equine rhinitis A virus (E2A), Foot-and-mouth disease virus (F2A), Porcine teschovirus-1 (P2A), and Thosea asigna virus (T2A) and tested for stoichiometric production and separation of fluorescent proteins and Shaker potassium channel. We selected the amino acid sequence ATNFSLLKQAGDVEENPGP from the porcine teschovirus-1 for use in mammalian cells (Jin Hee Kim et al., 2011), and EGRGSSLTCDVVEENPGP from the Thosea asigna insect virus for use in Drosophila cells (Jin Hee Kim et al., 2011). We compared the codon optimized CHYSEL peptides with the original viral CHYSEL peptides, and found that both of them resulted in a significant fraction of un-separated, fusion product (Figure 2). Codon optimization often creates a larger proportion of un-separated product, indicating that codon optimization could be worse for protein quantitation. Thus, we surmised that codon optimization could speed up ribosomal activity causing it to ignore the separation event between the final glycine and proline of the CHYSEL peptide. We tested DNA sequences that were selected for non-favored codons to decrease translation speed, which we found to enhance reliable separation (Figure 2) (Novoa & Ribas de Pouplana, 2012; Zhou et al., 2011). DNA sequence chosen for the PQR in mammalian cells (P2A-derived with glycine and serine linker, codon variation 3) was:

GGAAGCGGAGCGACGAATTTTAGTCTACTGAAACAAGCGGGAGACGTGGAGGAAAACCCTGGACCT. DNA sequence chosen for the PQR in Drosophila cells (T2A-derived with
glycine and serine linker, codon variation 2) was:

GGAAGCGGAGAAGGTCGTGGTAGTCTACTAACGTGTGGTGACGTCGAGGAAAATCC
TGGACCT. We also tested whether extended CHYSEL sequences, 30 amino acids in total length from the separation point, might enhance separation (Luke et al., 2008). We found that extended viral CHYSEL sequences created a proportion of fusion product compared with shorter, 19 amino acid codon de-optimized CHYSEL sequences (Figure 2). Mutated PQR sequences that failed to separate were used as linkers for fusion protein experiments. All viral sequences were generated using gene synthesis into a pUC57 vector (BioBasic, Markham, ON), and cloned into pCAG for mammalian experiments or pJFRC7 for Drosophila melanogaster experiments. GFP, RFP, and BFP constructs were based on superfolderGFP, TagRFP-T, and mTagBFP2, respectively. SuperfolderGFP and TagRFP-T were chosen for their relatively fast maturation times, 6 min and 100 min, and photostability, respectively (Pédelacq, Cabantous, Tran, Terwilliger, & Waldo, 2006; Shaner et al., 2008). For GFP and RFP protein concentration and fluorescence intensity measurements, proteins were purified from E. coli using GFP-specific chromatography columns (Biorad, Hercules, California), and protein concentrations were measured using a Bradford assay with a NanoDrop 2000 (Thermo Fisher). Samples were serially diluted and thin samples were imaged on glass slides to reduce any non-linear effects using a standard fluorescence microscope (see Image Acquisition). ShakerGFP cDNA (R. Blunck, Université de Montréal) and hs-PER::YFP (M.W. Young, Rockefeller University) were kind gifts, and all other plasmids were obtained through Addgene (Cambridge, MA). GFP::RFP fusion proteins were verified using immunoblotting (Figure 2), and imaging experiments verified that these large proteins were excluded from the nucleus.
Cell Culture

HEK293, Neuroblastoma-2A (N2A), and cells were cultured at 37°C under 5% CO₂ in Dulbecco's Modified Eagle Medium (Wisent, St-Bruno, QC) and H-Cell (22c10) (Wisent, St-Bruno, QC), or for *Drosophila melanogaster* S2 and Kc cells, at 25°C in Ex-Cell 420 Medium (Sigma-Alorich). Media for mammalian cells were supplemented with 10% fetal bovine serum (FBS) (Wisent), and 100 units/mL penicillin (Life Technologies, Carlsbad, CA) and 100 μg/mL streptomycin (Life Technologies). Cells were transfected with 5 μg of plasmid DNA in 35 mm dishes using Lipofectamine3000 (Life Technologies). For genome editing experiments, 800 ng of CRISPR-Cas9 plasmid DNA were co-transfected with 800 ng of repair template circular plasmid in 12-well plates. After 2–7 days, cells were non-enzymatically dissociated and seeded on glass coverslips and prepared for imaging and electrophysiology experiments.

Immunoblotting

Immunoblot experiments were performed four times. One billion cells were placed into lysis buffer (25mM HEPES, 150mM NaCl, 1mM EDTA, 1% Triton-X) with SIGMAFAST protease inhibitor tablet solution (Sigma-Aldrich). Protein concentrations were measured using a bicinchoninic acid protein assay (Pierce, Rockford, IL) and 30–40μg of protein was loaded into a NuPAGE Novex 12% Bis-Tris Gel (Life Technologies). Proteins were separated by electrophoresis and transferred to a polyvinylidene fluoride membrane using Invitrogen iBlot dry transfer (Life Technologies). The membrane was blocked in 5% BSA in PBS-T and incubated with the following antibody dilutions: 1:1000 anti-RFP rabbit polyclonal (R10367, Life Technologies), 1:2000 anti-GFP rabbit polyclonal (A6455, Life Technologies), and 1:5000 anti-actin JLA-20 mouse monoclonal (Developmental Studies Hybridoma Bank, Iowa City, Iowa).
Secondary antibodies used were 1:10,000 HRP-conjugated Donkey anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc.) and HRP-conjugated goat anti-mouse IgG (Abcam, USA). All antibodies were dissolved in 5% BSA in PBS-T. Membranes were imaged using the Pierce ECL Chemiluminescence Detection Kit for HRP (Thermo Scientific, USA). The ratios of band intensity of GFP and fusion products were normalized to actin and quantified using ImageJ.

**Image Acquisition**

Fluorescence and brightfield microscopy was performed using a Zeiss AxioScope A1, an Olympus laser scanning confocal microscope FV1000, and a Perkin Elmer UltraView spinning disk confocal Leica DMLFSA microscope. All images were acquired at 512 x 512 pixels using a 40× water objective, N.A. 1.0 (epifluorescence) or 60× oil, N.A. 1.4, or 63× water, N.A. 0.9, objectives (confocal) corresponding to an 215 x 160 μm or 120 x 110 μm field of view, respectively. Fluorescence emission was detected using a charge-coupled device (CCD) camera (MRm) for the Zeiss and (OrcaER, Hamamatsu) Leica microscopes, and photomultiplier tubes for the Olympus microscope. All image acquisition parameters were fixed for each imaging channel for exposure time, excitation intensity, gain, and voltages. Cells that were dimmer or brighter than the fixed initial acquisition dynamic range were not included for analysis. We verified that shifting the acquisition window across fluorescence intensity ranges produced linear correlations throughout the range. In co-transfection of GFP and RFP experiments, cells that were non-fluorescent in either the green or red channel were not imaged; therefore the $R^2$ values for our co-transfection experiments are likely to be overestimates of the true $R^2$ value.
**Image Analysis**

Images were selected for analysis on the basis of identification of single cells and low background. Images presented in all the figures were adjusted for contrast and brightness for clear presentation. Image analysis was performed blind to genotype. Fluorescence pixel intensities were measured in several random regions of interest (ROIs) within the cell using a custom written program in MatLab (MathWorks, Natick, MA) or ImageJ. Average pixel intensities were calculated from three ROIs of 10x10 pixels for measurements within the cytoplasm and nucleus, or from five ROIs of 3x3 pixels for membrane and mitochondrial measurements. For *Drosophila* small lateral ventral neuron analysis, six ROIs of 6x6 pixels were measured from six neurons per lobe, and six animals per time point were chosen. All signal intensities were background subtracted from the average of three 10x10 ROIs surrounding the cell. Cell volume was not taken into account as comparisons were made between either different fluorescence signals within the same cells (Figures 3, 4, 5, and 6) or different intensities of the same fluorescence across cells with comparable sizes (Figures 8, 9, and 10).

**Electrophysiology**

Standard whole cell voltage clamp was used to record potassium currents from HEK293 cells. During recordings, cells were maintained for 1–2 hours at 25°C in extracellular solution consisting of 140 mM NaCl, 10 mM CaCl₂, 5 mM KCl, 10 mM HEPES, and 10 mM glucose at pH 7.4, 319 mOsm. Patch electrodes were pulled from standard wall borosilicate glass (BF150-86-10, Sutter instruments, Novato, California) with 3–5 MΩ resistances. The intracellular pipette solution was 150 mM KCl, 2 mM MgCl₂, 1 mM CaCl₂, 2 mM EGTA, 20 mM HEPES,
and 20 mM sucrose at pH 7.23, 326 mOsm. Whole cell potassium currents were low pass filtered at 5 kHz and measured using an Axopatch 200B amplifier (Axon instruments, Sunnyvale, California), and recorded using a DigiData 1200 with pClamp9 software (Molecular Devices). All pipette and cell capacitances were fully compensated. Cells were held at -80 mV and then given +10 mV steps of 35 ms. The steady-state current elicited at +30 mV was used for analysis. Consistent cell capacitance, and membrane and access resistances were verified before and after recordings.

**Statistical Analysis**

The linear correlation between red and green fluorescence intensities or between red fluorescence and current density were calculated by using Pearson correlation coefficient ($r$). A best fit linear regression line was then generated with a Pearson product-moment correlation coefficient, $r$ value (and a corresponding coefficient of determination, $R^2$ value), to determine the strength of correlation between factors in each condition. The significance of the obtained $r$ was calculated by t-test. To verify that the strength of the linear correlation between GFP and RFP fluorescence from $PQR$ constructs is statistically no different from the fusion of one GFP to one RFP whereas the correlation obtained from the cotransfection of GFP and RFP is significantly distinct from all of them, we calculated the 95% confidence interval of the $r$ values from the fusion positive control. We then used the confidence interval as a standard to evaluate the significances of all other $r$ values obtained from different constructs. For the electrophysiology data, both of the skewed variables, fluorescence intensity and current density, were square root transformed to meet normality assumption for Pearson correlation coefficient ($r$). The $r$ value of
the square root of surface GFP fluorescence versus the square root of current density was used as positive control for generating the 95% confidence interval.

**Drosophila melanogaster Circadian Experiments**

To generate the UAS–RFP-PQR-PER::YFP construct, PER::YFP was amplified from hs-PER::YFP, ligated with the RFP-PQR fragment, and inserted into the pUAST vector. Transgenic fly lines were created using P-element transgenesis (Bestgene Inc, Chino Hills, California). The UAS–RFP-PQR-PER::YFP flies were crossed to the per-Gal4 driver line, P{GAL4-per.BS}3. Crosses were maintained at 25°C in a 12 hour light-dark cycle incubator and newly eclosed F1 progeny were entrained for three days before collection. Six female flies were selected for each time point (6 AM and 6 PM, or zeitgeber time ZT0 and ZT12, respectively). Flies were fixed in 3.7% paraformaldehyde in 0.2 M carbonate-bicarbonate buffer, pH= 9.5 at 4°C for 12 hours. Fly brains were then dissected, mounted on slides through PermaFluor mounting medium (ThermoScientific), and imaged using confocal microscopy.

**Drosophila melanogaster Dendritic Complexity Experiments**

The pJFRC-20XUAS-IVS-RFP<sub>nl5</sub>-PQR-cut construct was created by genomic extraction of the cut coding region from the fly UAS-cut (Grueber, Jan, & Jan, 2003). The cDNA was ligated to RFP<sub>nl5</sub>-PQR, and the resulting construct was cloned into the pJFRC7 vector (Pfeiffer et al., 2010). The transgenic fly w<sup>+</sup>; P{20XUAS-IVS-RFP<sub>nl5</sub>-PQR-cut}attP was created by PhiC31 integrase-mediated transgenesis (Bestgene Inc). Homozygous flies w<sup>+</sup>; P{20XUAS-IVS-RFP<sub>nl5</sub>-
*PQR-cut\attP*, were crossed to homozygous *w*; *; 221-Gal4, *UAS-mCD8::GFP* to selectively express *RFP\_nls-PQR-cut* in class I da neurons. Crosses were maintained at 18°C and wandering third instar larvae were used for imaging. Larvae were dissected in phosphate-buffered saline and the anterior end, gut, tracheal tubes, and fat bodies were removed prior to imaging. Class I ddaE living neurons were imaged using a Fluoview FV1000 confocal laser scanning microscope (Olympus). Neuronal morphology was visualized using the membrane-bound mCD8::GFP and Cut protein levels were determined by ROI analysis of nuclear red fluorescence intensity. Complete dendritic arbors were reconstructed and the number of terminal branches and total dendritic length were computed using the NeuronJ plugin in Fiji.

**Results**

**Creation and verification of Protein Quantitation Reporters (PQRs)**

We first modified and tested different CHYSEL sequences for efficient separation of upstream and downstream proteins in culture cell lines using western blotting. Specifically, we optimized and deoptimized the codon usage at specific residues of P2A and T2A (Jin Hee Kim et al., 2011) and evaluated their performance in human HEK293T and *Drosophila* S2 cells, respectively. Our results showed that codon deoptimized P2A and T2A produce less uncleaved fusion proteins than codon optimized and the original viral peptides (Figure 2). We then chose those activity-optimized CHYSEL sequences for use in mammalian and Drosophila systems and collectively called them *Protein Quantitation Reporters (PQRs)*.
RFP and GFP fluorescence is directly proportional to its protein concentration over several orders of magnitude

To use fluorescence output as a reliable readout for protein quantification, we have to verify that fluorescence intensities of GFP and RFP are directly proportional to their protein concentration as suggested in the literature (Furtado & Henry, 2002). To confirm this, we serially diluted purified GFP and RFP proteins into different concentrations and measured their corresponding fluorescence intensities using standard epifluorescence microscopy. Our results show that both GFP and RFP fluorescence indeed linearly correlate with their protein concentration with coefficients of determination, $R^2 = 0.97$ and $0.99$ for GFP and RFP, respectively.

PQR can produce discrete proteins in a stoichiometric ratio

To verify the equimolar production of the upstream and downstream proteins by PQR (Figure 1a), we quantified the amount of GFP and RFP produced in single HEK293T cells through the measurement of the green and red fluorescence intensities. Since the fluorescence intensity (i.e. brightness) is directly proportional to its protein concentration (Furtado & Henry, 2002) (Figure 1c), the ratio of GFP to RFP fluorescence is equivalent to that of protein amounts. We first quantified green and red fluorescence intensities from cells expressing a fusion protein GFP::RFP, in which one molecule of GFP was attached to one molecule of RFP through a mutated unfunctinal PQR peptide. We obtained a linear correlation between GFP and RFP fluorescence intensities with a coefficient of determination, $R^2 = 0.74$, $t(72)=14.3$, $p<0.001$, 95% CI: $R^2 [0.62, 0.83]$ (Figure 3a, see also materials and methods). The 95% confidence interval of the $R^2$ from the fusion protein GFP::RFP was used as the standard of linearity because the
encoded GFP and RFP are theoretically produced in one-to-one ratio. We then measured green and red fluorescence from cells cotransfected with plasmids encoding GFP and RFP and got a weaker correlation with coefficient of determination, $R^2 = 0.37$, $t(57)=5.77$, $p<0.001$ (Figure 3b). When PQR was inserted between GFP and RFP for coexpression of the two fluorophores, we got coefficient of determination, $R^2 = 0.78$, $t(75)=16.19$, $p<0.001$ for GFP-PQR-RFP and $R^2 =0.66$, $t(75)=12.06$, $p<0.001$ for RFP-PQR-GFP, respectively (Figure 3c, d). The $R^2$ values of GFP-PQR-RFP ($R^2 =0.78$) and RFP-PQR-GFP ($R^2 =0.66$) both fell within the 95% confidence interval of the $R^2$ value of the fusion positive control; while the $R^2$ value of the cotransfection of GFP and RFP ($R^2 =0.37$) was outside of the confidence interval. The linear correlation of RFP and GFP fluorescence in the two PQR constructs is not due to the presence of uncleaved GFP-PQR-RFP or RFP-PQR-GFP fusion proteins as no fusion products were detected in western blot (Figure 7a).

Furthermore, to determine whether PQR can produce independently functioning proteins stoichiometrically, we expressed spectrally distinct fluorescent proteins tagged with different subcellular localization signals separated by PQR linker (YFPmito-PQR-CFPnucleus-PQR-RFPcytoplasm) in HEK293T cells. The three fluorescent genes were co-transcribed into a single polycistronic RNA and subsequently split cotranslationally and targeted to their respective subcellular compartments. Yellow fluorescence from mitochondria, cyan fluorescence from nuclei, and red fluorescence from cytoplasm linearly correlated with one another with corresponding correlation of determination ($R^2$) ranged from 0.54 to 0.69 (Figure 4). The fluorescence intensity from non-expected fluorescence channels were not detected above background signals, confirming that uncleaved fusion proteins were not formed and all the proteins targeted to the appropriate compartments.
PQR can determine gene to phenotype relationship in single HEK293T cells

To demonstrate that PQR can correlate fluorescence output to cellular phenotypes directly proportional to the protein abundance, we measured ion channel density by using whole cell patch clamp electrophysiology. We expressed *Drosophila* Shaker potassium channel, which has one molecule of GFP embedded in the N terminal inactivation domain, attached to RFP through PQR (*ShakerGFP-PQR-RFP*) in HEK293T cells (Batulan, Haddad, & Blunck, 2010). The embedded GFP removed the fast inactivation feature commonly found in wild-type shaker channel so that steady state current can be obtained and used for calculation (*Figure 5a*). Also, the surface green fluorescence output can be used as an internal control for tracking the channel distribution and be used to determine the concentration of shaker channels expressed on the cell surface (*Figure 5a*). We recorded the steady state whole cell potassium current at +30 mv and normalized it to cell capacitance (proportional to cell size) to obtain current density (pA/pF) for each cell (*Figure 5a*). The green surface was imaged by using spinning disc confocal microscopy for obtaining clear cut image while the red cytoplasm was acquired by epifluorescence microscopy because of the insufficient 568 nm confocal laser intensity. The current density was used to correlate with the brightness of GFP or RFP fluorescence in each cell. We first used Pearson product-moment correlation method to generate the best fit regression line for analyzing correlation between GFP or RFP fluorescence intensity and corresponding current density and obtained non-normally distributed residuals in both sets. We then transformed both fluorescence and current variables by square rooting all the data points. The square root of surface green fluorescence linearly correlated with the square root of the obtained current density with a Pearson product-moment correlation coefficient, $r=0.84$ ($R^2=0.70$), $t(23)=7.34$, $p<0.001$, 95% CI: $r[0.66, 0.93]$. This was used as positive control to assess linearity of regression line of
other comparisons. The square root of cytoplasmic red fluorescence and the square root of current density also correlated with a correlation coefficient, \( r=0.71 \) (\( R^2=0.51 \)), \( t(24)=5.00, p<0.001 \), which fell within the 95% confidence interval of the \( r \) value of the positive control. Both groups of residuals resulting from the two regression lines were normally distributed.

**PQR can be used with different microscopy methods**

To demonstrate that protein quantitation can be carried out by using different microscopy methods, we imaged each *shakerGFP-PQR-RFP* expressing HEK293T cells using different excitation methods (e.g., metal-halide lamp, mercury vapor lamp, gas lasers) with the same ROIs in each cell and found linear correlation between Kr 568 nm and mercury lamp excited red fluorescence with correlation of determination (\( R^2 = 0.90 \)).

**PQR can measure protein concentration in single neurons in animals**

**PQR functionality was verified in single clock neurons**

To first demonstrate the applicability of PQR in single neurons, we need an *in vivo* system that produces predictable and quantitative changes in protein levels. To the best of our knowledge, *Drosophila* circadian gene, *period (PER)* is one of the best candidates since it has a well-characterized protein production and degradation cycle. *Period (PER)*, together with other clock genes, control the circadian rhythm through complex transcriptional and translational mechanisms (Hardin, Hall, & Rosbash, 1990, 1992; Reddy et al., 1984). Specifically, in small ventral lateral pacemaker neurons (sLNv), the protein levels of *period (PER)* has a cyclic change in a 24-hour periodicity as it is consistently synthesized, shuttled into the nucleus to feedback regulate its own transcription, and then exported into cytoplasm and degraded in proteasomes. The consistent changes in activity and concentration inside pacemaker cells set up the
endogenous clock (circadian rhythm) of the *Drosophila* brain. To measure the levels of PER in single sLNv neurons, a fusion protein of PER and yellow fluorescent protein (PER::YFP) separated by PQR was expressed specifically in clock neurons through tissue specific transcription factor *period-Gal4*. Six fly brains were dissected at each time point (6am and 6pm) and red and yellow fluorescence from six sLNv neurons were imaged through laser scanning confocal microscopy. The result showed that the perinuclear yellow fluorescence intensity increased and decreased in a 24 hour cycle as PER::YFP is continually produced and degraded in sLNv neurons. On the other hand, RFP has a slightly longer half-life (~26 hours) than PER::YFP (24 hours), thus as it was cotranslated and split from PQR::YFP, the cytoplasmic red fluorescence also cycled at the early time points but accumulated over several days until it saturated the acquisition setting.

**PQR can determine quantitative relationships between protein amount and cellular phenotype in single living neurons**

After demonstrating the applicability of *PQR* in measuring protein levels in animals, the next step is to answer a biologically relevant question: Can we use *PQR* to determine a quantitative relationship between cellular phenotypes and protein concentration in single cells? To test that, a system producing quantifiable molecular processes or cellular phenotypes that are sensitive to the corresponding protein levels is needed. The dendritic arbor complexity of *Drosophila* dendritic arborization (da) neurons is a well-characterized and quantifiable phenotype which has been shown to be regulated by the expression levels of different transcription factors (Grueber et al., 2003; W. Li, Wang, Menut, & Gao, 2004; Sugimura, Satoh,
Estes, Crews, & Uemura, 2004). Da neurons are a family of four different sensory neurons (Class I to IV), classified according to the increase of dendritic arbor complexity (Grueber, Jan, & Jan, 2002). The homeodomain-containing transcription factor cut has been implicated in regulating dendritic arbor complexity in a protein level-dependent manner, in which class I, II, IV, and III have undetectable, low, medium and high Cut protein levels, respectively. The ectopic expression of cut in class I da neurons has been shown to promote dendritic branching that leads to complex dendritic arbors typically seen in class III or IV da neurons (Grueber et al., 2003). However, it is still unclear how ectopically expressed Cut induces the outgrowth of dendritic arbor, such as in a dose-dependent manner, or perhaps as a simple binary on-off switch.

Here, we sought to determine the relationship between dendritic arbor complexity and the ectopic Cut protein concentration in living class I da neurons through the employment of PQR. The class I-specific driver 221-Gal4, UAS-mCD8::GFP to selectively express UAS-RFPnls-PQR-cut in class I da neurons, in which the expression level of endogenous cut is undetectable (Figure 9a). Wandering third instar larvae were dissected and GFP and RFP positive class I ddaE live neurons were imaged in situ. The membrane-bound mCD8::GFP was used for visualizing and evaluating dendritic arbor complexity and the nuclear red fluorescence output was used to determine cut protein expression levels. The data show that dendritic arbor complexity (evaluated by the number of dendritic terminals or total dendritic length) increased as the growth of red fluorescence output in a logarithmic function ($y = 7.7\ln(x) + 22.9$ and $y = 203.7\ln(x) + 2025.7$, respectively) until the Cut-induced dendritic branching effect was saturated (Figure 9b). Collectively, it appears that cut regulates dendritic arbor complexity of class I da neurons in a level-dependent manner.
**Discussion**

Protein quantitation ratioing (PQR) is based on the equimolar expression of a protein of interest and a fluorescent reporter that allow for protein quantitation through the output of fluorescence (i.e. brightness) in single living cells (Figure 1a,b). Theoretically, the quantification is taken at the level of translation, any downstream processes such as folding, maturation, posttranslational modification, and turnover rate can potentially affect the protein level measurement based on the fluorescence output of mature fluorophores. The variation in protein maturation and posttranslational processes can change the slope of the linear relationship between fluorescence and the levels of the protein of interest (Figure 1b), but the fluorescence is still proportional to the protein concentration. Effectively, for the majority of endogenous proteins, their concentrations have been shown to be mainly controlled by mRNA levels and translation rate; only a small fraction of contribution (~5%) from protein stability and degradation (Schwanhäusser et al., 2011). Our result with *Drosophila* Shaker potassium channel is in agreement with the assumption and finding in literature. Shaker potassium channel is a homo-tetrameric channel which comprises four molecules to form a single functional channel. Its level on the cell surface is very dynamically regulated with a turnover rate of several days, but has complex and comparatively rapid internalization and insertion rates on timescales of minutes to hours (de Souza & Simon, 2002; Jugloff, Khanna, Schlichter, & Jones, 2000; Zhao, Sable, Iverson, & Wu, 1995). The RFP used is TagRFP-T, a monomeric fluorophore with half-life of 26 hours (Corish & Tyler-Smith, 1999; Shaner et al., 2008). The linear correlation between cytoplasmic RFP and surface shakerGFP fluorescence intensities suggests that protein levels are insensitive to those contrasting posttranslation processes, and demonstrates that PQR can be used to quantify protein with complicated maturation process and long turnover rate. (Figure 5d).
Nevertheless, for certain proteins such as transcription factors, cell cycle regulators, and circadian proteins, in which tightly regulated transcription, translation, and even degradation are required to keep the protein levels in check, the resulting deviation caused by the contrasting degradation rates between the fluorophore and the protein of interest could be more pronounced than expected. In the case of the clock gene period (PER), the mRNA levels at some point can drop to the level at which protein synthesis is completely shut down for maintaining the normal circadian rhythm. In this short period of time, the roles of mRNA levels and translation rates in determining the protein concentration diminish and could be overtaken by the degradation rates. This effect manifested well in our circadian experiment (Chapter 3) in which a cyclic change of PER::YFP levels within a certain fixed range was strikingly distinct from the continual accumulation of RFP over several days of experiment. This paradoxical result explains the fact that the “5% error rule” in concentration determination concluded by Schwanhäusser et al. should not be generalized to all other proteins in any cell types, particularly those with certain characteristics such as a temporally regulated production and turnover (Circadian proteins, cell signalling proteins, transcription factors, and proteins with cell cycle specific functions) or stimulated insertion and internalization (ion channels, surface signal receptors). An important caveat to note when interpreting the result of the marginal (5%) predicting power of degradation rate observed by Schwanhäusser et al. is that the whole experiment was performed on a population of exponentially growing, nonsynchronized, and unperturbed cell lines. The obtained data set represents average values from a population of cells, which cannot be translated into single-cell condition as certain unique characteristics such as heterogeneity and dynamism are often masked in a population level. Moreover, the mRNA levels and translation rates in those proliferating cells are expected to be continuously elevated to meet the demand of proliferation.
In contrast, when measuring the protein levels of a tightly regulated clock gene in single synchronized and non-dividing clock neurons, the effect of degradation augments when the protein production ceases. That is to say, given that quantification of the contribution of protein production and turnover rates to steady state protein levels is based on a population scale, the conclusion from Schwanhäusser et al. might not be supported by data obtained from individual cells. In comparison, when pCAG-shakerGFP-PQR-RFP was expressed in HEK293T cells, the correlation between GFP and RFP fluorescence intensities demonstrate that the error caused by the contrasting turnover rates is negligible. This marginal influence of turnover rates in protein level determination may be attributed to the use of a constitutively active promoter (pCAG) in driving the expression of RFP and shakerGFP, in which the continuously maintained mRNA levels and translation rate dictate the protein concentration determination as concluded by Schwanhäusser et al.

**Codon deoptimization improves the separation efficiency**

Our western blot results show that in P2A, the codon deoptimized (i.e., can retard translation) protein quantitation reporter has better activity than even the longer version (30aa) of original viral P2A in HEK293T cells. The 30 aa version of F2A has been shown to possess better separation efficiency as it might better interact with the exit tunnel of ribosome. Similar improvement has been also observed in other 2A-like sequences (Luke et al., 2008). Thus the 30 aa version of viral P2A was expected to have the optimal separation since the working principle of different CHYSELS has been suggested similar (conserved C terminal residues: DxExNPGP) (Luke et al., 2008). The result of our data with which P2A separation efficiency also depends in
part on codon usage merits another topic of investigation as no one else has yet reported the effect of codon usage in CHYSEL activity. Whether it is simply due to the ribosome decoding speed or other mechanisms is currently unknown and is beyond the scope of this thesis. More 2A and 2A-like sequences need to be included for analysis to see if it is a generic effect or only specific to certain versions of CHYSELS, as the activity improvement is less pronounced in T2A in *Drosophila* S2 cells (Figure 2a). But given that the separation in 20 aa T2A is already superior (~99%) to many other versions of 2A (Luke et al., 2008), the marginal enhancement in T2A activity cannot be as perceivable as others. All of the PQR constructs we used in this thesis had been verified in western blotting and no uncleaved fusion proteins were detected (Figure 7).

Our data show that *PQR* can produce two fluorophores with linear correlation in fluorescence intensity indistinguishably from a fusion positive control and significantly better than the cotransfection of *GFP* and *RFP* (Figure 3a, b, c, and d). Cotransfection of one plasmid encoding a fluorescent reporter and the other encoding the protein of interest is a common assay for tracking transfected cells and the brightness of fluorescence is often used to qualitatively determine the target protein expression level. Our data demonstrates that the use of a cotransfected reporter is not a reliable measurement for the expression level of the target protein (Figure 3b). The robustness of PQR in producing independently functioning proteins is further demonstrated in HEK293T cells co-expressing multiple fluorophores. Three separate fluorescent proteins were produced stoichiometrically and located to the destined subcellular compartments (Figure 4).
PQR can relate cellular phenotype as a function of protein concentration

We furthermore confirm the utility of PQR to studying gene-to-phenotype relationships in single living cells by measuring ion channel concentration and comparing it to fluorescence intensity. The linear relationship between red fluorescence (protein level) and current density (phenotype) was statistically indistinguishable from the linearity between the embedded GFP fluorescence and current density, which is supposed to be produced in a 1 to 1 ratio (Figure 5b, e). This proof of concept experiment demonstrates that PQR can relate cellular phenotypes (current density) to the levels of protein concentration (RFP fluorescence) (Figure 5c). Also, the RFP fluorescence images were acquired through epifluorescence microscopy, but the linear relationship between Kr 568 nm laser versus mercury lamp excited red fluorescence allows the use of different microscopy methods for image acquisition (Figure 6).

The study of gene to phenotype relationship in transgenic animals

The circadian gene period (PER) is a gene with very tight temporal regulation in transcription, translation, and degradation that results in its predictive and consistent protein level variations at different time points during the day (Figure 8b). We measured RFP and YFP fluorescence to see the correlation between the regular PQR fluorescent reporter RFP and the exogenous circadian protein PER::YFP in a time-series experiment in which the value at each time point represents the average measurement of six animals. Clearly, at the early time points, the accumulation of RFP integrates the cyclic production of PER::YFP until the RFP fluorescence saturates the acquisition setting because the degradation of RFP is not temporally regulated. However, the cyclic production of PER::YFP at late time points can be read out by
resetting the acquisition parameter for the RFP fluorescence (Figure 10). This demonstrates the robust sensitivity of PQR technique in measuring protein level changes in single neurons. None of the correlation was contributed by the fusion RFP-PQR-PER::YFP as uncleaved product is nonexistent (Figure 7c). The lack of pronounced fluctuation in YFP at late time points (Day 9) might be attributed to the less precise regulation in circadian rhythms as animals aged (Driver, 2000).

One of the challenges of PQR is studying protein level-sensitive phenotypes in which a subtle change of low abundant proteins will induce substantial phenotypic variation, such as the transcription factor Cut-induced dendritic branching in class I da neurons. To overcome these issues, we sequestered the red fluorescent protein into nuclei so that the originally unresolvable differences in protein levels become distinguishable. The logarithmic relationship between RFP fluorescence (cut protein levels) and the dendritic arbor complexity (phenotypes) is in agreement with the assumption that low levels of transcription factor can cause profound effects.

Practical perspective about the use of PQR in biology

Protein quantitation ratioing (PQR) can be a convenient tool to relate cellular phenotypes as a function of protein concentration; as a result, many protein level-mediated cellular functions can be precisely delineated at single-cell levels. PQR is a fast and easy assay since the protein levels can be determined by measuring the fluorescence output through standard fluorescence microscopy. It has unprecedented resolution and sensitivity as we can measure protein concentration in single cells. It has potential utility for in vivo quantification as the process of protein quantitation is non-invasive and can be conducted in living cells.
More precise spatial and temporal measurements of protein kinetics can be obtained through the use of photoswitchable fluorescent proteins to allow for subcellular activation and quantitative imaging of newly synthesized fluorescent molecules. Through the use of fluorescence correlation spectroscopy (FCS), even subtle changes in absolute fluorescent protein levels at a given subcellular compartment can be measured (Kim et al., 2007; Levin & Carson, 2004). Although the very nature of PQR, or CHYSEL in general, is a protein translational reporter but not a direct protein level quantification assay, the accuracy and resolution it offers is still overwhelmingly better than any other currently available options for gene expression measurement such as qPCR, protein assay, or western blotting as they are indirect, lack single-cell resolution, and destructive to samples. Unlike other CHYSEL sequences, all of the PQRs we used in all the experiment do not produce uncleaved products after several rounds of activity optimization (Figure 7). To further remove the residual PQR appendages from the C terminus of the upstream protein, the inclusion of a furin cleavage site has been adopted. But since endogenous cellular furins exist predominantly in the Golgi apparatus, the cleavage of the residual PQR peptide might not be effective for proteins not going through secretory pathway. In our hands, when a furin cleavage site is added to the end of shakerGFP in order to cleave off the PQR peptide, the production of shakerGFP and downstream RFP is both attenuated through a still uncharacterized mechanism. This downregulated expression might be attributed in part to the interference on PQR activity, since the addition of a furin cleavage site has been reported to impair 2A function in HEK293T cells (Verrier et al., 2011). The fusion protein resulting from the ineffective 2A activity, shakerGFP-furin-PQR-RFP, may be degraded immediately after production, as numerous fluorescent puncta accumulated inside the cells. To address this issue, one can place the gene of interest after PQR or employ a shorter version of the linker (e.g., using
the 18 aa PQR as opposed to a 30 aa CHYSEL) if an authentic C-terminus is required. Overall, we believe that PQR can shift the current paradigm of protein quantification from inaccurate and laborious protein assays or western blotting to a more precise and easy procedure.
Figures.

Figure 1, Protein Quantitation Ratioing (PQR) can determine relative protein concentration in single living cells.

**a**

Stoichiometric protein translation can quantitate protein amounts. Insertion of a Protein Quantitation Reporter (PQR) between a fluorescent reporter (GFP) and a gene of interest creates a polycistronic mRNA for co-transcription and co-translation of GFP and the gene of interest. The PQR construct allows for one molecule of GFP to be synthesized for every one protein of interest synthesized. Because the fluorescence output of GFP is directly proportional to the concentration of GFP, then the fluorescence intensity of a cell can be used to quantitate the concentration of the protein of interest.
\[ F(\text{GFP}) \propto [\text{GFP}] \]

\[ [\text{Gene}] \propto [\text{GFP}] \]

\[ [\text{Gene}] \propto F(\text{GFP}) \]
b, Linear relationships between fluorescence output, fluorescent protein concentration, and protein of interest concentration allow for Protein Quantitation Raitoing. Because the fluorescence output of GFP is directly proportional to its concentration (top panel and c), then using a PQR will produce a stoichiometric ratio between GFP and the protein of interest (middle panel), therefore enabling the fluorescence intensity of GFP to be used as a measure of the protein of interest concentration (bottom panel). Any (linear) differences in post-translational processing, maturation, or insertion rates of the protein of interest or GFP will change the slope of the relationship (dotted gray lines). For example, if at steady state there are 11 functional molecules of a Shaker K+ channel for every 41 functional molecules of GFP, the relationship will still be linear. Importantly, protein concentration is predominantly controlled by translation, with very small contribution from protein stability and degradation.
GFP Fluorescence Intensity vs Concentration

$R^2=0.97$

RFP Fluorescence Intensity vs Concentration

$R^2=0.99$
The fluorescence intensity of GFP and RFP is a linear function of its concentration over five orders of magnitude. Purified GFP (top panel) and RFP (bottom panel) were imaged using standard fluorescence microscopy. Pixel intensities are plotted in arbitrary units (a.u.) in log10. Coefficient of determination, R2 values from a simple linear regression model were calculated from the averages of five experiments. Error bars are S.D.
Figure 2. Wildtype viral CHYSEL sequences produce un-separated fusion proteins; PQR sequences produce reliable separation of proteins. We modified and synthesized different viral CHYSEL sequences to screen for separation efficiency using immunoblots (representative examples shown) and stoichiometric production of proteins using quantitative imaging. Anti-GFP antibody was used to detect GFP (middle blots) versus fusion product of unseparated RFP and GFP (top blots). Anti-Actin (bottom blots) was used to normalize pixel intensities of fusion product (numbers underneath top blots). We added glycine and serine N-terminus linkers to all synthesized CHYSEL sequences, for example on the 2A-like sequences from *Thosea asigna* virus (T2A) and to the 2A sequences from *Porcine teschovirus-1* (P2A).
a, wildtype T2A viral codon usage or codon optimization produces fusion protein production, whereas codon de-optimization enhances separation efficiency. Separation efficiency for each CHYSEL construct was tested using immunoblotting of RFP–CHYSEL–mCD8::GFP constructs transfected into Drosophila S2 cells for T2A-derived sequences. Manipulation of the T2A peptide sequence by adding glycine and serine linkers still produced a large fraction of fusion protein (arrowhead in Lane 2, “viral” T2A). When we turned to manipulating codon sequence usage, we found that codon optimization produced equivalent or worse amounts of fusion products (arrowhead in Lane 5, T2A variant 3, 100% codon optimized) compared to the viral CHYSEL sequence, along with diminished amounts of separated mCD8::GFP. Codon de-optimization of specific amino acids reduced the proportion of fusion product (Lane 3, T2A variant 1 is 60% codon de-optimized and Lane 4, T2A variant is 45% codon de-optimized). T2A variant 2, 45% codon de-optimization (asterisk), produced close to the background levels of the untransfected S2 cells lane. T2A mutant constructs (Lane 6) that produced fusion products were used as positive controls. Numbers at the left side of the upper blot indicate protein size in kDa.
b. Codon de-optimization of specific CHYSEL residues produces reliable separation of proteins. We used HEK293 cells to test codon de-optimization of different CHYSEL residues using RFP–CHYSEL–GFP constructs derived from P2A sequences. We found that ~50% codon de-optimization of sequences (Lane 6, P2A variant 3), without altering the final four codons, allows for greatest separation efficiency. P2A variants 1, 2, and 3, changing 100%, 80%, and 50%, respectively, of the codons (except for the last 4 codons) produced decreasing amounts of fusion product and increasing amounts of separated GFP. All codon percentage change numbers do not include the glycine serine linker codons, which were required in all constructs (including
“viral” sequences) to avoid large amounts of fusion products within the proteasome. Thus, using viral CHYSEL sequences (both 19 aa and 30 aa) will not work as Protein Quantitation Reporters, as these sequences leave a large fraction of uncleaved fusion protein product (arrowheads) that will contaminate any results of quantitation, and any experiments where fusion products are undesirable.
**Figure 3**, Protein Quantitation Reporters can correlate fluorescence intensity with protein amount.

**a**, Red and green fluorescence intensities of HEK293 cells expressing a fusion protein of GFP and RFP (GFP::RFP) were linearly correlated with a coefficient of determination, $R^2 = 0.74$ ($n = 74$ cells, $P < 0.001$).

**b**, Co-transfection of GFP and RFP produced a weak correlation between fluorescence intensities ($n = 59$ cells, $P < 0.001$).
c, d. Insertion of a *PQR* between GFP and RFP produces red and green fluorescence intensities that were linearly correlated. $R^2$ values for *GFP-PQR-RFP* (c) and *RFP-PQR-GFP* (d) ($n = 77$ cells for both) were not significantly different from the fusion protein data ($P > 0.05$).
Figure 4, Protein Quantitation Reporters can produce discrete and functional fluorescent proteins with correlated fluorescence intensity.

a, Three different fluorescent proteins expressed from a single PQR construct in HEK293T cells were located exclusively to the appropriate subcellular compartments. We used mitochondrial (mito) and nuclear (nls) localization signals on YFP and CFP, respectively, with RFP remaining cytosolic.

b–d, Image analysis of HEK293T cells expressing $YFP_{\text{mito}} PQR - CFP_{\text{nls}} PQR - RFP$ demonstrates linear correlations between fluorescence intensities within different subcellular compartments. Fluorescence intensities within the nucleus compared to the cytoplasm were consistently the most highly correlated (a).
Figure 5, PQR fluorescence intensity is linearly correlated with potassium channel current density.
**a**, **b**, $K^+$ channel current density was linearly correlated with green fluorescence intensity in cells expressing the Shaker $K^+$ channel fused to GFP, with a coefficient of determination, $R^2 = 0.73$ ($n = 28$ cells, $P < 0.001$). Whole cell patch clamp recordings were performed on HEK293 cells and voltage steps of $+10$ mV were given to generate an I-V curve (e). Steady-state current was measured at $+30$ mV and current density (pA/pF) was calculated using the membrane capacitance. Scale bar is 25 µm.

**c**, **d**, Red fluorescence intensities were correlated with $K^+$ channel current density and green fluorescence in cells expressing *ShakerGFP-PQR-RFP*. $R^2$ values for current density to RFP, and GFP to RFP were not significantly different from the current density to GFP positive control data ($P > 0.05$). These correlations were not due to unseparated RFP fusion products since green fluorescence was restricted to the membrane, and red fluorescence remained cytoplasmic (images in e). All fluorescence intensities are plotted in arbitrary units (a.u.).
**Figure 6.** Different fluorescence microscopy methods can be used with PQR

**Red Fluorescence Intensities from Kr Laser Versus Hg Lamp Excitation**

\[ n = 46, \quad R^2 = 0.90 \]

Fluorescence output remains linear between different excitation sources and microscopy methods. HEK293 cells expressing *ShakerGFP-PQR-RFP* were imaged using a spinning disk confocal microscope. Red fluorescence intensities were highly correlated between excitation methods using a Kr 568nm laser and Hg lamp with \( R^2 = 0.90 \) (\( n = 46 \) cells, \( P < 0.001 \)). All excitation methods, fluorophores, and microscopy methods tested produced \( R^2 \) values \( \geq 0.90 \) (representative example shown).
Figure 7, No fusion products produced from all PQR constructs

(a) HEK293 Cells

(b) HEK293 Cells

(c) Kc Cells

No RFP\textsuperscript{Nh}-PQR-cut fusion

No RFP-PQR-PER\textsuperscript{::}YFP fusion

RFP\textsuperscript{Nh}
Western blots were performed on all PQR constructs used in experiments to demonstrate the absence of fusion protein products confounding our analysis. (a) HEK293 cells expressing fusion protein $\text{GFP::RFP}$, $\text{GFP}$ and $\text{RFP}$ plasmids co-transfected, $\text{GFP-PQR-RFP}$, and $\text{RFP-PQR-GFP}$ were lysed 5 days after transfection. Collected protein content was analyzed using anti-GFP antibody. Cells that were transfected with either $\text{GFP-PQR-RFP}$ or $\text{RFP-PQR-GFP}$ produced low or undetectable amounts of fusion product (upper arrow). (b) HEK293 cells expressing Shaker$\text{GFP-PQR-RFP}$ were analyzed in Western blots using anti-RFP antibody, and no fusion product was detected. (c) Kc cells expressing $\text{RFP-PQR-\text{PER::YFP}}$ and $\text{RFPnls-PQR-cut}$ were analyzed in Western blots using anti-RFP antibody. No fusion proteins were produced.
Figure 8, PQR can relate cellular phenotype as a function of protein concentration.
a, PQR can detect cyclic increases in protein concentration over time. *RFP-PQR-PER::YFP* was used to quantitate changes in PER transcription factor levels in single neurons in the animal. An image of the *Drosophila* brain is shown with RFP and PER::YFP expression restricted to the small lateral ventral neurons (dotted box and right panels) using *Per-Gal4* to drive *UAS–RFP–PQR-PER::YFP*. Red fluorescence within the neurons remained in the cytoplasm, and yellow fluorescence was peri-nuclear. Scale bars are 100 µm (left panel) and 10 µm (right panels).

b, Red fluorescence increased cyclically in neurons over time scales of days. Flies were entrained on a 12 hour light-dark cycle and red and yellow fluorescence intensities were measured within single neurons at zeitgeber time 0 (sun symbol) and 12 (moon symbol) (*n* = 6 cells/6 animals/time point). Yellow fluorescence intensities cycled every 24 hours without accumulating beyond a fixed value, reflecting the rapid lifetime of PER. Red fluorescence intensities were also cyclical, but gradually increased over several days, reflecting the integrated amount of PER produced over time. Error bars are S.E.M.
Figure 9, PQR can relate cellular phenotype as a function of protein concentration.
a, PQR in single living neurons is used to quantitatively relate dendritic complexity with Cut protein levels. Dendritic complexity of *Drosophila* da neurons is regulated by the transcription factor Cut. Wildtype class I da neurons (left panel) have relatively simple dendritic arbors. Expression of *RFP*<sub>nls</sub>-*PQR-cut* within class I neurons increases dendritic branch number and total dendritic branch length (middle and right panels). Red fluorescence within the nucleus (inset in middle and right panels) reflecting Cut protein levels indicates that Cut controls dendritic growth in a concentration dependent manner. Posterior is up and dorsal to the right in all three panels. Scale bar is 30 µm.

b, Dendritic complexity is logarithmically dependent on Cut protein concentration. The average number of dendritic branch terminals and total dendritic length in wild-type neurons is indicated by the solid grey lines (± 1 S.D., dashed lines) in the left and right panel, respectively.
Red fluorescence intensity values cycled in phase with yellow fluorescence at Day 5 and 6 when measured with a lower acquisition setting. Small lateral ventral neurons in the *Drosophila* brain expressing *Per-Gal4* to drive *UAS–RFP-PQR-PER::YFP* were analyzed for yellow and red fluorescence.

**Table 1, CHYSEL sequences used in the study**

<table>
<thead>
<tr>
<th>CHYSEL peptides</th>
<th>CHYSEL DNA sequences</th>
<th>CHYSEL Oligo Peptide Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>vT2A 18aa</td>
<td>GAGGGCGAGAGGAGATCGCTAAACATGGCCGTGAGGAGAGAATCCCGGACCT</td>
<td>EGRGSSLTCGDEVENPGP</td>
</tr>
<tr>
<td>T2A Variant 1</td>
<td>GAGGGCGAGGAGATCCGTGACATCGGTGAGGAGAATCCCGGACCG</td>
<td>EGRGSSLTCGDEVENPGP</td>
</tr>
<tr>
<td>T2A Variant 2</td>
<td>GAAGGCTGAGGTGAGCTAAGCTGAGGAGAATCCCGGACCT</td>
<td>EGRGSSLTCGDEVENPGP</td>
</tr>
<tr>
<td>T2A Variant 3</td>
<td>GAAGGCTGAGGTGAGCTAAGCTGAGGAGAATCCCGGACCT</td>
<td>EGRGSSLTCGDEVENPGP</td>
</tr>
<tr>
<td>T2A Mutant</td>
<td>GAGGGCGAGAGGAGATCCGACATCGGTGAGGAGAATCCCGGACCT</td>
<td>EGRGSSLTCGDEVENAP</td>
</tr>
<tr>
<td>vP2A 30aa</td>
<td>GCTATGACTGTTGAGTGGCATTACAGGAGCGGGAGGAGGAGAATCCCGGACCT</td>
<td>AMTVMAFQGPAGANFSLLKQAQAGDEVENPGP</td>
</tr>
<tr>
<td>vP2A 19aa</td>
<td>GCTACTAATTCAAGCCTGAGCAGGAGGAGGAGGAGAATCCCGGACCT</td>
<td>ATNFSSLLKQAQAGDEVENPGP</td>
</tr>
<tr>
<td>P2A Variant 1</td>
<td>GCGAGCAATTGCTAATGAGGAGGAGGAGGAGAATCCCGGACCT</td>
<td>ATNFSSLLKQAQAGDEVENPGP</td>
</tr>
<tr>
<td>P2A Variant 2</td>
<td>GCGAGCAATTGCTAATGAGGAGGAGGAGGAGAATCCCGGACCT</td>
<td>ATNFSSLLKQAQAGDEVENPGP</td>
</tr>
<tr>
<td>P2A Variant 3</td>
<td>GCGAGCAATTGCTAATGAGGAGGAGGAGGAGAATCCCGGACCT</td>
<td>ATNFSSLLKQAQAGDEVENPGP</td>
</tr>
<tr>
<td>P2A Mutant</td>
<td>GCTACTAATTCAAGCCTGAGCAGGAGGAGGAGGAGAATCCCGGACCT</td>
<td>ATNFSSLLKQAQAGDEVENAP</td>
</tr>
</tbody>
</table>
Chapter 3.

The quantification of endogenous proteins in single living cells

Introduction

In the previous chapter, we demonstrate the utility of the protein quantitation ratioing (PQR) technique in single-cell protein level measurement both in cell lines and animals. In this chapter, we further extend the applicability to quantify endogenous proteins by inserting protein quantitation reporters into genomic loci.

Why quantify endogenous proteins?

Normally, genes are expressed from their genomic loci and regulated by the promoters and nearby enhancers. The interpretation of results from expressing an exogenous gene from a plasmid (Shaker potassium channel in HEK293T cells) or an extra copy of a gene integrated somewhere in the genome through transgenesis (period or cut flies) is complicated by the fact that different regulatory pathways are used to control gene expression or even protein kinetics. Indeed, ectopic expression of genes from recombinant plasmids often leads to non-physiological levels, like the overexpression of shaker potassium channels in HEK293T cells. Even if only a single copy of a gene is integrated into a random or specific locus, the new promoter, nearby enhancers, or silencer elements cannot guarantee the native function of the gene as it would be expressed from its endogenous locus. Therefore, to decipher gene-to-phenotype relationships, we have to measure proteins expressed from their endogenous loci.
Targeting *PQR* into genomic locus through the CRISPR-Cas9 gene editing technique

To measure endogenous protein expression, we have to integrate *protein quantitation reporters* (*PQRs*) into their genomic loci. For a given genome, *PQR-RFP* is knocked in to the end of the protein coding region right in front of the stop codon. As a result, the endogenous gene and *PQR* are co-transcribed by the endogenous promoter into a single polycistronic RNA with regulatory elements such as 3’ untranslated region (UTR) that preserve its function. The precise integration is attained by using the Regularly Interspaced Short Palindromic Repeats-Cas9 (CRISPR-Cas9) gene editing technique. Specifically, a custom-designed guide RNA (gRNA) guides the endonuclease Cas9 to make a double-strand break at a predetermined genomic locus. DNA double-strand break is repaired through either the error prone non-homologous end joining or homology-directed repair (HDR). HDR is based on recombination with the homologous chromosome or an exogenous template with homologous sequences. *Protein quantitation reporters* can be inserted into the genomic locus through HDR with an exogenous repair DNA template containing 5’ and 3’ homology arms flanking *PQR*. Common caveats associated with CRISPR-Cas9 (e.g., off-target effect) can be avoided by employing the double Cas9 nickase strategy that nicks DNA at two different targets to increase the gene-editing specificity. Another solution is to use a truncated gRNA (17 nucleotides complementary sequence rather than the commonly used 20 nucleotides) which has been shown to decrease off-target mutagenesis by 5,000 fold without sacrificing the on-target genome editing efficiencies (Fu, Sander, Reyon, Cascio, & Joung, 2014). The concern of repetitive cutting after HDR can be obviated by using donor DNA with mutated protospacer adjacent motif (PAM) sequences that render the new knock-in sequences unrecognizable by CRISPR-Cas9.
**Ribosomal Protein L13A (RPL13A)**

Ribosomal protein L13A (RPL13A) is one of the many ribosomal proteins composing the 60S subunit. It is a widely used reference gene in RT-qPCR thanks to its consistently moderate to high expression levels (Mane, Heuer, Hillyer, Navarro, & Rabin, 2008). We chose to insert a PQR and measure the expression level of ribosomal protein L13A (RPL13A) for numerous reasons: First, it has a relatively more constant expression level and can be used as an internal standard for protein level measurement. A consistent RPL13A fluorescence from a second channel (e.g., RFP, BFP_{nls}) would allow for normalization across cells or experimental conditions and for correction for optical effects such as spherical aberrations, optical distortions, and imaging depths, often encountered in *in vivo* experiments. Second, it has been shown that RPL13A is dispensable for the functionality of the ribosome so that any interference imposed on it through the PQR integration might not affect the physiology of host cells (Chaudhuri et al., 2007). Finally, it is known that ribosomal proteins possess several extra-ribosomal functions such as transcriptional and translational regulation; thus the expression levels of ribosomal proteins can be used to normalize for global cellular transcriptional and translational levels.

**Creation of RPL13A-PQR-XFP knock-in Drosophila**

A knock-in animal with *PQR-XFP* integrated into the Ribosomal Protein L13A (*RPL13A*) locus can be an invaluable resource for biologists. It allows the quantification of RPL13A expression in any cell type and developmental stage through fluorescence measurement. The RPL13A fluorescence represents the relative levels of RPL13A protein expression which can be used as a reference for normalization in *in vivo* imaging and cellular transcriptional and translational states. The knock-in *PQR-XFP* can be stably maintained in the *RPL13A* locus and
passed on from generation to generation provided that the knock-in is genetically stable and heritable. In comparison with mice, the generation of knock-in Drosophila lines can be more efficient and cost effective owning to the short life cycle and numerous genetic tools. Port et al. 2014 demonstrate the employment of the transgenic gRNA fly strategy as a robust protocol in germline and somatic genome editing creating. Specifically, a fly constitutively expressing a customized gRNA is crossed to another fly producing the Cas9 nuclease either in a ubiquitous or tissue restricted manner. Subsequently, the combination of gRNA and Cas9 forms the active CRISPR-Cas9 complex in the offspring to perform genome editing at predetermined loci. Their results show that a Cas9/gRNA expressing animals produce the most efficient biallelic editing in various somatic tissues and the highest germline transmission rate comparing with other protocols (Port et al., 2014).

Here in this chapter, we demonstrate the use of this transgenic gRNA strategy to create different RPL13A-PQR-XFP knock-in Drosophila lines, and experiments that characterize them.

Materials and Methods

Cell Culture

HEK293 and Neuroblastoma-2A (N2A) cells were cultured at 37°C under 5% CO₂ in Dulbecco's Modified Eagle Medium (Wisent, St-Bruno, QC) and H-Cell (22c10) (Wisent, St-Bruno, QC), or for Drosophila melanogaster S2 and Kc cells, at 25°C in Ex-Cell 420 Medium (Sigma-Alorich). Media for mammalian cells were supplemented with 10% fetal bovine serum (FBS) (Wisent), and 100 units/mL penicillin (Life Technologies, Carlsbad, CA) and 100 μg/mL
streptomycin (Life Technologies). Cells were transfected with 5 µg of plasmid DNA in 35 mm dishes using Lipofectamine3000 (Life Technologies). For genome editing experiments, 800 ng of CRISPR-Cas9 plasmid DNA were co-transfected with 800 ng of repair template circular plasmid in 12-well plates. After 2–7 days, cells were non-enzymatically dissociated and seeded on glass coverslips and prepared for imaging and electrophysiology experiments.

**Immunoblotting**

Immunoblot experiments were performed four times. One billion cells were placed into lysis buffer (25mM HEPES, 150mM NaCl, 1mM EDTA, 1% Triton-X) with SIGMAFAST protease inhibitor tablet solution (Sigma-Aldrich). Protein concentrations were measured using a bicinchoninic acid protein assay (Pierce, Rockford, IL) and 30–40µg of protein was loaded into a NuPAGE Novex 12% Bis-Tris Gel (Life Technologies). Proteins were separated by electrophoresis and transferred to a polyvinylidene fluoride membrane using Invitrogen iBlot dry transfer (Life Technologies). The membrane was blocked in 5% BSA in PBS-T and incubated with the following antibody dilutions: 1:1000 anti-RFP rabbit polyclonal (R10367, Life Technologies), and 1:5000 anti-actin JLA-20 mouse monoclonal (Developmental Studies Hybridoma Bank, Iowa City, Iowa) for the top (RFP) and bottom blots (Actin) in Figure 2, respectively. Secondary antibodies used were 1:10,000 HRP-conjugated Donkey anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc.) and HRP-conjugated goat anti-mouse IgG (Abcam, USA). All antibodies were dissolved in 5% BSA in PBS-T. Membranes were imaged using the Pierce ECL Chemiluminescence Detection Kit for HRP (Thermo Scientific, USA). The ratio of band intensity of GFP and fusion products were normalized to actin and quantified using ImageJ.
Image Acquisition

Fluorescence and brightfield microscopy was performed using a Zeiss AxioScope A. All images were acquired at 512 x 512 pixels using a 40× water objective, N.A. 1.0 (epifluorescence) or 60× oil, N.A. 1.4, or 63× water, N.A. 0.9, objectives (confocal) corresponding to an 215 x 160 µm or 120 x 110 µm field of view, respectively. Fluorescence emission was detected using a charge-coupled device (CCD) camera (MRm).

Genome Editing using CRISPR-Cas9

Guide RNAs were designed as 20 bp DNA oligonucleotides and cloned into pX330 (Addgene 42230), and co-transfected with a circular PQR repair template using Lipofectamine 3000 (Life Technologies). All CRISPR-Cas9 guide RNAs were tested for activity using SURVEYOR Nuclease and SURVEYOR Enhancer S (Transgenomics) on extracted genomic DNA. Re-annealed products were analyzed on 4%–20% Novex TBE polyacrylamide gels (Life Technologies). Repair templates were constructed by placing PQR-XFP between homology arms specific to human, mouse, or Drosophila RPL13A. The homology arms lacked the RPL13A promoter, which prevented expression of the PQR-XFP until in-frame genomic integration within an active coding gene. Left and right homology arms were 1.0 kb for the human genome, 1.5 kb for the mouse genome, and 1.4 kb for the Drosophila genome. Cellular fluorescence from PQRs was observed four days post-transfection.
Validation of *PQR* Genomic Insertion

Genotyping experiments were performed in experimental duplicate. Integration of *PQR* into the endogenous *RPL13A* genomic locus was validated by genomic DNA extraction six days post-transfection and genotyping using primers outside and within the homology arms of the repair template. The 5’ and 3’ ends were probed with two sets of primers and the endogenous *RPL13a* locus was PCR amplified. Restriction digests were then performed on PCR products at sites specific for *PQR*. All genomes were sequenced to identify the *PQR* and genomic junctions.

*RPL13A* targeting Guide RNA (gRNA) expressing fly generation

A U6:3gRNA plasmid was first constructed and the genome targeting was verified in *Drosophila* Kc cells as shown in Chapter 4. High quality plasmid was then prepared and used for making gRNA expressing flies at the Bestgene inc. The obtained gRNA flies were balanced and then crossed to *nos-cas9* flies to ensure a germline restricted editing in order to minimize toxicity caused by somatic mutagenesis on essential gene *RPL13A*. Embryos expressing the two components, gRNA and Cas9, were collected for injection with different repair templates to make different knock-in animals.

Different *RPL13A-PQR-XFP* knock-in flies generation

Different repair templates were constructed based on *Drosophila RPL13A* sequences with different *PQR-XFP* as described in chapter 4.
Results

PQRs can be inserted into any genomic locus to quantitate endogenous RPL13A protein levels

As a proof of principle experiment, we generated DNA repair templates with PQR flanked by 5’ and 3’ RPL13A homology arms to integrate PQR fluorescent reporter into RPL13A genomic loci at human, mouse, and Drosophila genomes. Repair templates with red fluorescence protein (PQR-RFP) or blue fluorescent reporter tagged with a nucleus localization signal (PQR-BFP<sub>nls</sub>) were cotransfected with plasmids encoding organism-specific CRISPR-Cas9 to insert PQR into RPL13A loci of HEK293T, Neuroblastoma-2A (N2A), and S2 cells, respectively. Fluorescent cells were identified and isolated for enrichment 5 days after transfection. A pure knock-in cell line was obtained by isolating a single fluorescent cell into each well of a 96-well plate through a custom-made mouth pipette. All knock-in cell lines were first verified by genotyping, using primer pairs with one primer outside of the repair template sequence and the other within the PQR sequence, as well as PQR sequence-specific restriction digestion of the genotyping PCR product, and finally Sanger sequencing on the PCR product. Western blotting results confirmed that no fusion proteins of RPL13A and PQR were produced (Figure 2a,b). The RT-qPCR results further demonstrate that PQR insertion only subtly decreases the transcript number of RPL13A in comparison to wild type control (Figure 2c,d).

Generation and evaluation of RPL13A-PQR-RFP<nsub>nls</nsub> knock-in Drosophila

A series of different RPL13A targeted gRNAs under the control of the U6:3 promoter in pCDF vector were first verified in Drosophila Kc cells (Figure 1c). The one with the highest editing efficiency, based on the proportion of fluorescent cells, was selected for making the
gRNA fly. To minimize lethality due to the induced NHEJ somatic mutation on RPL13A, the balanced gRNA fly was crossed to a fly expressing Cas9 through the nos promoter, which is active solely in male and female germlines. The repair template with Red Fluorescent Protein tagged with a nucleolus localization signal was chosen for better in vivo light penetration and signal enhancement, respectively. Circular RPL13A-PQR-RFPnols repair templates were injected into embryos expressing active Cas9 and gRNA to create germline transmissible knock-ins (Bestgene, Inc). G0 flies from the injected embryos were crossed to one another and the resulting F1 larvae with red nucleoli across the whole body can be identified and isolated under a standard epi-fluorescence or confocal microscope (Figure 3c,d). The correct knock-in was verified by genotyping and Sanger sequencing. The F1 flies were then outcrossed to balance and then homozygose the RPL13A-PQR-RFPnols allele at the third chromosome.

**Red fluorescence intensities vary between different tissues but stay constant among the same type of cells**

The pattern of red fluorescence was first characterized in dissected wandering third instar larvae, in which red nuclei were observed throughout the entire animal with various intensities between different tissues, implying different levels of cellular transcription and translation. Indeed, for cells undergoing rapid proliferation for the purpose of development, such as in the body wall and gut, the fluorescence intensities were significantly stronger than those from slow-growing tissues like neurons (Figure 3d). On the other hand, for the same types of cells, the variation in fluorescence intensity was considerably smaller in comparison with the striking differences between different tissues.
Discussion

The insertion of PQR fluorescent reporter into genomic loci is a useful strategy for quantifying endogenous protein expression. Moreover, the preservation of the authentic genomic context and regulatory elements, such as the endogenous promoter, enhancer, splicing sites, and UTR, can provide a more accurate way to determine gene-to-phenotype relationships. The physiological expression levels of most endogenous proteins in single cells are below the detection limit of commonly used measurement methods including protein assays, ELISA, and western blotting. Although single-cell mass spectrometry and western blotting on endogenous proteins have been developed and performed successfully, the required cell lysis and probing preclude a time-lapse and immediate phenotype-to-protein level correlation analysis. By sequestering fluorescent reporters into small and geometrically uniform compartments, such as nuclei or nucleoli, the originally undetectable signal can be enhanced to a quantifiable level.

The advancement of the CRISPR-Cas9 gene editing technology has made gene editing accessible for research labs. CRISPR mediated knock-out or knock-in has been successfully achieved in a variety of organisms (Sander & Joung, 2014). Theoretically, by constructing repair templates with homology arms and designing locus-specific gRNAs, PQRs can be integrated into any genomic loci to measure the protein expression in dividing eukaryotic cells. To attain temporal or tissue specificity, an inducible PQR reporter with the use of tissue specific Cre or Flp recombinases can switch on the fluorescence specifically in the cell or tissue of interest. Because homologous recombination is only active during the S to G2-phase of cell cycle (Bee, Fabris, Cherubini, Mognato, & Celotti, 2013; Saleh-Gohari & Helleday, 2004), the insertion of PQRs can be accomplished only in proliferating cells (e.g., cell lines or embryos). For non-dividing cells, such as post-mitotic neurons and cardiac myocytes, PQR integration will be
difficult unless homologous recombination activity could be somehow stimulated. To conduct protein quantification in postmitotic cells, a knock-in organism is the only solution. As genome editing techniques become more commonplace, the hurdle of knock-in organism creation is expected to become lower. Increasing numbers of providers offer CRISPR knock-in animal services for two widely used genetic model organisms, *Drosophila* and mouse.

**Creation of RPL13A-PQR-RFP knock-in *Drosophila***

We created an *RPL13A-PQR-RFPnols* knock-in fly with red nuclei in all somatic cells. This fly can benefit different fields of biology, such as providing a measurement for cellular transcriptional and translational states during development, serving as a reference gene to correct for optical effects in *in vivo* imaging, or being employed to monitor the process of metamorphosis during pupal stages as the fading red fluorescence indicates the cessation of cell proliferation and the self-destruction of tissues.

The survival of this fly without obvious defects in growth rates and behaviors addresses numerous concerns regarding the use of PQR in endogenous protein quantification. First, the extended mRNA transcript has been expected to retard translation which might lead to insufficient protein production. This effect cannot be easily evaluated as no other single-cell protein quantification techniques are available to measure the expression of the wild-type RPL13A allele. It could, however, be considered innocuous to the health of the animal as no defective phenotypes have thus far been observed. Second, as discussed in chapter 2, the residual PQR peptide at the C terminal of RPL13A might raise reservations in adopting this technique. The existence of healthy knock-in flies obviates the concern of this issue.
However, although these knock-in flies appear to be healthy without any detectable phenotypes, the influence of the integrated $PQR$-$RFPnols$ on the health of animals cannot be completely excluded as the data collected thus far are from heterozygous animals. A homozygous $RPL13A$-$PQR$-$RFPnols$ fly is required to prove a fully functioning $RPL13A$-$PQR$-$RFPnols$ allele.

**Future perspectives**

To fully exploit the advantages of this knock-in animal, we are aiming to create more flies with different reporters integrated into the $RPL13A$ locus. For example, an $RPL13A$-$PQR$-$BFPnols$ fly with blue nuclei can provide researchers another option when choosing the fluorophore to use. In addition, by crossing the $RPL13A$-$PQR$-$RFPnols$ to $RPL13A$-$PQR$-$BFPnols$ flies, the resulting offspring expressing both red and blue fluorescence from different $RPL13A$ alleles can be used to investigate the differential gene expression regulation between alleles over the course of an experiment. Most diploid eukaryotic organisms, including *Drosophila*, are expected to express both alleles at a given locus. Surprisingly, monoallelic expression, in which only one of the two alleles is expressed, has been demonstrated as a widespread feature of gene expression in human genome. Using a genome-wide approach, Gimelbrant et al. 2007 showed that among the 4,000 assessed human genes from a clonal cell line, over 300 were subject to random monoallelic expression, such that cells express either the maternal allele or the paternal allele. The distinction between the maternal and paternal allele expression in the test is based on the SNP containing regions with heterologous sequences. In other words, this approach cannot detect the differential expression between two alleles with identical sequences. In our case, different colors of protein quantitation reporters are inserted
into each allele of RPL13A with identical sequences (verified by Sanger sequencing). Any biased allele expression on a single-cell scale can be detected instantly through fluorescence microscopy over the course of development.

Additionally, a fly with a conditional reporter is also being created, RPL13A-FRT-stop-FRT-PQR-RFP, in which a copy of the RPL13A 3’UTR with the transcriptional stop signal flanked by FRT is incorporated between the RPL13A coding sequence and PQR. This conditional allele can provide both spatial and temporal resolutions of protein level measurement. With the numerous genetic tools for tissue specific expression of Flp in Drosophila, this conditional fly can be of great benefit to biologist from different disciplines.
Figures

Figure 1, *PQR*s can be inserted into any genomic locus to quantitate endogenous protein levels
a, Insertion of a $PQR$ before the final stop codon of the endogenous gene maintains the mRNA production fidelity and the 3’ untranslated region (UTR) for all isoforms of the mRNA with the $PQR$. A site-specific DNA double-strand break is created using the CRISPR-Cas9 system. This break is repaired by the cell using homologous recombination, and in the presence of an exogenous repair template with appropriate homology arms, the locus is replaced with the $PQR$ edited version. Colored nucleotide sequences represent genomic sequencing results of an edited mouse $RPL13A$ gene with a $PQR$-RFP insertion.

b–d, Targeted genome editing technologies allow for insertion of a $PQR$ into genes in different genomes. Different repair templates and guide RNAs for CRISPR-Cas9 were designed for the $RPL13A$ gene in human (b), Drosophila (c), and mouse (d) genomes. Edited $RPL13A$ genes with $PQR$ produced RFP or BFP with a nuclear localization signal (BFP$_{nls}$). $PQR$ insertion was verified using genomic PCR genotyping with primer pairs (A and B) that spanned $PQR$ and outside the homology arms, followed by genomic sequencing. Scale bars are 100 µm.
Figure 2. Knock-in of PQR into endogenous loci does not produce fusion proteins nor significantly alter the mRNA expression.
(a and b) Genome-edited HEK293 and N2A cells were lysed and protein content was analyzed using immunoblots against RFP. RFP protein bands were observed, but no fusion protein products were detected. (c and d) Quantitative real-time PCR analysis of knock-in PQR cells shows that relative levels of mRNA of the PQR-edited gene are not changed. PQR-specific mRNA was measured and normalized to GAPDH mRNA levels and compared as fold changes from the levels in untransfected cells. For experiments in HEK293T cells, qPCR experiments were performed in duplicate with technical replicates in duplicate, on the RPL13A locus, whereas experiments in N2A cells were performed in quadruplicate with technical replicates in duplicate on the RPL13A locus.
Figure 3, RPL13A-PQR-RFPnols knock-in Drosophila

**a**, A knock-in Drosophila expressing the PQR-RFPnols reporter from the endogenous RPL13A locus. Colored nucleotide sequences represent genomic sequencing results of an edited Drosophila RPL13A gene with a PQR-RFPnols insertion. **b**, A heterozygous RPL13A-PQR-RFPnols fly was verified using genomic PCR genotyping with primer pairs (A and B in a) that spanned PQR-RFPnols and outside of the homology arms. For primers spanning outside of
homology arms (Primer B forward and Primer A reverse), *PQR-RFPnols* allele can produce PCR product size of 3.9 kb (upper arrow), while WT allele resulted in a PCR product size of 3 kb (middle arrow). When primer set A or B was used for genotyping, only the knock-in fly produced PCR amplicon of 1.8 kb (bottom arrow). c, d, red nuclei were observed from all the cells in both embryos and larval body walls. Da neurons labeled in green were used to visually delineate each segment of an entire larval body. Scale bars are 50 µm.
Chapter 4.

Potential application of PQR---Knock-in stable cell line creation

Introduction

Generally speaking, a stable cell line refers to an immortalized homogeneous cell population which consistently produces a recombinant gene indefinitely in culture. It has wide utility in biotechnology including cellular screening for chemical compounds or RNAi, reporter gene assays, cytokine assays, and antibody production. In clinical applications, the use of stable cell lines to produce different pharmaceutical proteins, such as growth hormone, insulin, and interferon, is widespread and has replaced inefficient and expensive protein acquisition procedure from native host animals (Wurm, 2004). Carrying out the recombinant protein production in prokaryotic or lower eukaryotic systems might be more economical and rapid; it is, nonetheless, more desirable to express clinically-used recombinant proteins in mammalian cells as they confer correct folding, assembly, and post-translational modification. Essentially, cultivated mammalian cells, such as Chinese hamster ovary (CHO), baby hamster kidney (BHK), and human embryonic kidney HEK293T cells, have become the dominant system (60-70% proteins used in the clinic) for producing recombinant pharmaceuticals for clinical use (Wurm, 2004).

Typically, to create a stable cell line producing a specific recombinant protein, an expression vector encoding a recombinant gene is transfected into a parental cell line and then permanently integrated to the host cell’s genome. First, the transgene delivery can be achieved by calcium phosphate transfection, lipofection, electroporation, or virus-mediated transduction.
Subsequently, multiple copies of the delivered transgene will randomly integrate into the genomes. To confer selective advantages to the transgenic cells, another gene encoding either a drug resistant protein or an essential metabolic enzyme is also either introduced through co-transfection as another plasmid or placed together with the gene of interest in the same vector as a dual-promoter or bicistronic construct. Following integration, a selection pressure (e.g., antibiotics, or essential amino acid deficient medium) is applied to eliminate all untransfected wild-type cells; as a result, only the cells expressing the selectable marker gene (often wrongly assumed to be co-expressed with the recombinant gene of interest) can survive and grow into colonies in the culture vessel. Multiple rounds of single cell isolation and expansion are required to obtain a stably maintained homogeneous population. Different independent cell clones need to be further verified through genotyping, sequencing, western blotting, and activity assays to ensure a proper expression of the functional protein of interest.

However, although these widely-employed protocols are well-established and simple, they suffer from numerous shortcomings. First, multiple random integrations of transgenes at different loci often lead to an uncontrollable expression, as different integration loci can differentially influence the expression of the recombinant gene, an issue known as positional effect. Also, the copy number of the integrated gene is unaccountable and can vary from cell to cell, which often results in a heterogeneous population of surviving cells after selection. Second, a continuous presence of selection pressure is required to maintain the transgene, as cells tend to remove randomly inserted foreign genetic elements that affect genome integrity. Some selective agents, such as antibiotics, cannot be fully detoxified by the encoded resistant markers, leading to accumulated mutations in the host cells (Oliva-Trastoy, Trastoy, Defais, & Larminat, 2005). These selection agents are also harmful and need to be eliminated from the purified recombinant
proteins before being used in other applications. The purification process can be time-consuming and increase cost. Third, the use of strong exogenous promoters, such as pCAG and pCMV, often drives the expression of the recombinant gene into an elevated and non-physiological level. Overexpression of recombinant proteins can result in a considerable burden on the host cells, especially when multiple copies of the transgene are inserted.

**Generation of knock-in stable cell lines using multicistronic vector and gene editing technology**

Considering the inherent limitations of current methods, we present a distinct protocol to create stable cell lines at known integration loci and copy numbers of the gene of interest (GOI), using CRISPR-Cas9 gene-editing technology. In addition, to allow for quantification of the recombinant protein expression, a *PQR* fluorescent reporter is attached to the transgene so that the expression of the recombinant can be determined by the fluorescence intensity. The genetically attached *PQR* reporter and the **GOI**, PQR-GFP-PQR-GOI, are flanked by locus-specific homologous sequences, in such a way that they can be inserted together into the predetermined genomic locus through homologous recombination facilitated by CRISPR-Cas9. Moreover, rather than being driven by an exogenous promoter and integrated as multiple copies at random loci, a CRISPR-Cas9 mediated gene insertion can place the gene of interest under the control of a given endogenous promoter with all the auxiliary regulatory elements, providing the opportunity to modulate the gene expression on the basis of targeted genomic loci. Finally, the preservation of the native genomic integrity as well as the physiological level of transgene expression can minimize the burden imposed on the host cells, thus ensuring a permanent maintenance of the integrated transgene without the need of a continuous selection pressure.
Materials and Methods

Cell Culture

HEK293T cells were cultured at 37°C under 5% CO₂ in Dulbecco's Modified Eagle Medium (Wisent, St-Bruno, QC) and H-Cell (22c10) (Wisent, St-Bruno, QC), or for Drosophila melanogaster S2 and Kc cells, at 25°C in Ex-Cell 420 Medium (Sigma-Alorich). Media for mammalian cells were supplemented with 10% fetal bovine serum (FBS) (Wisent), and 100 units/mL penicillin (Life Technologies, Carlsbad, CA) and 100 μg/mL streptomycin (Life Technologies). Cells were transfected with 5 μg of plasmid DNA in 35 mm dishes using Lipofectamine3000 (Life Technologies). For genome editing experiments, 800 ng of CRISPR-Cas9 plasmid DNA were co-transfected with 800 ng of circular stable cell line destination template. 7 days later, 200 μg/ml of Zeocin™ was added into culture medium to kill all the untransfected cells. After 1~2 weeks under Zeocin treatment, survived colonies were isolated individually and then expanded into larger populations for further verification. The correct PQR and transgene integrated cells were verified by genotyping, restriction digestion, and Sanger sequencing as described in chapter 4.

Immunoblotting

Immunoblot experiments were performed four times. One billion cells were placed into lysis buffer (25mM HEPES, 150mM NaCl, 1mM EDTA, 1% Triton-X) with SIGMAFAST protease inhibitor tablet solution (Sigma-Aldrich). Protein concentrations were measured using a bicinchoninic acid protein assay (Pierce, Rockford, IL) and 30–40μg of protein was loaded into a NuPAGE Novex 12% Bis-Tris Gel (Life Technologies). Proteins were separated by
electrophoresis and transferred to a polyvinylidene fluoride membrane using Invitrogen iBlot dry transfer (Life Technologies). The membrane was blocked in 5% BSA in PBS-T and incubated with the following antibody dilutions: 1:1000 anti-RFP rabbit polyclonal (R10367, Life Technologies) and 1:5000 anti-actin JLA-20 mouse monoclonal (Developmental Studies Hybridoma Bank, Iowa City, Iowa). Secondary antibodies used were 1:10,000 HRP-conjugated Donkey anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc.) and HRP-conjugated goat anti-mouse IgG (Abcam, USA). All antibodies were dissolved in 5% BSA in PBS-T. Membranes were imaged using the Pierce ECL Chemiluminescence Detection Kit for HRP (Thermo Scientific, USA).

**Genome Editing using CRISPR-Cas9**

Guide RNAs were designed as 20 bp DNA oligonucleotides and cloned into pX330 (Addgene 42230), and co-transfected with a circular PQR repair template using Lipofectamine 3000 (Life Technologies). All CRISPR-Cas9 guide RNAs were tested for activity using SURVEYOR Nuclease and SURVEYOR Enhancer S (Transgenomics) on extracted genomic DNA. Re-annealed products were analyzed on 4%–20% Novex TBE polyacrylamide gels (Life Technologies). Stable cell line destination templates were constructed by placing PQR-RFP<sub>not</sub>-PQR-loxp-Zeocin<sup>R</sup>-PQR-loxp-MCS between homology arms specific to human <i>RPL13A</i>. The homology arms lacked the <i>RPL13A</i> promoter, which prevented expression of the transgene until in-frame genomic integration within an active coding gene. Left and right <i>RPL13A</i> homology arms were 1.0 kb as in chapter 3. Cellular fluorescence from PQRs was observed five days post-transfection.
Results

Identifying a genomic locus for integration in human (HEK293T) cells

The first step is to determine an integration locus, depending on the intended expression levels, consistency of expression, and expression patterns, such as in an inducible manner like heat shock protein genes or a constitutively expressed one like housekeeping genes. Data published by Schwanhäusser and colleagues (2011) provide the levels of gene expression in both mRNA and protein levels from a list of around 5,000 endogenous genes from mouse fibroblast NIH3T3 and 2,000 genes from human breast cancer MCF7 cell lines. These endogenous gene expressions in two relevant organisms provides a reference for the identification of an appropriate gene targeting site (Schwanhäusser et al., 2011). Given its moderate to high expression levels and dispensable role for cell survival, RPL13A was chosen to be the targeting locus to express the recombinant gene of interest and drug resistant marker.

Construction of a PQR multicistronic knock-in vector

We constructed a repair template based on human RPL13A genomic sequences and a multi-cistronic insert, in which RPL13A homologous sequences at both 5’ and 3’ ends flank the multicistronic gene of insert, PQR-RFPnols-PQR-loxP-ZeocinR-PQR-loxP-MCS. PQR red fluorescent reporter with a nucleolus localization signal (PQR-RFPnols) is connected to the “floxed” (loxP flanked) zeocin antibiotic resistant gene, followed by multiple cloning sites (MCS) for accommodating other genes of interest. Red fluorescence from nucleoli is used to quantify the expression levels of all the encoded proteins from the RPL13A locus. ZeocinR confers cell resistance to the antibiotic Zeocin, which induces cytotoxicity by binding and cleaving DNA. The product of ZeocinR can bind to Zeocin in a one-to-one ratio so as to prevent the DNA cleavage. ZeocinR was used as the selection marker owing to its relatively smaller size (375
bases versus 1 kilobase in hygromycin) and higher reliability in stable cell clone selection (Lanza, Kim, & Alper, 2013; our own unpublished data). PQR linkers were cloned in frame between all the encoding genes to split them into discrete proteins. To prevent vector recombination, different PQRs with different synonymous codon compositions were used. The flanking loxP sites allow us to remove the selection gene after the establishment of the stable cell line.

Generation of a stable and pure HEK293T cell line expressing a fluorescent reporter and a drug resistant protein

The first step is to identify the optimal Zeocin concentration for the specific cell line used. To do that, a dose response kill curve, in which cells were subjected to different concentrations of Zeocin treatment, was established for both HEK293T and mouse N2A cells. The drug selection concentration was determined by the minimal concentration of Zeocin needed to kill all the healthy wild-type HEK293T or N2A cells over the course of 1 week. For HEK293T, 200 µg/mL of Zeocin was used.

The repair template (PQR-RFP<sub>nols</sub>-PQR-loxP-Zeocin<sup>R</sup>-PQR-loxP-MCS) and corresponding CRISPR-Cas9 reagents were transfected into HEK293T cells. 5 days after transfection, 200 µg/ml of Zeocin were added into the medium to kill all the untransfected cells. After 2 weeks of drug selection, several colonies of Zeocin-resistant cells started growing, from which around 40% have red nuclei (RFP positive). Single red fluorescent cells were isolated and expanded into different batches under continuous 200 µg/ml of Zeocin treatment. The correct stable knock-in cell clones were verified by genotyping with one site of primer priming at genomic region outside of homology arm and the other within the RFP sequence. These results
suggest that RFP_{nols} and Zeocin were correctly co-expressed and functionally active in the selected stable cells, but the drug resistance and fluorescent nuclei can also be obtained from an unprocessed RPL13A-PQR-RFP_{nols}-PQR-Zeocin fusion protein. To eliminate this possibility and confirm the functionality of PQR, protein lysates from the stable cells were analyzed in western blotting probed by RFP specific antibody. A single band with correct RFP_{nols} size at 30 kDa demonstrates that all PQR peptides function normally in producing separate functional proteins (Figure 2). We conclude that the correct knock-in of the RPL13A-PQR-RFP_{nols}-PQR-Zeocin^{R} at RPL13A locus confers Zeocin resistance to the red nucleolus cells and the vector can be further engineered to make different cell lines expressing different recombinant proteins.

**Discussion**

In chapter 4, we describe the use of the CRISPR-Cas9 gene editing technology to create knock-in PQR cell lines for measuring endogenous protein expression. Here in this chapter, we provide another example of the PQR technology for making stable cell lines continuously expressing recombinant proteins from endogenous genomic loci. Traditionally, stable cell lines are made through integration of transgenes and accompanied drug resistant markers into random genomic loci, both of which are driven by exogenous promoters. Transgenes that are usually driven by constitutively active promoters (e.g., pCAG or pCMV) and influenced by unwanted position effects often renders the expression of the gene of interest uncontrollable. Notably, overexpression of the exogenous genes, as well as the modification of genome integrity due to the multiple random insertions, can potentially impair the homeostasis of host cells and retard their growth. The resulting underprivileged recombinant cells would very likely either remove the transgenes or be outcompeted by the wild-type counterparts once the selection pressure is
withdrawn. That is to say, a continuous presence of selection pressure is required to maintain a transgenic cell clone.

On the other hand, with the locus-specific knock-in and the preservation of the genomic context integrity, the selected homogeneous knock-in cell lines could be maintained without a continuous selection pressure (Figure 1c). This is particularly advantageous given that Zeocin cannot be fully detoxified by the antagonistic agent encoded by ZeocinR (Oliva-Trastoy et al., 2005). Chronic exposure during prolonged selection process has been shown to cause mutagenesis in the established recombinant cell lines (Oliva-Trastoy et al., 2005). The precise targeting of the transgene at the intended locus can obviate the time-consuming characterization of the isolated cell clones, shortening the selection process to 2~3 weeks and allow for early removal of Zeocin from the culture medium.

**Expression of recombinant proteins at a physiological and quantifiable level**

All the obtained knock-in HEK and N2A cells have comparable growth rates as their wild-type counterparts, suggesting that the physiological levels of expression of RFP fluorophore and ZeocinR do not stress the host cells. The preservation of native cellular phenotypes and molecular processes is particularly beneficial as any impairment to the cellular biology would be considered undesirable when it comes to experiments such as drug screening, gene function studies, and other applications in which a normal homeostatic cellular milieu is needed. In contrast, when the transgene is driven by an exogenous promoter and integrated into the genome in multiple copies, the overexpression of recombinant protein might cause unwanted side effects that lead to incorrect interpretation of the experimental results.
Furthermore, the accompanied RFP or other PQR fluorescent reporter provides the opportunity to quantify recombinant protein expression across conditions in which the level of \textit{RPL13A} housekeeping gene expression is either up- or downregulated. Given that RPL13A level indicates the global transcription and translation level of the cells, measuring RFP intensity reveals not only the expression level of the recombinant protein but also a global regulation of cellular homeostasis (Lindström, 2009).

\textbf{Future Perspectives}

Our experimental results demonstrate the utility of this PQR knock-in based method of recombinant cell generation. Several questions remain to be answered and further experiments need to be conducted in order to demonstrate the full spectrum of utility. First, can the recombinant gene of interest be correctly expressed from the gene locus and execute its function normally? Second, if the expression level of recombinant gene is insufficient for some purposes, what other genes can be used to increase expression? Third, how many copies of transgene are incorporated (i.e., homozygous or heterozygous) in the genome? This copy number issue will also affect the observed recombinant protein expression levels.

For the first and second questions, the RFP fluorophore and Zeocin\textsuperscript{R} marker have been expressed and functioning successfully and we are currently testing two representative recombinant genes, \textit{GFP} (717 bases) and \textit{shakerGFP} (2.6 kilobases), in the context of \textit{PQR-RFP}_{\text{nols}}-\textit{PQR-loxP-Zeocin}^\text{R}-\textit{PQR-loxP-GFP/shakerGFP} and verify their functions in HEK293T cells. So far, the GFP fluorescence could not be detected in either cytoplasm (GFP) or cell surface (shakerGFP) although genotyping results indicate a successful knock-in. We reasoned
that the moderate expression levels of RPL13A might be insufficient to drive the expression above detection threshold or the extended transcript might impede the transcription or translation. To address the low expression issue, we are presently constructing and testing a beta actin (ACTB) -based repair template to upregulate the recombinant protein expression, as ACTB has been shown to have one of the highest protein expression levels (Schwanhäusser et al., 2011) and produces strong fluorescence intensity from the gene locus in knock-in stem cells (Kimura et al., 2015).

To the third question, our genotyping results suggest the coexistence of wild type and transgene knock-in RPL13A alleles. Given that the cell lines are originating from single cells, the heterogeneity is not due to a mixed population of heterogeneous cells but a pure cell line with a heterozygous knock-in genome.

In comparison with the conventional random integration-based procedure for stable cell line generation, a site specific targeting of the recombinant transgene with protein quantitation reporter has several invaluable applications. First, a site specific insertion prevents the unpredictable position effect commonly encountered by the random integration approach. Second, a multicistronic vector encoding the transgene and a drug resistance gene simplifies the transfection and selection process. Third, the continuous presence of selection pressure is not required for maintaining the integrated transgene in the cell line, minimizing the chance of drug induced-mutagenesis. Fourth, the physiological expression level of the recombinant protein minimizes the disruption of global cell transcription and translation which provides a better system for experiments when a normal cellular homeostatic condition is required. Fifth, the included PQR can be used to not only quantify the recombinant protein expression but also monitor the entire cellular transcriptional and translational levels based on the RPL13A
fluorescence. Finally, by selecting an expression level of various endogenous genes, the levels of transgene expression can be predetermined by selecting the appropriate gene loci for targeting. Overall, this site specific integration protocol for stable cell line creation can be an important application of PQR and also set the stage for recombinant protein production in a controllable and quantifiable way.
Figures.

Figure 1, Stable cell lines created through CRISPR-Cas9 mediated insertion

a. A stable cell line with \( PQR \)-RFPnols and Zeocin\(^R \) flanked by loxP sites integrated in \( hRPL13A \) locus.

b. \( PQR \) and Zeocin\(^R \) insertion were verified using genomic PCR genotyping with primer pairs (A and B) that spanned \( PQR \) or Zeocin\(^R \) and outside the homology arms (correct PCR amplicons are arrow pointed), followed by genomic sequencing.

c. A population of obtained red nucleolus and Zeocin-resistant cells can be maintained indefinitely in culture without the fading of red fluorescence.

Scale bar is 100 \( \mu \)m.
Figure 2, No fusion proteins were produced from the maintained stable cell lines

The production of separate proteins was verified in western blot probed with RFP antibody (lower arrow). No fusion products were detected from the stable recombinant cell lysates (indicated by upper arrow). Anti-actin antibody was used as a loading control.
Chapter 5.
Conclusion and future directions

In this thesis, I describe the use of the protein quantitation ratioing (PQR) technique to conduct protein quantification in single living cells. The functionality and utility in biology are demonstrated in Chapter 2, 3, and 4. A potential application as a new method for stable cell line creation can find valuable utility in the biopharmaceutical industry, such as for recombinant protein production, drug screening, and other cell-based functional assays. The translation-based protein quantification is at the level of protein synthesis, suggesting that any distinct posttranslational processes, including protein stability and degradation rate, can deviate the true protein levels from the values estimated through the output of fluorescence. This sort of quantification error might be negligible for most proteins since the intracellular protein abundance has been shown to be predominantly controlled by mRNA levels and translation rates, while only 5% variation is contributed by turnover rates (Schwanhäusser et al., 2011). We confirmed this translation rate-dominated protein-level control at the single-cell scale using fluorescence microscopy. Single HEK293T cells expressing shakerGFP-PQR-RFP, which encodes two proteins (shakerGFP and RFP) with strikingly different maturation and degradation rates, produced linearly correlated surface GFP and cytoplasmic RFP fluorescence intensities (Figure 5d in Chapter 2). However, this effect might hold only in proliferating cells in which considerable levels of mRNA and elevated translation rates are maintained. In non-dividing cells, such as circadian clock neurons, the distinct gene regulatory feature of the gene of interest (PER) can render its accumulated protein levels (PER::YFP) contrastingly different from that of the fluorophore (RFP) (Figure 8b in Chapter 2).
Based on the data from our circadian experiments, PQR is typically not appropriate as an absolusion quantification of the protein of interest, but rather a relative protein level measurement between different cells. Despite this intrinsic limitation, the single cell resolution and non-invasive characteristics can still prove it to be a powerful technique in cell biology. It provides a rapid, sensitive, and cost effective way of quantifying protein expression. Given that other alternative assays for protein quantification, such as the classical quantitative western blotting, immunofluorescence, and intracellular flow cytometry, are also relative measurements, as the tissue processing, permeabllization, and the antibody-to-antigen based procedure per se can profoundly affect the accuracy of true cellular protein level measurement, the PQR technique will be a more accurate and useful way of intracellular protein quantification.

PQR has broad application in biomedical and biotechnological research, such as measuring protein production in single cells over time for drug screening or cell signalling assays. The fact that it is not physically tagged to the protein of interest confers it the ability to quantify secreted proteins like antibodies and hormones. The development of gene-editing techniques (CRISPR-Cas9, TALEN, and Zinc-finger nucleases) could potentially facilitate the use of PQR in quantifying proteins expressed from endogenous loci. For example, by integrating PQR into the genomic loci of the genes encoding antibodies or insulin, the B cells or beta cells that are actively producing antibodies or insulins, respectively, can be identified and isolated on the basis of the corresponding fluorescence output. For pharmaceutical applications, the production of a recombinant protein can be ensured if the protein of interest is placed as the upstream gene such that the translation of one molecule of downstream fluorophore must be preceded by the translation of one molecule of the protein of interest.
Furthermore, PQR is a valuable technique to help advance the field of single-cell analysis. Although quantitative PCR allows for single cell transcript measurement, the lack of an effective and non-destructive single-cell protein level measurement restricts our insights regarding the function of the gene of interest, due to the poor correlation between mRNA and protein levels on a population scale. The relationship in single cells has not yet been characterized but is expected to be more disparate given the heterogeneous and dynamic features of single-cell biology. PQR is a non-invasive and single-cell based technique, such that the protein levels from a single cell can be determined first through fluorescence microscopy, then the number of transcripts from that cell can be quantified using single-cell quantitative PCR.

In conclusion, we have demonstrated the utility of PQR and believe that it will be of great benefits to molecular biologists given its broad application and expansion possibility. The field of single-cell analysis can see its immediate usefulness as a tool for protein quantification. The observed codon biased dependent separation improvement is an original finding and merit another topic of study to unravel its underlying mechanism.
References


Gilda, J. E., & Gomes, A. V. (2013). Stain-Free total protein staining is a superior loading control to b-actin for Western blots. *Analytical Biochemistry, 440*(2), 186–188.


Lang, K. J. D., Kappel, A., & Goodall, G. J. (2002). Hypoxia-inducible factor-1alpha mRNA contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia. *Molecular Biology of the Cell, 13*(5), 1792–801.


