# Investigating the Nuclear Roles of Iron Regulatory Protein-1A (IRP1A) in *Drosophila melanogaster*

by

Minyi Yan

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Molecular Biology and Genetics

Department of Biological Sciences
University of Alberta

**Abstract**

Iron is an indispensable micronutrient for almost all organisms. Either the lack or excess of iron can result in various human diseases. Therefore, to maintain cellular iron homeostasis, the IRP/IRE (iron regulatory protein/iron responsive element) system tightly regulates iron concentrations. Under conditions of iron depletion, IRP1 binds to the IRE located in mRNAs, thereby enhancing or inhibiting the translation of iron-related proteins. This binding leads to a net increase of iron availability within cells. When iron levels are sufficient, IRP1 incorporates an iron-sulfur cluster, referred to as holo-IRP1, which has aconitase activity and is unable to bind RNA. Considering the RNA-binding and cytosolic aconitase activity, IRP1 was believed to exclusively localize to the cytoplasm ever since its discovery ~30 years ago. However, recent studies have revealed tissue-specific nuclear localization of fly and human IRP1. In addition, further investigation via RNA-Seq analysis demonstrated that overexpression of IRP1 preferentially downregulated iron- and ecdysone-related genes in *Drosophila*. Given these unexpected discoveries, which challenged the traditional IRP/IRE paradigm, it is of great interest to explore the nuclear roles of IRP1A.

Taking advantage of the CUT & Tag methodology, I generated a chromatin-binding profile for *Drosophila* IRP1A, revealing distinct binding patterns in the prothoracic gland (PG) and fat body chromatin. Additionally, I found that the target genes of IRP1A in these two tissues fell into different functional categories. In the PG, IRP1A exhibited a tendency to bind to loci encoding homeotic genes (e.g., *fushi tarazu*, *engrailed,* and *drumstick*). In the fat body, on the other hand, IRP1A showed associations with genes related to ecdysone response and pupation, including *EcR (ecdysone receptor), Eip75B (Ecdysone-induced protein 75B), broad, Fbp1 (fat body protein 1), Lsp1 (larval serum protein 1),* and *Blimp-1*. This implies that IRP1A potentially plays a role in

ecdysone signaling and the pupation process. Furthermore, I characterized the interactomes of IRP1A and IRP1B for the PG. IRP1A appeared to interact with chromatin remodeler complexes such as Moira (MOR) and Posterior Sex Combs (Psc), suggesting its potential involvement in the maintenance of active and repressed states of homeotic genes via the Trithorax group (TrxG) and Polycomb group (PcG) proteins, respectively. Additionally, this may also imply that IRP1A indirectly influences the transcript profile of iron/ecdysone-related genes via affecting the expression of homeotic genes. Both IRP1A and IRP1B were found to bind to spliceosomal proteins Imp (IGF-II mRNA-binding protein) and CG10333, suggesting that they may affect the transcriptome via their interactions with the spliceosome machinery. To further investigate the functions of IRP1A and IRP1B, I generated CRISPR knock-in and knock-out lines, which, for example, would allow to repeat the CUT & Tag analysis with endogenously controlled genes rather than transgenes. Moreover, I will compare the significance of cytosolic and nuclear of IRP1A, explore the essentiality of conformational interconversion in IRP1A, as well as identify the role of IRP1B in iron metabolism via assessing the survival rate of IRP1A/B mutants.

**Acknowledgements**

I would like to express my deep gratitude to my esteemed supervisor, Kirst King-Jones, who has been an invaluable mentor throughout my journey. Not only has he consistently offered professional advice, but he has also guided me through countless obstacles in various circumstances. His support has allowed me to develop into a more responsible researcher, an experience that I truly treasure.

I would also like to extend my thanks to Martin Srayko, my committee member, for providing constructive suggestions during our meetings. His valuable input has greatly enhanced my research from multiple perspectives.

Furthermore, I would like to express my appreciation to my fellow lab colleagues, particularly Wen Liu, our postdoc. Over the years, I have approached him with numerous questions, and he has always responded patiently and generously. His assistance has been indispensable, and I am sincerely grateful to him. A special thanks goes to Andrew Simmons for his remarkable help whenever I encountered coding challenges. His support has been instrumental in overcoming these obstacles.

I would also like to acknowledge the GTA program of the Department of Biological Sciences for providing me with funding, as well as the Molecular Biology Service Unit (MBSU) team for granting me access to their facilities. Their contributions have been vital to the progress of my research, and I am grateful for their support.

# Table of Contents

**List of Tables**

**List of Figures**

**Abbreviations**

| | |
|---|---|
| **2-ME** | 2-mercaptoethanol |
| **20E** | 20-hydroxyecdysone |
| **5′ss** | 5′ splice site |
| **7DC** | 7-dehydrocholesterol |
| **α-KG** | α-ketoglutarate |
| **αFTZ-F1** | fushi tarazu transcription factor 1, α isoform |
| **µL** | microliter |
| **ABC** | ATP-binding cassette |
| **ALA** | aminolevulinic acid |
| **ALAD** | aminolevulinic acid dehydratase |
| **ALAS** | aminolevulinic acid synthase |
| **attB** | bacteria attachment site |
| **attP** | phage attachment site |
| **acetyl-CoA** | acetyl coenzyme A |
| **BRGC** | brain-ring gland complex |
| **BDSC** | Bloomington Drosophila Stock Center |
| **BPS** | bathophenanthrolinedisulfonic acid disodium |
| **br** | broad |
| **BS** | branch site |
| **Cat** | catalog |
| **ChIP-Seq** | chromatin immunoprecipitation followed by sequencing |
| **CIA** | cytosolic iron-sulfur cluster assembly |
| **CIA1** | cytosolic iron-sulfur cluster assembly 1 |
| **CIAPIN1** | cytokine-induced apoptosis inhibitor 1 |
| **ConA** | concanavalin A |
| **co-IP** | co-immunoprecipitation |
| **CPOX or CPO** | coproporphyrinogen oxidase |

| | |
|---|---|
| **CRISPR** | clustered regularly interspaced short palindromic repeats |
| **CUT & Tag** | Cleavage Under Targets & Tagmentation |
| **cm** | centimeter |
| **Cyp6t3** | cytochrome P450 6t3 |
| **DCytb** | duodenal cytochrome b |
| **DMSO** | dimethyl sulfoxide |
| **DMT1** | divalent metal transporter 1 |
| **dib** | disembodied |
| **drm** | drumstick |
| **EcI** | ecdysone importer |
| **EcR** | ecdysone receptor |
| **en** | engrailed |
| **Eip75B** | ecdysone-induced protein 75B |
| **FAC** | ferric ammonium citrate |
| **FAD** | flavin adenine dinucleotide |
| **FB** | fat body |
| **Fbp1** | fat body protein 1 |
| **FBS** | fetal bovine serum |
| **FBXL5** | F-box and leucine-rich repeat protein 5 |
| **Fer1HCH** | ferritin heavy chain |
| **Fe-S** | iron-sulfur clusters |
| **FMN** | flavin mononucleotide |
| **FPN** | ferroportin |
| **ftz** | fushi tarazu |
| **g** | gram |
| **h** | hour |
| **H2Av** | histone H2A variant |
| **$H_2O_2$** | hydrogen peroxide |

| | |
|---|---|
| **HD** | homeodomain |
| **Hox** | homeotic |
| **H3K27me3** | trimethylation of histone 3 lysine 27 |
| **H3K4me3** | trimethylation of histone 3 lysine 4 |
| **Hsp23** | heat shock protein 23 |
| **HP1** | heterochromatin protein 1 |
| **Imp** | IGF-II mRNA-binding protein |
| **IF** | immunofluorescence |
| **IRE** | iron responsive element |
| **IRP** | iron regulatory protein |
| **IOP1** | iron-only hydrogenase-like protein 1 |
| **kDa** | kilodalton |
| **L** | liter |
| **LIP** | labile iron pool |
| **Lsp1** | larval serum protein 1 |
| **M** | molar |
| **Mco1-4** | multi-copper oxidase 1-4 |
| **min** | minute |
| **ML** | milliliter |
| **mm** | millimeter |
| **MS** | mass spectrometry |
| **MOR** | Moira |
| **Mvl** | malvolio |
| **NADPH** | reduced nicotinamide adenine dinucleotide phosphate |
| **NDOR1** | NADPH-dependent diflavin oxidoreductase 1 |
| **NES** | nuclear export signal |
| **NGS** | next-generation sequencing |
| **NO** | nitric oxide |

| | |
|---|---|
| **nemy** | no extended memory |
| **nvd** | neverland |
| **pA-Tn5** | protein A-Tn5 |
| **PBG** | porphobilinogen |
| **PBGD** | porphobilinogen deaminase |
| **PBS** | phosphate-buffered saline |
| **PcG** | polycomb group |
| **PCR** | polymerase chain reaction |
| **PG** | prothoracic gland |
| **phm** | phantom |
| **POI** | protein of interest |
| **PPOX or PPO** | protoporphyrinogen oxidase |
| **PRC1** | polycomb-repressive complex 1 |
| **pre-mRNAs** | precursor message RNAs |
| **PREs** | polycomb-response elements |
| **Psc** | posterior sex combs |
| **PTTH** | prothoracicotropic hormone |
| **qPCR** | quantitative real-time PCR |
| **RNAi** | RNA interference |
| **RNA Pol II** | RNA polymerase II |
| **RNS** | reactive nitrogen species |
| **ROS** | reactive oxygen species |
| **RT** | room temperature |
| **S2** | Schneider 2 |
| **s** | second |
| **sad** | shadow |
| **SCF** | SKP1-CUL1-F-box |
| **SdhB** | succinate dehydrogenase subunit B |

| | |
|---|---|
| **SDS-PAGE** | sodium dodecyl sulfate–polyacrylamide gel electrophoresis |
| **SEACR** | Sparse Enrichment Analysis for CUT&RUN |
| **shd** | shade |
| **snRNPs** | small nuclear ribonucleoproteins |
| **spok** | spookier |
| **STEAP3** | six-transmembrane epithelial antigen of the prostate 3 |
| **sro** | shroud |
| **TCA** | tricarboxylic acid |
| **TES** | transcription end site |
| **Tf** | transferrin |
| **TFs** | transcription factors |
| **TfR1** | transferrin receptor 1 |
| **TrxG** | trithorax group |
| **Tsf1** | transferrin-1 |
| **TSS** | transcription start site |
| **usp** | ultraspiracle |
| **UROD** | uroporphyrinogen decarboxylase |
| **UTR** | untranslated region |
| **V** | volt |
| **WB** | western blot |
| **X-S** | sulfur-containing factor |

# Chapter 1 General introduction

## 1.1 The importance of iron

Iron is a crucial micronutrient needed for nearly all organisms [1]. In humans, around 200 billion red blood cells are produced every day, utilizing ~6 grams of hemoglobin and ~20 mg of iron [2]. Additionally, iron is the fundamental substrate for two important cofactors, namely heme and iron-sulfur clusters (Fe-S) [3, 4]. In addition to the involvement in hemoglobin, heme is also required to produce steroid hormones in humans, because heme-binding cytochrome P450 proteins are needed for this biosynthetic pathway [5]. Similarly, the biosynthesis of steroid hormone ecdysone in *Drosophila melanogaster* requires a lot of heme as cofactors [3]. On the other hand, Fe-S clusters participate in various cellular processes such as iron regulation, where they facilitate conformational changes of IRP1s (iron regulatory protein 1) into aconitase [6]. Given the critical importance of iron, it is not surprising that its misregulation is associated with a variety of human diseases. For instance, an excess of iron in the brain could potentially serve as a pathological factor in the development of neurodegenerative disorders such as the Alzheimer's disease [7, 8], Parkinson's disease [9, 10], and Huntington's disease [11]. In addition disturbances in the heme biosynthesis pathway can lead to porphyria [12].

## 1.2 The prothoracic gland (PG) is a great model to study iron metabolism

### 1.2.1 The heme biosynthesis pathway

Utilizing glycine and succinyl-CoA as substrates, the synthesis of heme starts with the production of aminolevulinic acid (ALA) by ALAS (aminolevulinic acid synthase) (Figure 1.1). The condensation of two ALA molecules is then catalyzed by ALAD (aminolevulinic acid dehydratase) to produce the porphobilinogen (PBG). The PBGD (porphobilinogen deaminase) catalyzes the oligomerization PBG molecules, generating the hydroxymethylbilane, which undergoes ring closure to form the first cyclic intermediate uroporphyrinogen III. Subsequent modifications of the macrocycle side chains occur with the aid of UROD (uroporphyrinogen decarboxylase) and CPOX/CPO (coproporphyrinogen oxidase), leading to the formation of coproporphyrinogen III and protoporphyrinogen IX, respectively. The ring system is then aromatized via PPOX/PPO (protoporphyrinogen oxidase), giving rise to the production of protoporphyrin IX. Finally, iron is inserted into the protoporphyrin IX molecule, yielding heme [13, 14].

Currently, nine types of porphyria disorders have been documented in humans, each resulting from the accumulation or overproduction of heme precursors (Figure 1.1) [12, 15]. Among these types, congenital erythropoietic stands out with its distinct characteristics of pinkish fluorescent teeth and urine under the UV light [16, 17]. This phenotype is caused by the dysfunction of the URO3S enzyme in step four, which results in the accumulation of non-enzymatic pathogenic porphyrins (namely uroporphyrin I and coproporphyrin I) instead of the normal conversion of hydroxymethylbilane to uroporphyrinogen III (Figure 1.1) [12, 15, 16]. Those porphyrinogen rings can be isomerized to fluorescent porphyrins when exposing to the air and UV light [18].

Intriguingly, in *Drosophila*, the accumulation of the last four intermediates (compounds 5 to 8, indicated as red circles in Figure 1.1) results similarly in the accumulation of proto-porphyrins, which we termed the "porphyria-like phenotype" [19]. For example, the PG-specific RNAi (RNA interference) of PPOX in step seven causes the accumulation of protoporphyrinogen IX, leading to the porphyria-like phenotype [19]. This phenotype is characterized by tissue enlargement and red auto-fluorescence of the PG in flies. It is worth noting that the porphyria-like phenotype can arise not only from the incapability of ferrochelatase to insert iron into protoporphyrin IX, which is crucial in the last step of heme synthesis, but also from iron deficiency in the PG due to the genetic disruption in iron delivery. For instance, when IRP1A, the main iron regulator in flies, is knocked out/down in the PG, PG cells lose the ability to uptake sufficient iron under iron-depleted conditions, which thereby disrupts the synthesis of heme and leads to autofluorescence in the PG [19].

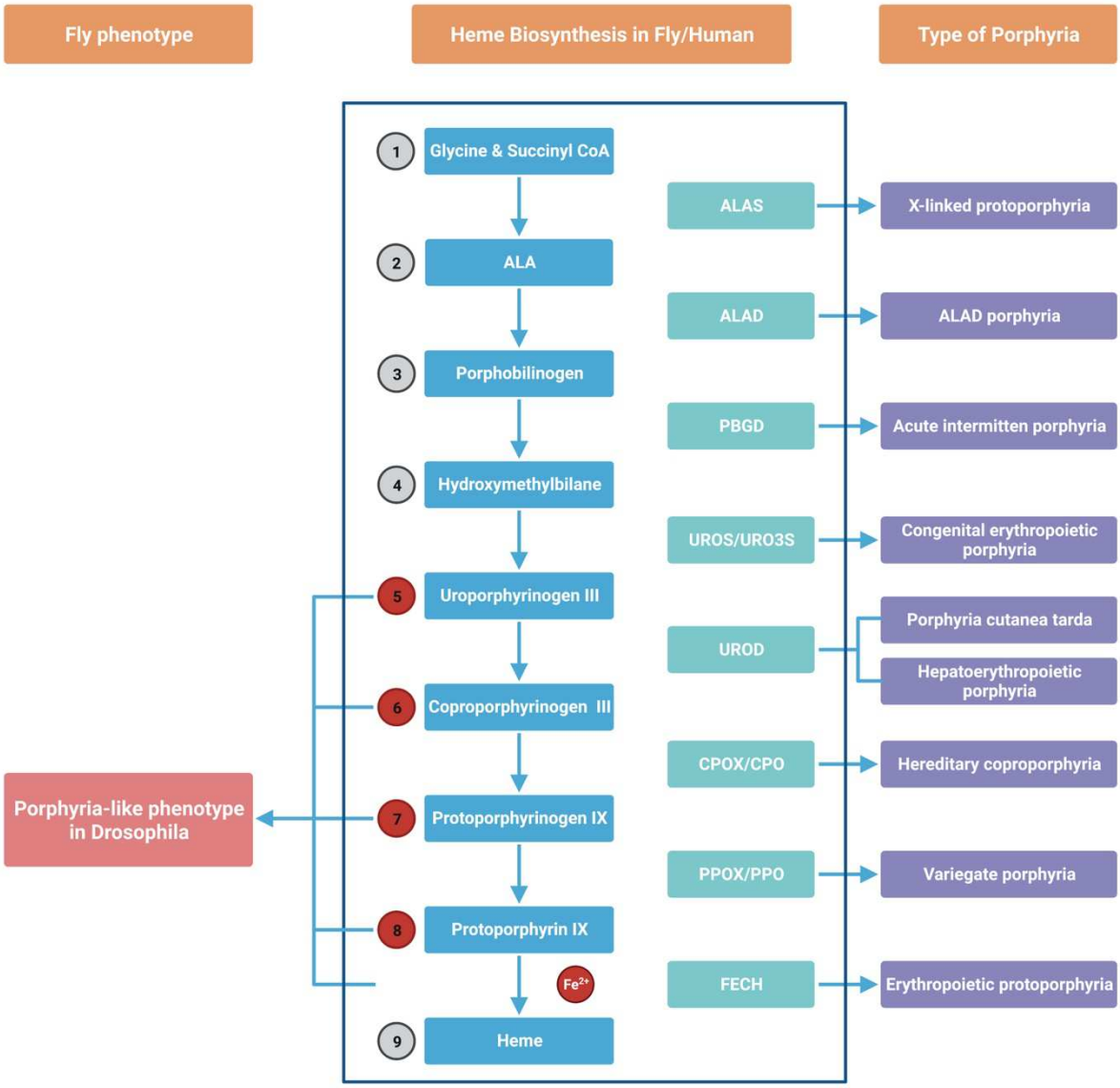



**Figure 1.1 The heme biosynthesis pathway and types of porphyria**

The heme biosynthetic pathway in *Drosophila* and humans is shown. Homologous enzymes with different names in flies and human are separated by a forward slash. In total, there are eight enzymes involved in this pathway, and disruptions in each of these enzymes lead to distinct types of porphyria in humans. Interestingly, in *Drosophila*, the accumulation of intermediate compounds 5 to 8 (indicated in red circles) due to mutations in corresponding downstream enzymes results in a porphyria-like phenotype, which is characterised by the presence of red auto-fluorescence and/or enlarged PGs. Notably, the failure of iron incorporation can also contribute to the porphyria-like

phenotype because Fe is essential in the final step of heme biosynthesis. ALA: aminolevulinic acid; ALAS: aminolevulinic acid synthase; ALAD: aminolevulinic acid dehydratase; PBGD: porphobilinogen deaminase; UROS/URO3S: uroporphyrinogen III synthase; UROD: uroporphyrinogen decarboxylase; CPOX/CPO: coproporphyrinogen oxidase; PPOX/PPO: protoporphyrinogen oxidase; FECH: ferrochelatase [13, 15].

### 1.2.2 The ecdysone biosynthesis pathway

The production of ecdysone, one of the most crucial hormones for *Drosophila* development and metamorphosis, predominantly occurs in the larval prothoracic gland [20]. The PG serves as the decision-making center, where external environmental cues and internal signals converge to regulate the timing of ecdysone production and release [20]. An important regulatory pathway, for example, involves the prothoracicotropic hormone (PTTH), a neuropeptide produced by the brain. PTTH binds to its receptor Torso and activates the transcription of Halloween genes through the Ras/Raf/ERK signaling pathway (Figure 1.2) [21].

The Halloween genes encode a subset of P450 enzymes that are required for ecdysone biosynthesis. This process starts with dietary cholesterol that is taken up by the PG cell, which is then converted to 7DC (7-dehydrocholesterol) by Neverland (Nvd) [3, 22]. Subsequently, the conversion of 7DC to 5β-ketodiol is catalyzed by the "black box" enzymes, named so because the intermediates remain unclear and because additional unidentified enzymes may act in this part of the pathway. The characterized black box proteins, likely include Shroud (Sro), Spookier (Spok), and Cyp6t3 (Cytochrome P450 6t3) [23]. Subsequently, Phantom (Phm) catalyzes 5β-ketodiol to 5β-ketotriol, followed by its transformation into to 2-deoxyecdysone via Disembodied (Dib). The synthesis of ecdysone is completed by Shadow (Sad), after which the ecdysone is released to the hemolymph or imported into peripheral tissues by ecdysone importer (EcI). In these tissues, ecdysone is transformed into its active form, 20-hydroxyecdysone (20E), through Shade (Shd) [22, 24, 25]. The 20E (hereinafter referred to as ecdysone) interacts with the heterodimer EcR/Usp (ecdysone receptor/ultraspiracle) to mediate the transcription of a wide range of genes, resulting in the physiological, morphological, and behavioral changes [20].

Ecdysone is generated and released in pulses, and each major pulse corresponds to a transition in developmental stages [26]. For instance, there is a significant increase in ecdysone levels prior to key developmental transitions such as embryo-to-larval transition, puparium formation, and eclosion. The synthesis of ecdysone requires an abundant supply of iron-sulfur clusters and heme co-factors: the enzyme Neverland requires Fe-S as a prosthetic group; heme acts as a cofactor for all the P450 enzymes involved in the ecdysone synthetic pathway. The synthesis of heme itself is dependent on iron availability as well (Figure 1.1) [19]. Consequently, the high

demand for iron in the PG makes it sensitive to changes in iron concentrations, which is one of the main reasons why it represents a great model for studying iron metabolism. Therefore, I will utilize the PG to study IRP1A in my research.

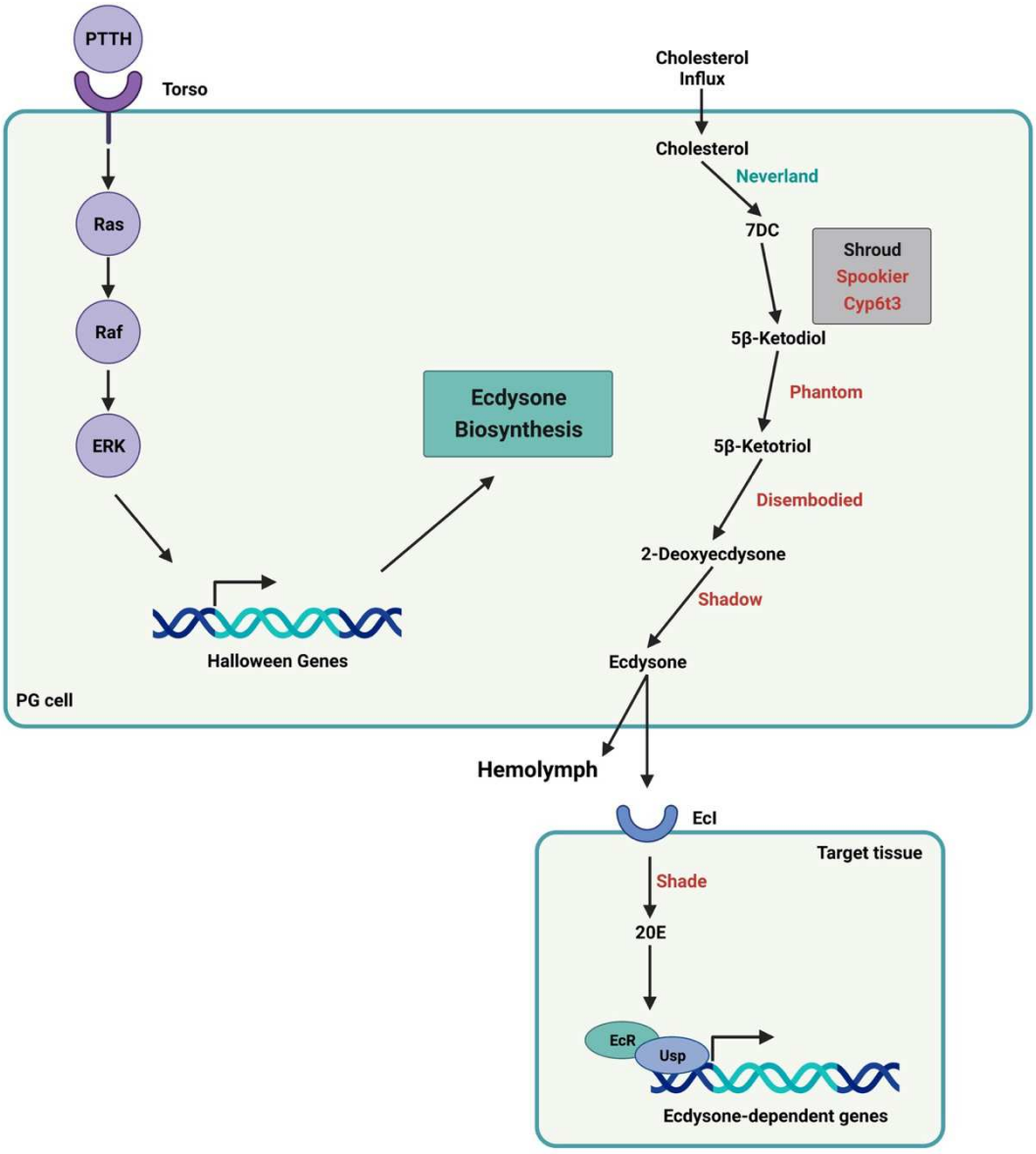



**Figure 1.2 The synthesis and action of ecdysone**

The "black box" is represented as a grey rectangle. The P450 enzymes are shown in red; while neverland, which requires iron-sulfur clusters as cofactor, is in green. PTTH: prothoracicotropic hormone; 7DC: 7-dehydrocholesterol; EcI: ecdysone importer; 20E: 20-hydroxyecdysone; EcR: ecdysone receptor; Usp: ultraspiracle.

## 1.3 Iron homeostasis in mammals

### 1.3.1 Iron uptake and traffic by mammals

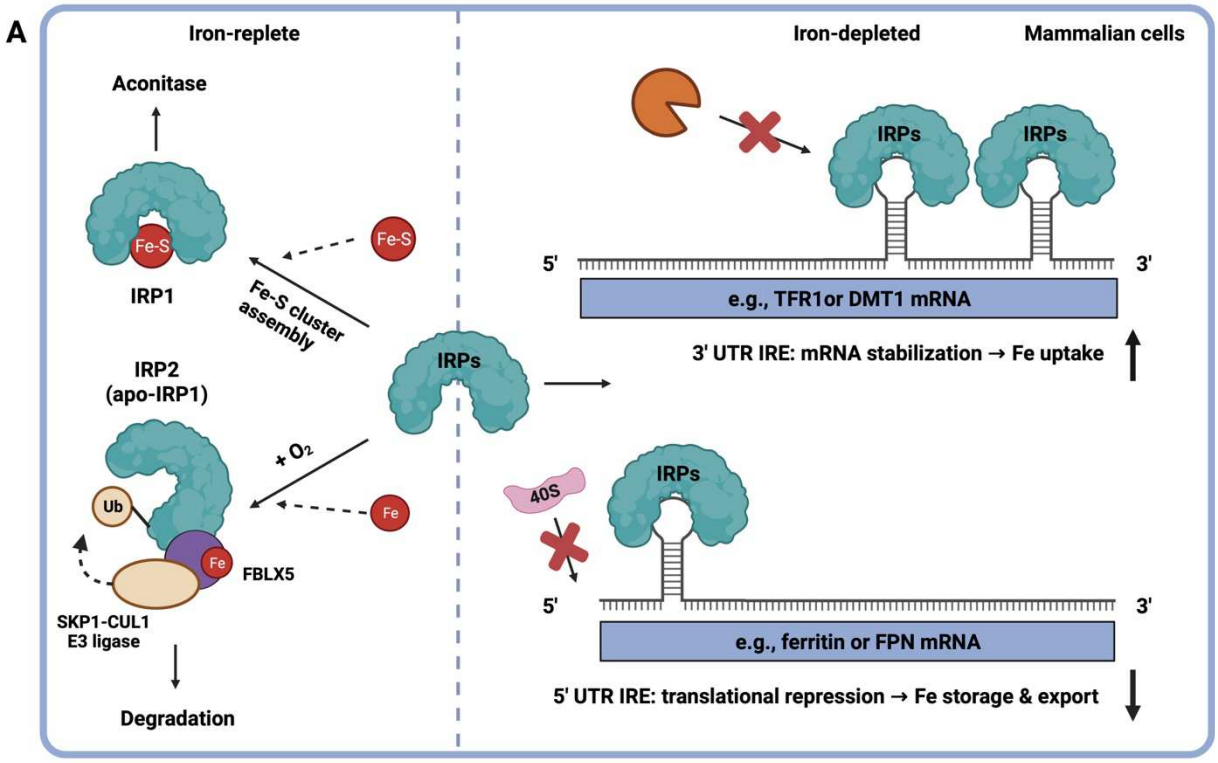
Iron as an essential micronutrient that can only be acquired through the diet in the forms of heme or nonheme iron [1]. Since the process of heme uptake remains poorly understood, it will not be discussed in this thesis. In mammals, the very first step of iron metabolism occurs in the duodenum, where dietary iron (mainly ferric iron) is reduced by the ferric reductase duodenal cytochrome b (DCytb) before its import via the iron importer DMT1 (divalent metal transporter 1) [1]. Once inside the enterocyte, imported iron accumulates in the labile iron pool (LIP), which serves as a reservoir for iron distribution towards utilization, storage, and export. In cells, an abundance of iron is utilized by mitochondria for the synthesis of heme and iron-sulfur clusters. Excess iron can be stored in ferritin, a protein complex with a cage-like structure that can accommodate approximately 4,000 ferric ions. Additionally, absorbed iron is exported and transported to other cells to fulfil their iron demands. The export process from the enterocyte involves the efflux of ferrous iron through ferroportin (FPN) and the oxidation of $Fe^{2+}$ by multicopper ferroxidase hephaestin or ceruloplasmin [1, 2, 27]. Exported $Fe^{3+}$ is bound by transferrin (Tf) and delivered to other cells, which acquire iron through transferrin receptor 1 (TfR1)-mediated endocytosis. In the acidic endosome, iron is released and reduced by STEAP3 (six-transmembrane epithelial antigen of the prostate 3). Subsequently, ferrous iron is transported across the endosome membrane by DMT1, entering the LIP for allocation [2].

### 1.3.2 Iron regulation in mammals

In mammals, the regulation of body iron bioavailability primarily relies on a hormone called hepcidin, which has no counterpart in the fly [27]. Hepcidin interacts with the FPN iron exporter when iron is repleted, triggering the internalization, ubiquitination, and degradation of FPN. As a result, iron efflux from enterocytes into the bloodstream is reduced. At the cellular level, on the other hand, iron concentrations are strictly controlled by the IRP/IRE system (Figure 1.3). There are two IRPs in mammals known as IRP1 and IRP2. When cellular iron levels are low, both IRPs, particularly IRP2, are able to bind to the IREs of mRNAs. The IRE is a hairpin structure that locates in either the 3′- or 5′-UTR (untranslated region) of mRNAs [28]. The interaction between IRPs and IREs either enhances or inhibits the translation of iron-related proteins, thereby

promoting increased iron uptake and decreased iron export. For instance, when IRPs interact with the IRE located in the 3′-UTR, it protects the mRNA from degradation and enhances the translation of proteins such as DMT1 and TfR1, thereby facilitating iron uptake. Furthermore, IRPs inhibit the translation of ferritin and ferroportin mRNAs by binding to their 5′-UTR IREs, which blocks the interaction between ribosomes and mRNAs. Consequently, iron storage and export are reduced. In summary, there is a net increase in iron availability when IRPs bind to IREs.

Conversely, in the presence of sufficient iron, IRP1 can bind to iron-sulfur clusters, triggering conformation changes in the IRP1, converting it into the holo-form with aconitase activity (Figure 1.3). Moreover, IRP2 (as well as apo-IRP1) is bound by FBXL5 (F-box and leucine-rich repeat protein 5), which recruits an SCF (SKP1-CUL1-F-box) E3 ligase, facilitating the ubiquitination and degradation of IRPs by the proteasome [27, 29, 30]. FBXL5 acts as an iron and oxygen sensor. It is stabilized by the interaction with iron and oxygen; otherwise, it undergoes degradation, resulting in increased levels of apo-IRP1 and IRP2 [31].


A
Iron-replete
Iron-depleted
Mammalian cells
Aconitase
Fe-S
IRP1
Fe-S cluster assembly
IRPs
IRP2
(apo-IRP1)
+ O2
Ub
Fe
FBLX5
SKP1-CUL1
E3 ligase
Degradation
IRPs
IRPs
5'
3'
e.g., TFR1or DMT1 mRNA
3' UTR IRE: mRNA stabilization → Fe uptake
40S
IRPs
5'
3'
e.g., ferritin or FPN mRNA
5' UTR IRE: translational repression → Fe storage & export

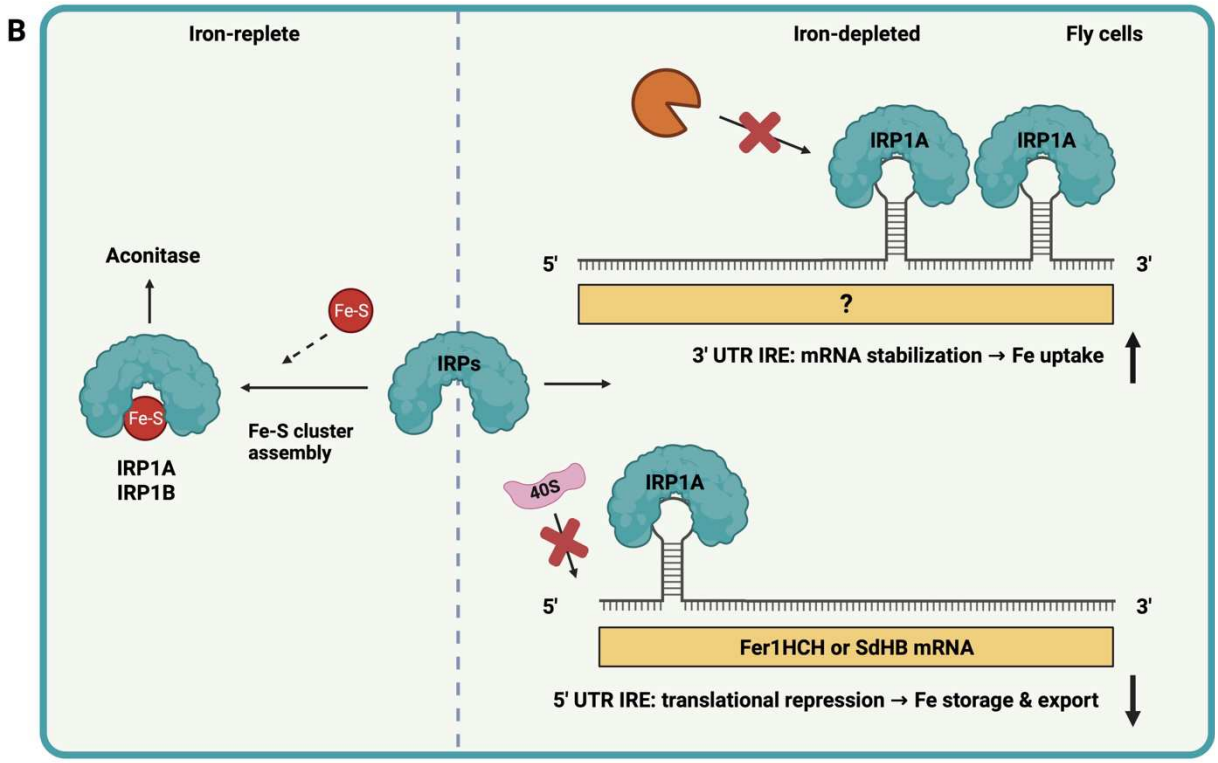

B
Iron-replete
Iron-depleted
Fly cells
Aconitase
Fe-S
IRP1A
IRP1B
Fe-S cluster assembly
IRPs
IRP1A
IRP1A
5'
3'
?
3' UTR IRE: mRNA stabilization → Fe uptake
40S
IRP1A
5'
3'
Fer1HCH or SdHB mRNA
5' UTR IRE: translational repression → Fe storage & export

**Figure 1.3 Comparison of the mammalian and the fly IRP/IRE system**

The IRP/IRE system in (A) mammals and (B) fruit flies is illustrated, respectively. (A) There are two IRPs in mammals, namely IRP1 and IRP2. When cellular iron levels are low, both IRP1 and IRP2 bind to IREs located in either the 3′-UTR or 5′-UTR of mRNAs, promoting or blocking the translation of iron-related proteins. Consequently, the levels of iron uptake proteins such as TFR1 and DMT1 increase, while the translation of iron storage (e.g., ferritin) and export (e.g., ferroportin) proteins is inhibited, resulting in a net increase in cellular iron concentration. Under iron-replete conditions, IRP1 undergoes conformational changes upon interaction with iron-sulfur clusters, transforming into a cytosolic aconitase. Moreover, FBXL5, when stabilized by the presence of iron and/or oxygen, recruits the SKP1-CUL1 E3 ligase, which leads to the ubiquitination and subsequent degradation of IRP2 (and apo-IRP1). (B) In *Drosophila*, IRPs are termed IRP1A and IRP1B, both of which are orthologs of human IRP1. Currently, there is no identified homolog of IRP2 in flies. Like human IRP1, IRP1A regulates the translation of iron-related proteins via the interaction with IREs and can become an aconitase by binding with Fe-S when iron is sufficient. However, IRP1B only possesses the iron-binding ability and functions solely as an aconitase. So far, only two IREs have been identified in fruit flies, specifically in the mRNAs encoding the heavy chain of ferritin (Fer1HCH) and SdhB. Both IREs are located in the 5′-UTR, while no 3′-UTR IREs have been found. Fe-S: iron-sulfur cluster; FBXL5 F-box and leucine-rich repeat protein 5; TFR1: transferrin receptor 1; DMT1: divalent metal transporter 1; FPN: ferroportin; Fer1HCH: ferritin heavy chain; SdhB: succinate dehydrogenase subunit B.

### 1.3.3 The structure of human IRPs

Both human IRP1 and IRP2 are classified as iron-sulfur cluster isomerases, comprising 889 and 964 amino acids, respectively. They are highly similar to each other, sharing 61% sequence identity and 75% similarity [32]. A major difference between these two proteins is that IRP2 harbors a cysteine-rich 73-amino-acid insertion in domain 1 [32-34]. Given their similarities, both IRP1 and IRP2 consist of four domains, with the first three domains compact together and link to the forth domain via a linker sequence (Figure 1.4 A) [35]. During conformational changes, domains 1 and 2 form a relatively stable core, while domains 3 and 4 move extensively [36].

Under conditions of low iron concentrations, IRP1 (and IRP2) adopts an L-shape conformation (the "open" state), which provides sufficient space to accommodate IREs (Figure 1.4 A and B) [36, 37]. The IRE is a stem-loop structure composed of approximately 30 nucleotides. The consensus structure of IREs features a six-nucleotide loop (CAGUGN), and a top helix and bottom helix separated by a bulged C in the middle (Figure 1.4 B) [28, 38, 39]. The top helix is 5-base pairs long, while the length of bottom helix varies among members of the IRE family. Through site-directed mutagenesis studies, arginine residues at position 536, 541, and 780 have been identified as crucial for the IRE-binding activity of IRP1 [40]. When R536 is mutated to glutamine, the RNA-binding affinity is significantly reduced, while mutations R541Q and R780Q completely abolish the RNA-binding capability. Furthermore, the crystal structure of the IRP/IRE complex has revealed that both the terminal loop and the bottom helix contribute to the interaction with IRP1 [36]. Conversely, when iron levels go up, domains 3 and 4 undergo a pincer-like movement, forming a narrow cleft that interacts with a cubane [4Fe-4S] cluster (Figure 1.4 A and C) [35, 36, 41]. Cysteine 437, 503, and 506 have been identified as critical residues for Fe-S binding [40]. Mutating any of these cysteine residues to serine renders the aconitase activity of IRP1 undetectable and leads to a constitutive RNA-binding state, making IRP1 unresponsive to the changes of iron bioavailability.

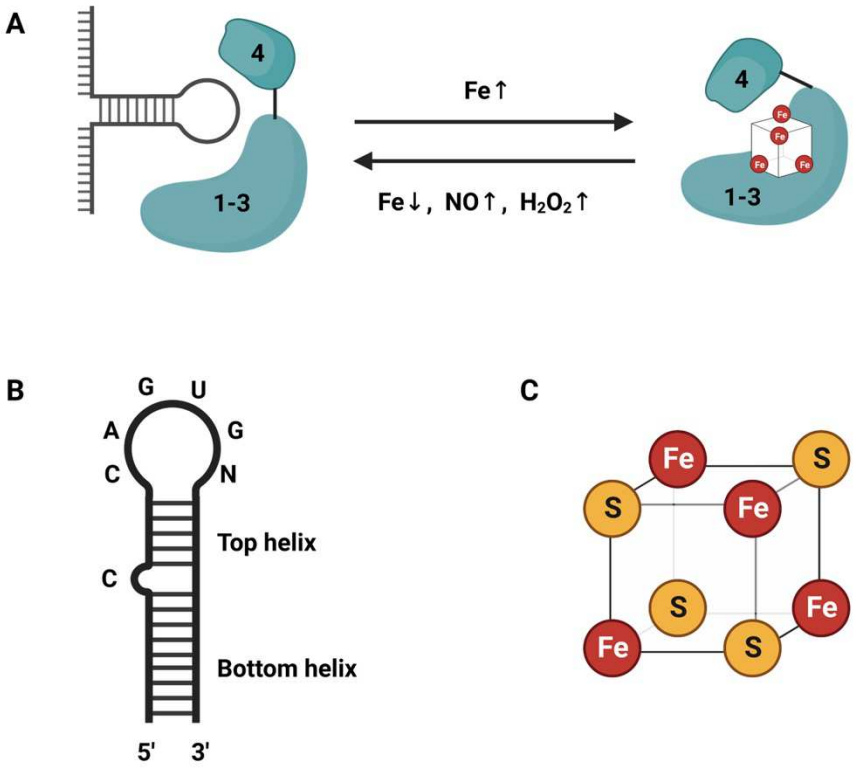



**Figure 1.4 The conformational transition of human IRP1**

(A) The structural interconversion of IRP1. Domains 1-3 of human IRPs form a compact structure and are connected to the fourth domain through a linker. When iron is depleted, IRP1 (and IRP2) adopts an L-shape structure, which opens up the space between domains 1-3 and domain 4, allowing for the accommodation of IRE. Conversely, when iron levels are sufficient, domains 3 and 4 move towards each other (pincer-like movement), forming a narrow cleft that interacts with a [4Fe-4S] cluster. Additionally, apart from iron deficiency, both nitric oxide (NO) and hydrogen peroxide ($H_2O_2$) can trigger the disassembly of the Fe-S, resulting in the enhanced IRE-binding activity of IRP1. (B) The structure of consensus IRE. The conserved structure of IREs is a ~30-nucleotide long hairpin that characterizes a loop containing six nucleotides CAGUGN, followed by the top helix and the bottom helix, which are separated by an unpaired C. (C) The structure of [4Fe-4S] cluster.

### 1.3.4 Cytosolic iron-sulfur cluster assembly (CIA)

Iron-sulfur clusters, mostly existing as [2Fe-2S], [3Fe-4S], and [4Fe-4S] forms, play an important role in numerous biological processes, including the tricarboxylic acid (TCA) cycle, DNA synthesis and repair, and iron homeostasis [42]. Unlike bacteria, which harbor three independent pathways to synthesize iron-sulfur clusters [41], the biogenesis of Fe-S is highly conserved across eukaryotic organisms (e.g., from yeast to human) [43]. The de novo synthesis of Fe-S clusters, which is essential for all eukaryotes, is carried out by the iron sulfur cluster machinery in mitochondria (or in mitochondria-related organelles called mitosomes) [42, 43]. An exception is Oxymonad *Monocercomonoides*, which has lost mitochondria during evolution [42]. This organism employs a simplified CIA machinery along with a bacteria-derived sulfur mobilization system to generate iron-sulfur clusters. In the mitochondria, the assembly of iron-sulfur clusters can be divided into three steps, where the [2Fe-2S] is initially synthesized by the scaffold protein ISCU2. Subsequently, this [2Fe-2S] is transferred to mitochondrial [2Fe-2S] proteins, and a sulfur-containing factor (X-S) is synthesized, which is critical for the synthesis of cytosolic and nuclear Fe-S. The proteins involved in the first two steps, which are necessary for assembling extra-mitochondrial iron-sulfur clusters, are collectively known as the core Fe-S assembly components. In the last step, [4Fe-4S] clusters are generated and delivered to mitochondrial recipient apoproteins, such as the mitochondrial aconitase, which is responsible for the TCA cycle [41, 42].

The newly synthesized X-S is exported to the cytoplasm by an ATP-binding cassette (ABC) transporter ABCB7 and provides sulfur source for the assembly of cytosolic and nuclear iron-sulfur clusters [44, 45]. The CIA process can be divided into two major steps, where the Fe-S is initially generated by the scaffold protein complex and subsequently inserted into target proteins by the CIA targeting complex (Figure 1.5) [46]. The first step of CIA involves two P-loop NTPase CFD1 and NBP35, which form a heterotetramer and serve as the scaffold for Fe-S assembly [47-49]. A [4Fe-4S] cluster bind to the C terminus of both proteins transiently, bridging two subunits of the complex and allowing efficient transfer of Fe-S to target proteins [50]. The second [4Fe-4S], on the other hand, binds to the N terminus of NBP35 tightly and is essential for the proper function of NBP35. Aside from X-S, the biosynthesis of cytosolic Fe-S also relies on electrons provided by an electron transfer chain, which comprises NADPH (reduced nicotinamide adenine dinucleotide

phosphate), the diflavin protein NDOR1 (NADPH-dependent diflavin oxidoreductase 1), and the Fe-S protein CIAPIN1 (cytokine induced apoptosis inhibitor 1) [51]. Upon the binding of NADPH to NDOR1, electrons are transferred to the FAD- (flavin adenine dinucleotide) binding domain of NDOR1 and then to the FMN- (flavin mononucleotide) binding domain. Subsequently, one electron is transferred to the [2Fe-2S] cluster within the CIAPIN1 protein and supplies the CIA machinery [46, 51].

During the second step, the newly assembled [4Fe-4S] clusters, which are loosely bound by the CFD1-NBP35 scaffold, are released and inserted into target proteins (Figure 1.5) [46]. A key protein called IOP1 (iron-only hydrogenase-like protein 1) acts as a connector between the early- and late-acting components of the CIA machinery. It potentially functions as an electron donor for the maturation and/or transfer of cytosolic Fe-S clusters [52, 53]. The pivotal event in the second step, Fe-S insertion, is carried out by the CIA targeting complex. This complex consists of CIA1 (cytosolic iron-sulfur cluster assembly 1), CIA2B and MMS19. Although the exact mechanism of how the CIA targeting complex build up is not yet fully understood, it is believed that CIA1 serves as the docking site for the assembly, while CIA2B and MMS19 interact specifically with different subsets of target apoproteins and facilitate the Fe-S transfer [54-56]. These three proteins form various binary and ternary complexes, enabling the maturation of a wide range of Fe-S proteins involved in nucleotide metabolism, translation, DNA repair, DNA replication, and so on [55, 56]. Notably, another isoform of CIA2 called CIA2A is specifically implicated in iron metabolism, and is involved in the Fe-S maturation of human IRP1 and the stabilization of IRP2 (Figure 1.5) [56, 57].

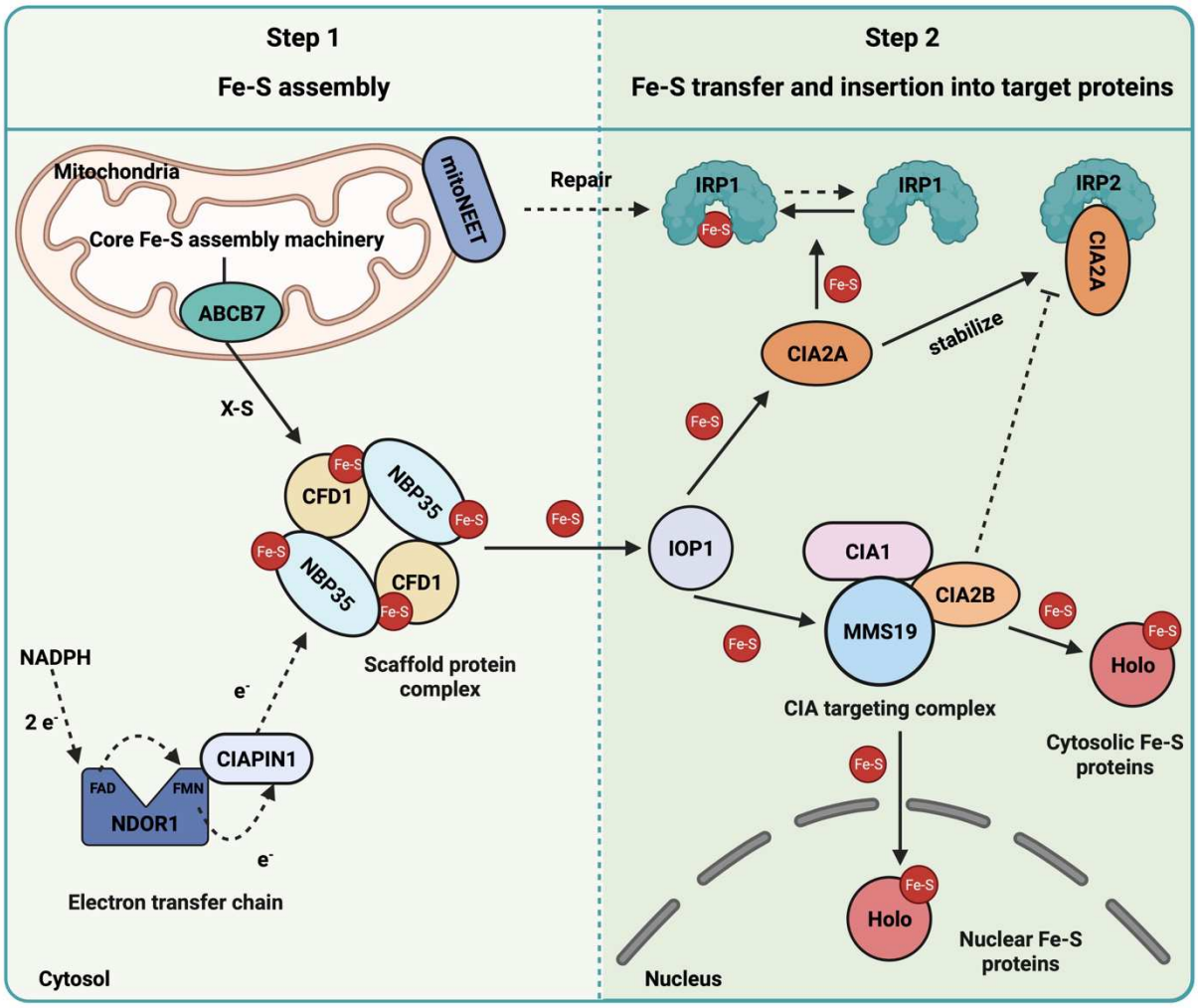



**Figure 1.5 Cytosolic iron-sulfur cluster assembly in humans**

In the first step of CIA, the sulfur-containing compound X-S that is synthesized by the core mitochondrial Fe-S assembly machinery is transported to the cytosol, providing a sulfur resource for the CIA process. Electrons required for the CIA are supplied through the electron transfer chain, which comprises NADPH, NDOR1, and CIAPIN1. The assembly of [4Fe-4S] clusters takes place within the scaffold protein complex, which is a heterotetramer composed of CFD1 and NBP35. These iron-sulfur clusters are loosely bound to the scaffold, ready to be efficiently transferred to target proteins. Another [4Fe-4S] cluster that interacts with the N terminus of NBP35 tightly is required for the function of NBP35. The second step of CIA involves IOP1, which connects the early- and late-acting CIA components. The insertion of Fe-S into apoproteins is facilitated by the

CIA targeting complex (comprising CIA1, CIA2B and MMS19), giving rise to the maturation of various cytosolic and nuclear Fe-S proteins. Additionally, CIA2A, another CIA2 in huma, plays a specific role in the maturation of cytosolic aconitase (holo-IRP1). It can also stabilize IRP2 through its binding, while CIA2B and MMS19 may affect the stability of IRP2. Furthermore, under oxidative stress, mitoNEET is able to repair the Fe-S in holo-IRP1, which is damaged by reactive species, such as NO and $H_2O_2$. Fe-S: [4Fe-4S] iron-sulfur clusters; $e^-$: electrons.

### 1.3.5 The CIA component plays a role in iron metabolism

The efficient knockdown of CIA1, CIA2A, and CIA2B via RNAi in HeLa cells imposed different effects on human IRP1, suggesting that CIA components played a role in iron homeostasis [46, 56]. Specifically, depleting CIA2A led to a reduction in both cytosolic aconitase activity and the protein level of IRP1, causing IRP1 to favor the IRE-binding state. Conversely, the RNAi of CIA2B slightly enhanced the aconitase activity and protein levels of IRP1, while CIA1 depletion displayed no noticeable effects. Under all these conditions, the RNA-binding activity of IRP1 barely changes. Taken together, CIA2A, but not CIA1 and CIA2B, was specifically responsible for the Fe-S maturation of human IRP1, whereas CIA2B inhibited this process in an unknown manner (Figure 1.5). Furthermore, the IRE-binding capability and protein levels of IRP2 was assessed under the context of CIA1, CIA2A, CIA2B, and MMS19 knockdown [56]. Similarly, reduced CIA1 had no impact on IRE-binding or protein levels of IRP2. However, surprisingly, knocking down CIA2A decreased both IRE-binding and protein levels of IRP2. Considering the high affinity interaction between CIA2A and IRP2, it was plausible that this binding stabilizes IRP2 (Figure 1.5). On the other hand, the IRE-binding activity and protein amount were increased by the knockdown of either CIA2B or MMS19, leading to reduced translation of ferritin and increased transferrin receptor binding level. This may imply that CIA2B and MMS19 affect the stability of IRP2 in an unidentified manner (Figure 1.5) [56].

### 1.3.6 IRP1 and IRP2 respond to oxygen levels differentially

As previously mentioned, with the presence of sufficient iron in mammals, Fe-S triggers the conformational transition of IRP1, enabling it to become a cytosolic aconitase (Figure 1.3) [58]. By contrast, IRP2 (and apo-IRP1), which solely possesses IRE-binding activity, is subjected to proteasomal degradation [59]. This degradation is also triggered by high oxygen levels. The regulation of IRP2 in response to iron and oxygen levels is mediated by the adaptor protein FBXL5, which contains an F-box domain and a hemerythrin domain. The F-box domain enables FBXL5 to interact with the SCF ubiquitin ligase complex, while the hemerythrin domain in the N terminus is able to regulate the stability of FBXL5 in an iron- and oxygen-dependent manner [60, 61]. In iron-replete and oxygenated cells, an iron-oxygen center within the hemerythrin domain binds to both Fe and $O_2$, stabilizing FBXL5 and leading to increased ubiquitination of IRP2, ultimately

promoting its degradation [30]. Conversely, when cells are treated with an iron chelator or grown under hypoxia conditions, the level of FBXL5 protein decreases, and thus IRP2 accumulates, resulting in an elevated level of IRE-binding activity [30, 61]. The IRE-binding ability of IRP1 is diminished under hypoxic conditions, and the protein amount of IRP1 remains unchanged [62]. This is attributed to the tendency of IRP1 to adopt the cytosolic aconitase form under conditions that reduces oxygen or ROS levels [33, 63].

### 1.3.7 NO and $H_2O_2$ regulate iron homeostasis via different manners

Due to the labile [4Fe-4S] cluster within IRP1, which is sensitive to oxidative environments, IRP1 has been identified as a sensor for both reactive nitrogen species (RNS) and reactive oxygen species (ROS) [32, 64]. The presence of exogenous or endogenous nitric oxide produced by nitric oxide synthetase triggers the disassembly of [4Fe-4S] cluster in holo-IPR1, resulting in the loss of aconitase activity of IRP1 and the enhancement of IRE-binding activity of apo-IRP1 (Figure 1.4) [65-67]. In contrast, the IRE-binding capability of IRP2 remains unaffected by NO exposure [68]. As a consequence, the mRNA and protein levels of TfR1 are increased by apo-IRP1. Interestingly, despite the expected decrease in mRNA levels upon IRP1 binding, exposure to NO actually increases the mRNA levels of ferritin's heavy and light chain and FPN. Knocking out either IRP1 or IRP2 cannot alter this upregulation effect, indicating that NO does not upregulate their mRNA levels through IRPs. This is further demonstrated by treating cells with a transcriptional inhibitor, which led to a decrease in the mRNA levels of ferritin and FPN, meaning NO enhances their mRNA levels transcriptionally [68].

On the other hand, treatment of cells with ROS, such as hydrogen peroxide, activates the IRE-binding function of IRP1 within a short period of time (~5 min) (Figure 1.4), while IRP2's IRE-binding activity remains unchanged [69]. By measuring the aconitase activity and released iron content from IRP1 concomitantly, it has been determined that the induction of IRE-binding capability is due to complete disassembly of the Fe-S cluster within IRP1 [70]. However, this activation of IRP1 solely affects the mRNA stability of TfR1 and increases TfR1 protein content, leaving ferritin mRNA levels unchanged but with reduction in protein level. This indicates that $H_2O_2$ regulates TfR1 specifically through IRP1 [71]. Furthermore, the involvement of phosphorylation and unknown signaling pathways in the induction of IRE-binding activity by

$H_2O_2$ has been suggested by blocking the activation of IRP1 with a phosphatase inhibitor (e.g., okadaic acid) or using cell extracts in these experiments [32, 33, 71, 72]. Interestingly, mitoNEET, a protein located in the outer membrane of mitochondria and possessing a Fe-S cluster resistant to oxidative stress, is able to repair the oxidatively damaged Fe-S in IRP1 via direct transfer of the iron-sulfur cluster (Figure 1.5) [73].

## 1.4 Iron metabolism in *Drosophila melanogaster*

### 1.4.1 Iron uptake and transport in flies

Although the importance of iron in flies and its role in ecdysone synthesis are well recognized, the detailed *Drosophila* iron metabolism pathway is not as clearly understood as the pathways in mammals. However, current data suggests that there are substantial similarities between fly and mammalian iron metabolism. For example, dietary ferric iron is first absorbed in the gut of fly by an iron importer Malvolio (Mvl). Nonetheless, the coupling ferric reductase has not yet been determined, even though three candidates are identified: CG1275 and nemy (no extended memory) and CG8399 [1, 74]. In terms of iron export from the enterocyte, fly counterparts for mammalian ferroportin have not been identified. Nevertheless, four genes, *MCO1-4* (*multi-copper oxidase 1-4*), are homologous to human hephaestin, a protein required for loading iron onto transferrin. Of these, only MCO3 is a potential candidate with a potent ferroxidase activity *in vitro* [74, 75].

Transferrin-1 (Tsf1), a homolog of human transferrin, participates in iron transport in flies, but its receptor has not been discovered in *Drosophila*. Interestingly, the role of ferritin in flies appears to be different from its role as an iron repository in mammals, as it may have a role in iron trafficking in flies [76, 77]. Ferritin is a nanocage-shape protein composed of 24 subunits of ferritin heavy and light chains. Its primary function in humans is to store iron, capable of accommodating up to 4500-5000 iron atoms [78, 79]. In contrast, in flies, when ferritin is specifically knocked down in the midgut, where iron absorption takes place, it leads to iron accumulation in the midgut but reduced iron content in the whole body, causing delayed development and low survival rate of flies [76]. These findings suggest a potential involvement of fly ferritin in the delivery of iron from the gut to other tissues.

### 1.4.2 IRP1A is the main regulator of cellular iron homeostasis in *Drosophila*

At the cellular level, iron concentration in *Drosophila* is also modulated by the IRP/IRE system (Figure 1.3). The IRPs in *Drosophila* are termed IRP1A and IRP1B, both of which are homologs of human IRP1, sharing 78% and 79% similarity with human IRP1, respectively [6]. There is no IRP2 homolog in flies. Similar to human IRP1, IRP1A is bifunctional, possessing both RNA-binding and aconitase activities. Under iron-deficient conditions, IRP1A binds to IREs to regulate iron metabolism. So far, only two IREs have been identified in flies: the heavy chain of ferritin and SdhB (Succinate dehydrogenase subunit B) [38, 80]. These two IREs locate in the 5′-UTR, and no 3′-IRE has been discovered thus far. Conversely, under iron-rich conditions, IRP1A incorporates Fe-S, undergoes conformational changes, and functions as a cytosolic aconitase (Figure 1.3). In contrast, even though IRP1B is highly similar to IRP1A, sharing 86% identity and 93% similarity, it can only bind to iron-sulfur clusters and act as aconitase, but it does not possess mRNA-binding activity [6, 81].

## 1.5 IRP1A/B show unexpected nuclear localization

### 1.5.1 The mutants of IRP1A/B

Fly IRP1A/B share a high degree of sequence identity with human IRP1, particularly in the conserved active sites responsible for iron-sulfur binding (Figure 1.6 A) and IRE binding (Figure 1.6 B) [81]. As mentioned in the previous section (1.3.3), cysteine at position 437 of human IRP1 plays a crucial role in iron-sulfur binding [40]. Mutating this cysteine to serine (hIRP1$^{C437S}$) abolishes the ability to bind to Fe-S clusters, rendering IRP1 constitutive IRE-binding [82, 83]. This dysregulated binding leads to elevated levels of transferrin receptor (TfR) and suppression of ferritin synthesis, causing iron overload and cellular toxicity. Similarly, corresponding mutants have been created in IRP1A and IRP1B, referred to as IRP1A$^{C450S}$ and IRP1B$^{C447S}$, respectively (Figure 1.6 A) [19]. Like hIRP1$^{C437S}$, these mutants exhibit impaired iron-binding ability and are predicted to primarily bind to RNA. Although the RNA-binding site is conserved between IRP1B and human IRP1 (Figure 1.6 B), IRP1B cannot exhibit IRE binding. Consequently, IRP1B$^{C447S}$ exists predominantly in the apo-form and cannot bind to RNA (Figure 1.6 A).

In human IRP1, three arginine residues (R536, R541, and R780) are essential for mRNA binding (Figure 1.6 B) [35, 40]. Substituting any of these residues with glutamine results in the loss of RNA-binding function, and IRP1 is expected to predominantly adopt the holo-form (aconitase). In IRP1A, all three corresponding arginine residues are mutated to glutamine (IRP1A-R549Q, R554Q, R793Q), forming the IRP1A$^{3R3Q}$ mutant (Figure 1.6 B) [19]. The counterpart mutant in IRP1B is called IRP1B$^{3R3Q}$ (IRP1B-R546Q, R551Q, R790Q). Both of these 3R3Q mutants are predicted to abolish IRE binding and primarily exist as the holo-form.

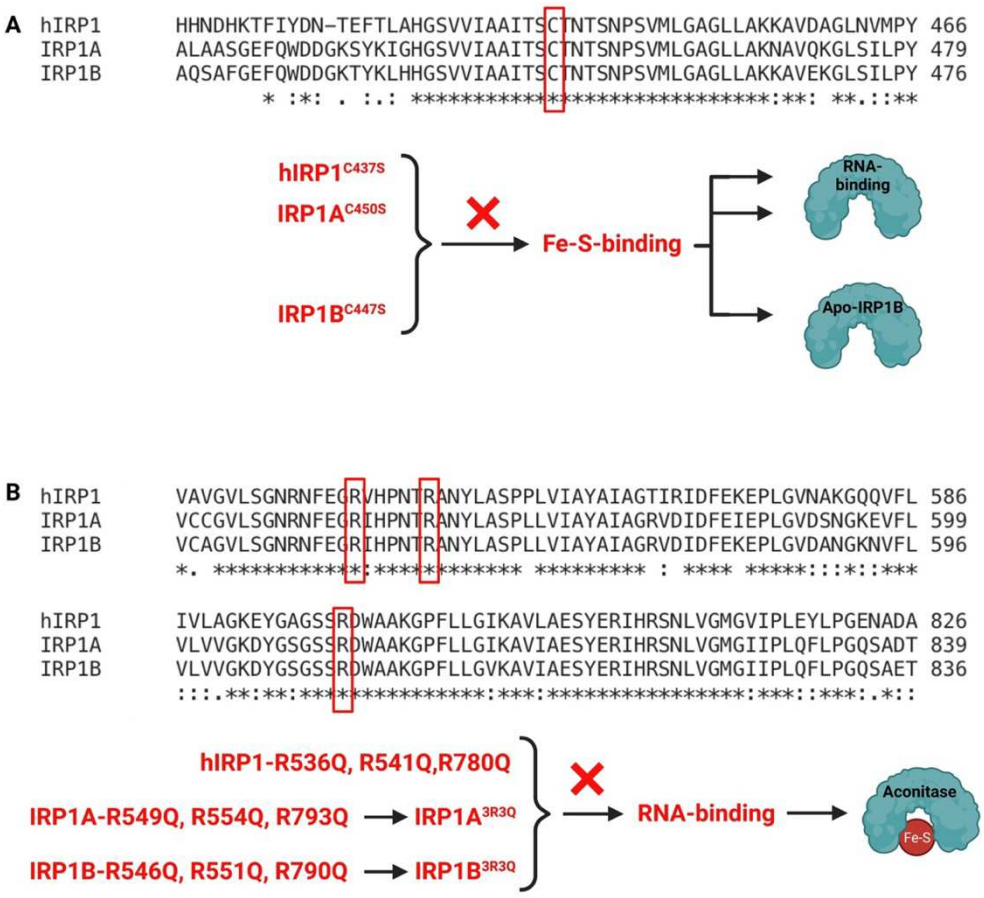



**Figure 1.6 The mutants of IRP1A/B**

The active sites for iron-sulfur binding (A) and IRE binding (B) are highlighted in red boxes. (A) The C437S point mutation in human IRP1 eliminates its ability to bind to Fe-S clusters, resulting in constitutive RNA binding. The counterpart mutants in IRP1A and IRP1B are referred to as IRP1A$^{C450S}$ and IRP1B$^{C447S}$, respectively. These mutants are predicted to be mainly apo-form. (B) Three arginine residues, R536, R541, and R780, are essential for the RNA-binding activity of human IRP1. Substituting any of these residues with glutamine abolishes the RNA-binding function, leading to the predominant adoption of the aconitase form by IRP1. Similarly, the corresponding mutations in IRP1A (R549Q, R554Q, R793Q, termed IRP1A$^{3R3Q}$) and IRP1B (R546Q, R551Q, R790Q, called IRP1B$^{3R3Q}$) result in the predominance of the aconitase form. Protein sequences were obtained from the UniProt database (https://www.uniprot.org) and the sequence alignment was done by Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

### 1.5.2 Only holo-IRP1A/B can enter nuclei

In the past 30 years, IRP1A was believed to be located only in the cytoplasm. However, this paradigm has been challenged by a new finding that fly IRP1A/B can enter the nucleus (Figure 1.7) [19]. One experiment using PG-specific overexpression of IRP1A indicated that wild-type IRP1A, which can interconvert between the apo- and holo-form, predominantly localizes to the nucleus of PG cells. On the other hand, $IRP1A^{C450S}$, which loses the Fe-S-binding function and thus assumes the apo-form, resided in the cytosol only [19, 40]. This was also true for IRP1B and its equivalent mutant $IRP1B^{C447S}$, with the wild-type IRP1B was found in both cytoplasm and nucleus of PG cells, $IRP1B^{C447S}$ was primarily cytosolic. Taken together, the *Drosophila* IRP1A/B are able to enter the PG nucleus and this special subcellular localization favors the holo-IRPs (represented as $IRP1A^{3R3Q}$ and $IRP1B^{3R3Q}$ in Figure 1.7). This was further proven by the fact that fly IPRs were purged from nuclei after rearing flies in BPS (bathophenanthrolinedisulfonic acid disodium) food for several generations [19]. The negatively charged BPS is an iron chelator that can compete for iron in cells and form an Fe(II)(BPS) complex [84]. When flies are fed on BPS food for generations, iron is depleted, thus forcing fly IRPs to remain in the apo-form [19].

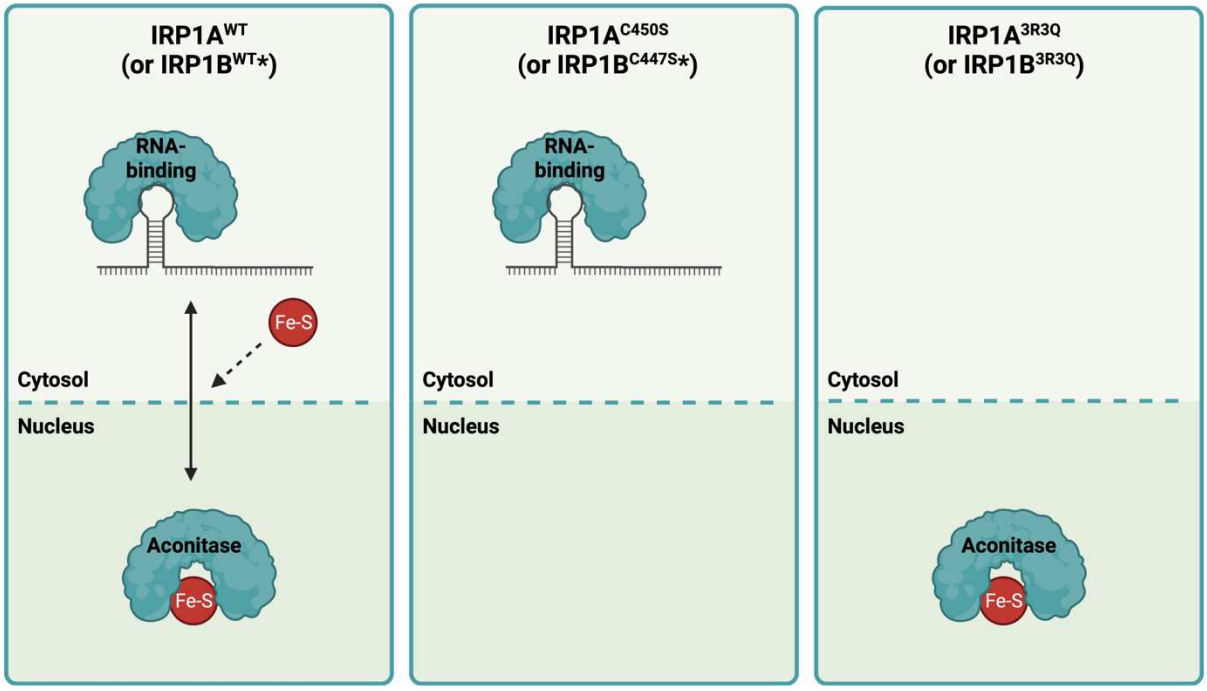



**Figure 1.7 Subcellular localization of fly IRPs**

The apo-form of IRP1A$^{WT}$ is RNA-binding and resides mainly in the cytoplasm. When bound to the [4Fe-4S] cluster (Fe-S), it undergoes conformational changes and becomes the holo-form, which has aconitase activity and can enter the nucleus. The apo- and holo-form of wild-type IRP1B have the same subcellular localization as IRP1A$^{WT}$, except for the apo-IRP1B cannot interact with RNA. The apo-variants, IRP1A$^{C450S}$ and IRP1B$^{C447S}$, are primarily cytosolic. In contrast, IRP1A$^{3R3Q}$ and IRP1B$^{3R3Q}$ with impaired IRE-binding activity tends to exist in the aconitase form and localizes mainly in the nucleus. WT, wild-type; *, apo-IRP1B$^{WT}$ and IRP1B$^{C447S}$ stay in the cytoplasm but do not possess RNA-binding activity.

### 1.5.3 The nuclear subcellular location of IRPs is tissue-specific

The startling observation that IRP1A/B can localize to PG nuclei raises the question as to whether this was true for all *Drosophila* tissues. By visualizing IRPs via immunofluorescence in a range of tissues, such as the fat body and the salivary gland, the nuclear localization was found to be tissue-specific[19]. For instance, in the fat body, IRPs behaved in the same way as in PG cells, distributing in the cytosol as well as in the nucleus. In contrast, in the salivary gland, IRP1A/B appear to be cytoplasmic. Moreover, a recent study revealed that human IRP1 is also capable of nuclear translocation in a cell- and iron-specific manner, entering Huh7 and HepG2 hepatoma cells, but not HeLa cancer cells [85].

Taken together, considering the essentiality of iron and the intriguing nuclear localization of fly IRPs, which is not limited to the fruit fly alone, it is of great interest to explore their nuclear role of the IRP1A/B holo-forms. Thankfully, the PG provides an excellent platform for investigating this subject. Through this investigation, we hope to uncover a novel iron regulation network in the nucleus, challenging the conventional IRP/IRE paradigm, and enriching our understanding of iron homeostasis.

## 1.6 Thesis outline

### 1.6.1 Investigating the molecular function of nuclear IRP1A

There are various possibilities regarding the role of IRP1A in the nucleus, including potential effects on gene regulation and RNA splicing (to be discussed further in the corresponding chapter). Given that most nuclear processes are tied to gene regulation, it seems plausible that IRP1A may play a role in regulating gene expression. Therefore, to begin with, I explored IRP1A's function at the molecular level by profiling its chromatin binding pattern.

### 1.6.2 Identifying the mechanism of nuclear IRP1A function

Upon identifying the chromatin loci bound by IRP1A, I focused my attention on how IRP1A accomplishes the interaction with chromatin. To understand this, I characterized the protein-protein interaction network of fly IRPs with IRP1B serving as a control to distinguish it from IRP1A.

### 1.6.3 Assessing the viability of IRP1A/B mutants

Apart from the molecular function and mechanism, I generated CRISPR knock-in and knock-out lines to gain deeper insights into the functions of IRP1A and IRP1B by measuring the survival rates of IRP1A/B mutants. These genetic tools provide a way to assess the importance of cytosolic and nuclear IRP1A by comparing the viability of *IRP1A*$^{C450S}$ and *IRP1A*$^{3R3Q}$ mutants. Additionally, the *IRP1A*$^{C450S/3R3Q}$ variant can be used to determine if the conformational transition of IRP1A is necessary for fly survival. Moreover, the role of IRP1B in iron metabolism can be identified through measuring the survival rate of IRP1B mutants (*e.g.,* IRP1B$^{+/-}$ and IRP1B$^{-/-}$) under an IRP1A knock-out context. Furthermore, the wild-type knock-in tagged line of IRP1A will allow the characterization of chromatin binding profile endogenously.

# Chapter 2 Materials and methods

## 2.1 *Drosophila* husbandry and maintenance

A list of fly lines that have been used in my project is provided in Table 2.1. All fly stocks were maintained at 18°C on a standard corn-based media. Two backup vials were kept and transferred to new food every four weeks.

Experimental flies were reared at 25°C with humidity around 50-60% on the "Nutrifly" media (Genesee Scientific, Catalog (Cat) # 66113). A standard recipe for this media was provided by the Bloomington *Drosophila* Stock Center (BDSC): 4.25 litters (L) Nutrifly food contains 67.5 grams (g) of yeast, 39 g of soy flour, 285 g of yellow cornmeal, 22.5 g of agar, 0.3 L of light corn syrup, and 18.8 milliliters (mL) of and propionic acid. To prepare 100 mL Nutrifly media, 17.8 g of Nutrifly powder was weighed and boiled in 100 mL autoclaved Mili-Q water for 5 to 10 minutes (min). After the food had cooled down to 50-60°C, 450 microliters (µL) of propionic acid (Sigma, Cat. # P1368) was added per 100 ml liquid fly food. After thoroughly mixing, the food was poured into vials, dishes, or bottles. Food was prepared 1 day before experiments allowing water vapours to escape and subsequently stored at 4°C for up to six weeks. For crosses, 30 flies were kept in one vial (or 200 flies in one bottle) with a male to female ratio of 1: 2.

## 2.2 Genomic DNA extraction

Pestles used for grinding were cooled in liquid nitrogen. A total of 50 flies were collected in a 1.5 mL tube and rapidly frozen in liquid nitrogen. The flies were ground to a fine powder using pre-cooled pestles, ensuring homogeneity. For each sample, 200 µL of DNAZol (Thermo Fisher Scientific, Cat. # 10503027) was added, and the sample was homogenized for 15 seconds (s) using a pestle. This step was repeated two more times. Following this, 400 µL of DNAZol was added to each sample, resulting in a total volume of 1 mL. The sample was vortexed for 15 s and left at room temperature for 5 min. Centrifugation was performed at 13,000 rpm for 15 minutes at 4°C. The resulting green viscous supernatant (~700 µL) was carefully transferred to a fresh 1.5 mL tube.

To extract the DNA, 700 µL of PCI (Phenol:Chloroform:Isoamyl Alcohol, 25:24:1; v/v; Invitrogen, Cat. # 15593031) was added to each sample in a 1:1 ratio. The samples were thoroughly vortexed or hand-shaken for approximately 20 seconds and then centrifuged at room temperature

for 5 min at 13,000 rpm. The solution separated into three distinct phases. The upper aqueous phase (viscous greenish aqueous phase), which contains the DNA, was cautiously removed without carrying over any phenol and transferred to a new tube (~200 µL). If the aqueous phase was not clear, this step was repeated until clarity was achieved. Subsequently, 200 µL of chloroform was added to each sample in a 1:1 ratio, mixed by inverting the tubes 10 times, and centrifuged for 5 min at 13,000 rpm at 4°C. The upper aqueous phase was then transferred to a new tube (~200 µL).

For DNA precipitation, 20 µL of 3 M (molar) sodium acetate (pH 5.5) was added to each sample in a 1:10 ratio and mixed thoroughly by inverting tubes. Next, 500 µL of cold 100% ethanol was added to each sample in a 2.5:1 ratio and mixed by shaking on a shaker for 3 minutes. The samples were then incubated on ice for 5 minutes and subsequently centrifuged at 13,000 rpm for 10 minutes at 4°C to pellet the DNA. The supernatant was carefully discarded, and the DNA pellet was washed by adding 0.8-1.0 mL of 75% ethanol and inverting the tube 3-6 times. After centrifugation for 2 minutes at 13,000 rpm and 4°C, the supernatant was discarded, and the DNA pellet was air-dried for 5 minutes at room temperature. Finally, the DNA pellet was dissolved in 200 µL of TE buffer or nuclease-free water by gently passing the pellet through a pipette tip.

### 2.3 Gibson assembly

DNA fragments were obtained by either PCR (polymerase chain reaction) or digestion of plasmids with proper restriction enzymes. Unless specified, all PCR reactions in this study were done by Q5 High-Fidelity DNA Polymerase (New England Biolabs (NEB), Cat. # M0491) following the protocol provided by NEB. After PCR or digestion, DNA fragments were purified using the QIAquick Gel Extraction Kit (Qiagen, Cat. # 28704) following the manufacturer's instructions. The concentration of DNA fragments was measured by a Nanodrop Spectrophotometer (NanoDrop ND-1000). For plasmids smaller than 10 kb, 200 ng of vector DNA was used. The amount of insert DNA was calculated by the NEB calculator (https://nebiocalculator.neb.com/#!/ligation) at a 3:1 ratio of insert DNA to vector DNA. For bigger plasmids (>10 kb), 300-400 ng of vector DNA was required. The volume ratio of DNA fragments to homemade Gibson Assembly Mastermix (see appendix A 1 for detailed recipe) was

1:3. To achieve a high transformation rate, the whole volume of Gibson mix did not exceed 20 µL. After addition of DNA, incubate Gibson assembly mix at 50 ˚C for 2-4 hours (h).

## 2.4 Preparation of ultra-competent cells

The ultra-competent cell protocol was adapted from Cold Spring Laboratory and Princetone University. To prepare 1 L of competent cells, the necessary reagents and media were prepared as described in appendix A 2.

Day 1: A single tube of competent cells (*DH5α Escherichia.coli (E.coli)*) was obtained from the laboratory stock and subsequently spread on an LB plate without antibiotics. The plate was then incubated at 37℃ for a duration of 16-20 hours (h).

Day 2: A single colony from the plate was inoculated into 5 mL of SOB medium without antibiotics, establishing a starter culture. The starter culture was incubated at 37℃ with vigorous shaking at 250-300 rpm for 8 h. Subsequently, 2.5 mL of this starter culture was inoculated into a 1 L culture, which was incubated at 18-22℃ with moderate shaking at 180-200 rpm for approximately 14 hours.

Day 3: The bacterial culture's optical density at 600 nm (OD600) was measured. Once the OD600 reached the range of 0.54-0.55, the culture flask was transferred to an ice-water bath and maintained for 10 minutes, ensuring uniform cooling through gentle shaking. The cells were then harvested by centrifugation at 3900 rpm for 10 minutes at 2℃. The cells were gently resuspended in 80 mL of ice-cold Inoue transformation buffer, followed by centrifugation at 3900 rpm for 10 minutes at 2℃. After discarding the buffer, the cells were gently resuspended in 20 mL of ice-cold Inoue transformation buffer, and the bacterial suspension was transferred to a pre-chilled 50 mL conical tube. To the suspension, 1.5 mL of DMSO (dimethyl sulfoxide, Sigma-Aldrich 276855) was added. The bacterial suspension was mixed by swirling and stored on ice for 10 minutes. Subsequently, the suspension was aliquoted into pre-chilled, sterile 1.5 mL tubes (50-100 µL) and immediately snap-frozen in liquid nitrogen. The competent cells were stored at -80℃ until needed.

## 2.5 Bacteria transformation

SOC media was prepared fresh according to the recipe provided in in A 2. The competent cells were retrieved from the -80°C freezer and allowed to thaw on ice for 20-30 min. For each transformation, 2 µL of the Gibson assembly mix or 10 pg - 100 ng of plasmids were added to 50-100 µL of competent cells. The contents of the tube were gently mixed by flicking the bottom of the tube with a finger a few times. The competent cell/DNA mixture was then incubated on ice for 20-30 minutes.

To induce heat shock, each transformation tube was placed into a 42°C water bath for 30-60 s (usually 45 seconds is ideal, but the optimal duration may vary depending on the specific competent cells being used). The tubes were promptly returned to ice and left for 2 minutes. Next, 950 µL of SOC media without antibiotic was added to the bacteria, and the mixture was incubated in a shaking incubator at 37°C for 1 h. The bacteria were then centrifuged at 4000 rpm for 2 minutes, and 900 µL of the media was discarded. The cells were resuspended, and the entire transformation mixture was plated onto a 10 cm LB agar plate containing the appropriate antibiotic. In the case of plasmid transformation, centrifugation was not necessary, and instead, 50-100 µL of the transformation mix was directly plated onto an agar plate with the suitable antibiotic. The plates were subsequently incubated at 37°C overnight.

For Gibson assembly transformations, 8 colonies were selected for colony PCR using specific primers. The PCR products were analyzed on an agarose gel to verify if the plasmid was constructed as expected. In the case of successful amplification of the desired band, at least 2 constructs corresponding to the confirmed colonies were expanded and sent for Sanger sequencing. Regarding plasmid transformations, 2 colonies were picked and cultured in 5 mL of LB media supplemented with the appropriate antibiotics for subsequent plasmid extraction. To confirm the presence of the desired plasmid, appropriate restriction enzyme(s) capable of generating a unique agarose gel pattern were employed for verification purposes.

### 2.6 Plasmid extraction

To extract plasmids from *E· coli*, two kits were used. For a small number of plasmids, a single colony was inoculated into 5 mL of LB media, with the addition of antibiotics if required. The culture was then incubated overnight in a 37°C incubator with shaking at 200 rpm. The GeneJET Plasmid Miniprep Kit from ThermoFisher Scientific (Cat. # K0503) was employed to

extract up to 20 µg of plasmids from this culture. Alternatively, when a larger quantity and higher quality of plasmids were needed for purposes such as transfection or embryo injection, a single colony was inoculated into 100 mL of LB media containing the appropriate antibiotics and incubated overnight. Plasmids were isolated with the QIAGEN Plasmid Midiprep kit (QIAGEN, Cat. # 10043). All extraction processes followed the guidelines provided by manufacturers.

## 2.7 Schneider 2 (S2) cell constructs for transfection

To generate constructs for S2 cell transfection, plasmids *pAFW* (Addgene), *pAMW* (Addgene) and *pAc5-STABLE2-neo* (Addgene, Cat. # 32426) were used. All plasmids ordered from companies are listed in Table 2.2. The full-length cDNAs of *IRP1A* and *H2Av* were PCR amplified from a reverse-transcribed cDNA library of $w^{1118}$ whole larvae. The cDNA of *EGFP* was amplified from *pAc5-STABLE2-neo*.

A plasmid harbouring a C-terminal 3X Flag coding sequence (*pAFW-C-3xFlag*) was derived from pAFW and then used to generate Flag-tagged IRP1A (*pAFW IPR1A-C-3xFlag*) by Gibson assembly. A list of plasmids generated in this study is shown in Table 2.3. *H2Av* and *EGFP* were subcloned into *pAMW* to encode C-terminal and N-terminal 6X Myc-tagged proteins (*pAMW H2Av-C-6xMyc*, *pAMW N-6xMyc-EGFP*), respectively. The tagged *IPR1A* and *H2Av* (or *EGFP*) were subcloned into *pAc5-STABLE2-neo* with the T2A sequence in between for co-expression (*pSTABLE2 IRP1A-C-3xFlag-H2Av-C-6xMyc*, *pSTABLE2 IRP1A-C-3xFlag-N-6xMyc-EGFP*). All primers used for S2 cell constructs are listed in appendix A 3.

## 2.8 Generating IRP1A$^{3R3Q}$ and IRP1B$^{3R3Q}$ overexpression lines

To generate the UAS-driven IRP1A$^{3R3Q}$ overexpression line, I first created two point mutations (R549Q and R554Q) in *IRP1A* using site-directed mutagenesis with customized primers, resulting in the intermediate *pAFW-IRP1A*$^{2R2Q}$ vector. R793Q was introduced into *pAFW-IRP1A*$^{2R2Q}$ in the same way to generate *pAFW-IRP1A*$^{3R3Q}$. The cDNA of *IRP1A*$^{3R3Q}$ was then amplified and ligated with backbone fragments cloned from *pBID-UASC-FG* (Addgene, Cat. # 35201) using the Gibson assembly (*pBID-UASC-FG-IRP1A*$^{3R3Q}$). This plasmid, which contained a mini-*white* marker, was sequenced, and sent to the GenetiVision Corporation (Houston, Texas, US) for embryo injection into $w^{1118}$ at attP40 (phage attachment site 40). attP40 is a docking site

in chromosome II of the fly genome. Recognized and catalysed by by φC31 integrase, the site-specific recombination can occur between an attP and attB (bacterial attachment site, present in the integrating plasmid) [86]. All fly lines constructed in this study are listed in Table 2.4. The *IRP1B*$^{3R3Q}$ transgenic line was established in a similar fashion. The wild-type cDNAs of *IRP1B* was amplified from a reverse-transcribed cDNA library of $w^{1118}$ whole larvae. Point mutations were first introduced into *pAFW-IRP1B*$^{2R2Q}$ and *pAFW-IRP1B*$^{3R3Q}$, followed by subcloning the *IRP1B*$^{3R3Q}$ cDNA into *pBID-UASC-FG* for embryo injection. All primers used in this section are listed in appendix A 4. After embryo injection, positive male flies carrying mini-white were crossed with $w^{1118}$ females and then sent to us with larvae. Male progenies with red eyes from the first generation then crossed with a balancer line ($w^{-}$*; CyO/Ser; Tb/TM6, Hu*). Flies with red eyes and curly wings in the second-generation were selected, which was the heterozygous line. The heterozygous flies were crossed to make the homozygous line, where the flies were with red eyes and non-curly wings.

## 2.9 Generating TurboID lines

Backbone fragments of TurboID plasmids were cloned from *pUAS-V5-TurboID-NES* (Addgene, Cat. # 116904) with the NES (nuclear export signal) sequence removed. Then the wild-type IRP1A cDNA was ligated with TurboID backbone fragments by Gibson assembly to generate *pTurboID-IRP1A*$^{WT}$. The cDNA of *IRP1A*$^{C450S}$ was cloned from the genomic DNA of *UAS-3F-IRP1A*$^{C450S}$ transgenic line, and then subcloned into the TurboID backbone plasmid to construct *pTurboID-IRP1A*$^{C450S}$. *pTurboID-IRP1B* was generated in the same way. Primers used are listed in appendix A 5. All these plasmids were sequenced and expanded by Midiprep, and then sent to the GenetiVision Corporation for embryo injection into $w^{1118}$ at attP40 docking site in the second chromosome. After embryo injection, positive male flies carrying mini-white were crossed with $w^{1118}$ females and then sent to us with larvae. Male progenies with red eyes from the first generation were then crossed with a balancer line ($w^{-}$*; CyO/Ser; Tb/TM6, Hu*). Progeny flies with red eyes and curly wings were picked, this was the heterozygous line. Balanced heterozygotes were inbred and flies with red eyes and non-curly wings were selected in the progeny to generate the homozygous line.

## 2.10 Generating CRISPR lines

### 2.10.1 Construction of gRNA plasmids

Sequences of target regions were obtained from FlyBase (http://flybase.org) and submitted to websites called DRSC Find CRISPRs (https://www.flyrnai.org/crispr/) and flyCRISPR Target Finder (https://flycrispr.org/target-finder/) to find optimal gRNA target sites. The forward and reverse gRNA sequences were designed into primers used to amplify gRNA scaffold from *pCFD5* (Addgene, Cat. # 73914) [87]. The backbone of *pCFD5* was prepared by restriction enzyme digestion using BbsI HF (NEB, Cat. # R3539S) at 37°C for 1 hour. The gRNA and backbone fragment were fused by Gibson assembly to generate gRNA constructs. See appendix A 6 for the full list of primers.

### 2.10.2 Construction of donor plasmids

Plasmid *pDsRed-attP* (Addgene, Cat. # 51019) carrying a DsRed marker was digested by EcoRI (Thermo Fisher Scientific, Cat. # FD0274) to produce a backbone for donor DNA construction. All homologous arms and target regions were amplified from genomic DNA of $w^{1118}$ adults. The 3xFlag tag was cloned from *pAFW*. For IRP1A knock-in lines, the homologous left arm and target region of wild-type IRP1A were first subcloned into *pDsRed-attP* (*pDsRed-left arm-IA*). Then the right arm was amplified using *Taq* 2X Master Mix (NEB, Cat. # M0270L) and inserted into *pDsRed-left arm-IA* to generate *pDsRed-IRP1A*$^{WT}$*-3F*. Point mutations were introduced by site-directed mutagenesis for *pDsRed-IRP1A*$^{C450S}$*-3F* and *pDsRed-IRP1A*$^{3R3Q}$*-3F*. In terms of IRP1A knock-out lines, the whole gene was deleted. See appendix A 6 for the full list of primers used in this section.

Corresponding gRNA and donor plasmids were verified by Sanger sequencing and sent to the GenetiVision Corporation for embryo co-injection into *nos-Cas9* line [88]. After embryo injection, positive male flies carrying DsRed were crossed with $w^{1118}$ females and then sent to us with larvae. Male progenies with red eyes under the UV light from the first generation then crossed with a balancer line ($w^{-}$*; CyO/Ser; Tb/Tm6, Hu*). *Tubby* pupae with redness under the UV light in the second-generation were selected, which was the heterozygous line. Balanced heterozygous flies were inbred, non-tubby pupae with DsRed were picked, this was the homozygous line. The IRP1A knock-out line can only be kept as heterozygotes and can be crossed with a Cre recombinase line to remove the DsRed marker.

## 2.11 Culturing S2 cells

The culture of *Drosophila* S2 cells was carried out following the guidelines provided in the Thermo Fisher Scientific "Schneider 2 (S2) Cells User Guide". All incubations were conducted in a 26°C–28°C incubator without the need for CO2. The complete medium for S2 cells consisted of Schneider's *Drosophila* Medium (Thermo Fisher Scientific, Cat. # 21720024) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Thermo Fisher Scientific, Cat. # A4766801). Prior to transfection or freezing, cell counting was performed to ensure accurate seeding and assess viability using a hemacytometer and trypan blue staining. Viability of S2 cells in culture was maintained at 95–99%. Fresh flasks or plates were always used for general maintenance when passaging cells. For transfection and selection purposes, cells were kept in the same culture vessel. The doubling time of S2 cells was approximately 24 hours when routinely passaged during the log to mid-log phase of growth.

To initiate cell culture from a frozen stock, a vial of cells was rapidly thawed at 30°C. Before complete thawing, the exterior of the vial was decontaminated with 70% ethanol. The contents of the cryovial were transferred into a sterile conical tube containing 4 mL of complete Schneider's *Drosophila* Medium at room temperature. The tube was centrifuged at 100 × g for 5–10 min, and the medium containing DMSO was aseptically discarded. The cell pellet was resuspended in 5 mL of fresh complete Schneider's *Drosophila* Medium and transferred to a T-25 cm2 flask. The flask was then incubated in a non-humidified, ambient air-regulated incubator or warm room at 28°C. The flask caps were loosened to allow for oxygenation and aeration. Incubation at 28°C continued until the cells reached a density of $0.6 \times 10^7$–$2.0 \times 10^7$ cells/mL, which typically took 3–4 days.

For passage of S2 cells, the cells started to clump at a density of approximately $5 \times 10^6$ cells/mL in serum-containing medium, but this did not affect their growth. Clumps could be broken up during the passage process. S2 cells should be subcultured to a final density of $2 \times 10^6$-$4 \times 10^6$ cells/mL. When removing cells from the flask, cells were detached by tapping the flask and pipetting. Cells were split at a dilution of 1:2 to 1:5 into new culture vessels, followed by the addition of complete Schneider's *Drosophila* Medium. Incubation at 28°C continued until the

density reached $0.6 \times 10^6$-$2 \times 10^6$ cells/mL. This process was repeated as necessary to expand cells for transfection or expression purposes.

To freeze S2 cells, a Freezing Medium consisting of 45% conditioned complete Schneider's *Drosophila* Medium containing 10% heat-inactivated FBS (used medium), 45% fresh complete Schneider's Drosophila Medium containing 10% heat-inactivated FBS, and 10% DMSO was prepared. It was important to use conditioned medium and heat inactivated FBS to optimize cell recovery. When the cell density in a 75-$cm^2$ flask reached between $1 \times 10^7$-$2 \times 10^7$ cells/mL, the cells were removed from the flask, resulting in approximately 12 mL of cell suspension. A sample of cells was counted using a hemacytometer to determine the actual cells/mL and viability (95–99%). The cells were pelleted by centrifugation at 1000 × g for 2-3 min in a table-top centrifuge at 4°C. The conditioned medium was reserved, and the cells were resuspended in 10 mL of PBS and pelleted again at 1000 × g for 2–3 minutes. The cells were resuspended at a density of $1.1 \times 10^7$ cells/mL in the Freezing Medium, and 1 mL of the cell suspension was aliquoted per vial. The cells were frozen using a controlled-rate freezer to -80°C, or alternatively, they were initially placed at 4°C for 30 minutes, then at -20°C for 30 minutes before transferring to -80°C storage.

### 2.12 S2 cell transfection

The transfection of S2 cells was conducted following the "Schneider 2 (S2) Cells User Guide" provided by Thermo Fisher Scientific, using the calcium phosphate transfection method. To ensure proper protein expression, a time course experiment was performed by assaying for expression at 2, 3, 4, and 5 days post-transfection.

Day 1: Cultured cells were prepared for transfection by seeding $3 \times 10^6$ S2 cells ($1 \times 10^6$cells/mL) in a 35-mm plate containing 3 mL of complete Schneider's *Drosophila* Medium. The cells were incubated at 28°C for 6-16 hours until the cell density reached $2 \times 10^6$-$4 \times 10^6$ cells/mL.

Day 2: For transient transfection, the following transfection mix was prepared for each 35-mm plate. In a microcentrifuge tube, Solution A was prepared by combining 36 µL of 2 M $CaCl_2$ and 19 µg of recombinant DNA, and the volume was brought to 300 µL using nuclease-free water. In a separate microcentrifuge tube, Solution B was prepared by adding 300 µL of 2X HEPES-Buffered Saline (50 mM HEPES, 1.5 mM $Na_2HPO_4$, 280 mM NaCl, pH 7.1). Solution A was

slowly added dropwise to Solution B with continuous mixing, ensuring thorough mixing through techniques such as vortexing or bubbling air through the solution. The addition and mixing process were continued until Solution A was fully depleted, which typically took 1-2 minutes. Continuous mixing was employed to ensure the production of a fine precipitate required for efficient transfection. The resulting transfection solution was incubated at RT for 30-40 min, during which a fine precipitate formed. The solution was mixed and added dropwise to the cultured S2 cells in the plate, ensuring proper mixing with each drop. The cells were incubated at 28°C for 16-24 h, and an extended incubation period of 36-48 hours was investigated to assess potential improvements in transfection efficiency.

Day 3: After transfection, the calcium phosphate solution was carefully removed from the cells, and the cells were washed twice with complete Medium. Fresh complete Schneider's *Drosophila* Medium was added to the plate, and the cells were replated into the same vessel. The plate was incubated at 28°C until needed.

Day 4+: For gene expression analysis, cells were harvested at 2, 3, 4, and 5 days post-transfection. The expression of the target gene was assessed through appropriate assays.

### 2.13 S2 cell immunostaining and imaging

Cells were directly grown, treated, fixed, and stained in multi-well plates, chamber slides, or coverslips. The liquid in the wells was aspirated, and the cells were covered with a 2–3 mm depth of 4% formaldehyde diluted in warm phosphate-buffered saline (PBS) solution. The 4% formaldehyde solution was prepared by diluting 1 portion of 16% formaldehyde (methanol-free, Polysciences, Inc., Cat. # 18814) in 4 portions of 1X PBS. The cells were allowed to fix for 15 minutes at room temperature. Subsequently, the fixative solution was aspirated, and the cells were rinsed three times with 1X PBS for 5 minutes each.

To block the specimen, a blocking buffer (1X PBS, 5% Normal Goat Serum (Abcam, Cat. # ab138478), and 0.3% Triton X-100) was applied and incubated for 60 minutes. Meanwhile, the primary antibody, as specified in Table 2.5 for antibodies used in this study, was prepared by diluting it according to the datasheet instructions in an antibody dilution buffer (1X PBS, 1% BSA, and 0.3% Triton X-100). The blocking solution was aspirated, and the diluted primary antibody

was applied to the specimen. Incubation was performed overnight at 4°C with gentle shaking. Following primary antibody incubation, the specimen was rinsed three times with 1X PBS for 5 minutes each. For fluorochrome-conjugated secondary antibody incubation, the specimen was incubated in the dark with the appropriate secondary antibody (Table 2.5). The secondary antibody was diluted in the antibody dilution buffer and incubated for 1-2 h at room temperature. The specimen was then rinsed three times with 1X PBS for 5 minutes each. To label the nucleus, the specimen was incubated with DAPI (Cell Signaling, Cat. # 4083) at RT for 15 min, and then rinsed three times with 1X PBS for 5 minutes each. Samples were mounted in mounting buffer (1X PBS, 50% glycerol). Pictures were taken on Nikon Eclipse 80i Confocal C2+ microscope/camera.

## 2.14 *Drosophila* tissue immunostaining and imaging

Larvae were washed in 1X PBS for 5 min, and tissues were dissected in clean 1X PBS and then fixed in 4% formaldehyde for 20 minutes at room temperature. Subsequently, the fixative solution was aspirated, and the cells were rinsed three times with 1X PBST (1X PBS, 0.3% Triton X-100) for 10 minutes each. To block the specimen, a blocking buffer was applied and incubated for 60 minutes. Meanwhile, the primary antibody, as specified in Table 2.5 for antibodies used in this study, was prepared by diluting it according to the datasheet instructions in an antibody dilution buffer. The blocking solution was aspirated, and the diluted primary antibody was applied to the specimen. Incubation was performed overnight at 4°C with gentle shaking. Following primary antibody incubation, the specimen was rinsed three times with 1X PBST for 10 minutes each. For fluorochrome-conjugated secondary antibody incubation, the specimen was incubated in the dark with the appropriate secondary antibody (Table 2.5). The secondary antibody was diluted in the antibody dilution buffer and incubated for 1-2 h at room temperature. The specimen was then rinsed three times with 1X PBST for 10 minutes each. To label the nucleus, the specimen was incubated with DAPI at RT for 20 min, and then rinsed three times with 1X PBST for 10 minutes each. Samples were mounted in mounting buffer. Pictures were taken on Nikon Eclipse 80i Confocal C2+ microscope/camera.

## 2.15 S2 cells co-immunoprecipitation (co-IP)

The required reagents were prepared following the instructions outlined in A 7. For sample collection, two groups of samples were collected: a blank plasmid (control) group and a gene-

expressing plasmid group (e.g., 3X Flag-tagged genes). S2 cells were seeded on day 1 and transfected with the corresponding plasmids (20 µg plasmid per well of a six-well plate) on day 2. On day 3, the culture medium was replaced with fresh medium, and the S2 cells were collected on day 4. Cells were then washed once in cold PBS and centrifuged at 2000 × g for 2 min. Before storage, excess PBS was removed. The samples were then snap-frozen using liquid nitrogen and stored at -80°C until a sufficient number of samples were obtained. For mass spectrometry, at least two wells of a six-well plate were collected for each sample, while for co-immunoprecipitation, one well of a six-well plate was sufficient for each sample.

The samples were removed from the -80°C freezer and thawed on ice. To each sample, 400 µL of cold lysis buffer containing a proteinase inhibitor cocktail was added. Homogenization was performed using a plastic homogenization tip or peptide on ice. The samples were kept on ice for 30 min, with homogenization repeated every 10 min. Subsequently, the lysate was centrifuged at 12,000 × g for 15 min at 4°C. While waiting for centrifugation, the anti-Flag M2 affinity gel was was resuspended well, and 20 µL of the anti-Flag M2 gel was pipetted into a spin column using a cut pipette tip. The gel was washed by adding 300 µL of lysis buffer (without cocktail). The spin column was quickly spun down, and the supernatant was discarded. This washing step was repeated four more times. For the sixth washing, the spin column bottom part was plugged, and 300 µL of lysis buffer (containing the cocktail) was added. The spin column was stored on ice until needed. Just before using the gel in the spin column, the plug and spin column were removed, and the supernatant was discarded. The plug was then placed back into the spin column.

After centrifugation of the lysate, the supernatant was transferred into the spin column. Fifty µL of the lysate was reserved to measure the protein concentration and used as the input sample. The spin column with the lysate was incubated on a rotating platform for 2 h at 4°C. Subsequently, the spin column was centrifuged at 2,000 × g for 1 minute, and the flow-through was discarded (with 50 µL reserved for western blot analysis). The column was washed using 300 µL of wash buffer 1 (1X buffer G, 0.1% Triton X-100, 5% glycerol). The spin column was centrifuged at 2,000 × g for 1 minute (or left for 2 minutes to allow buffer flow), and the flow-through was discarded. This washing step was performed a total of three times. The column was then washed using 300 µL of wash buffer 2 (1X buffer G, 5% glycerol). Again, the spin column

was centrifuged at 2,000 × g for 1 minute (or left for 2 minutes), and the flow-through was discarded. This washing step was also repeated three times.

To elute the bound proteins, 6X SDS loading buffer was prepared by mixing 0.33 M Tris (pH 6.8), 34% (vol/vol) glycerol, 10% (wt/vol) SDS, 0.09% (wt/vol) DTT, 0.12% (wt/vol) bromophenol blue in MiliQ water. 50 µL of elution buffer (1X SDS loading buffer containing 2-mercaptoethanol (2-ME)) was added to the spin column. The column was boiled at 95-100°C for 5 min, ensuring that the column cap was slightly released to prevent elution buffer leakage. A quick centrifugation at 12,000 × g for 10 s was performed to collect the eluate, which served as the final sample. To verify the IP results, a western blot analysis was performed using 6 µL of eluate, input, and flow-through samples. If the sample was not used immediately, it was aliquoted into the desired volume and stored at -20°C to avoid freeze-thaw cycles. When performing IP experiments using Myc-trap (Chromotek, yta-20), the same protocol was followed without any modifications, and 20 µL of beads was sufficient for each sample.

### 2.16 Whole larvae co-IP and mass spectrometry (MS)

The required reagents described in appendix A 7 were prepared accordingly. For sample collection, two groups of samples were collected: a control group ($w^{1118}$) and an experimental line carrying a tagged version of the target protein. Late L3 (40 h) larvae were collected in 1X PBS. The larvae were left in the buffer for 2-5 min, allowing them to expel food from their gut. Subsequently, the larvae were transferred to fresh 1X PBS two more times to ensure the cleanliness of their gut. Using clean forceps, the larvae were then transferred to 1.5 mL tubes containing ice-cold PBST (1X PBS, pH 7.4, with 0.1% Tween-20).

Fresh 1X PBS with 0.2% formaldehyde was prepared as the fixative solution. The PBST was removed from the collected samples as much as possible. The dissected tissue was then incubated in fresh fixative solution at room temperature on a rotary shaker for 10 min. Subsequently, the sample was allowed to sink to the bottom of the tube, and the fixative solution was promptly removed. To quench the crosslinking reaction, 0.25 M glycine in PBST was added to the sample and incubated on a shaker for 5 min. The quench solution was then removed, and the sample was washed three times with PBST by gently inverting the tube 5 times. Before storage,

the remaining PBST was removed, and the samples were snap-frozen in liquid nitrogen and stored at -80°C until the desired number (150 larvae for each sample) of samples was obtained.

The samples were removed from the -80°C freezer and thawed on ice. Simultaneously, the Dounce homogenizer was pre-rinsed with lysis buffer. Sample homogenization was performed using a Dounce homogenizer in 1 mL of lysis buffer on ice for approximately 10 strokes. Homogenization was repeated every 15 min for about 1 h. The lysate was transferred to 1.5 mL centrifuge tubes and incubated for 15 min to ensure thorough lysis. The lysate was then centrifuged at 16,000 x g for 30 min at 4°C. The anti-Flag M2 affinity gel was resuspended well, and 40 µL of anti-Flag M2 gel was pipetted into a spin column using a cut pipette tip. The gel was washed by adding 300 µL of lysis buffer (without proteinase inhibitor), followed by a quick spin down and discarding of the supernatant. This washing step was repeated four more times. For the sixth round, the bottom part of the spin column was plugged, and 300 µL of lysis buffer (with proteinase inhibitor) was added. The spin column was stored on ice until needed. Immediately before using the gel in the spin column, the plug and spin column were removed, the supernatant was discarded, and the plug was replaced.

After the lysate centrifugation, the supernatant was transferred into the spin column, taking care not to disturb the fat layer on top. Some lysate was kept aside to measure the protein concentration. The sample was then incubated on a rotating platform for 3 h at 4°C, followed by centrifugation at 12,000 x g for 10 s. The flow-through was discarded. The column was washed three times with 300 µL of wash buffer 1 (1X buffer G, 0.1% Triton X-100, 5% glycerol) by centrifugation at 12,000 x g for 10 seconds, discarding the flow-through after each wash. The column was then washed three times with 300 µL of wash buffer 2 (1X buffer G, 5% glycerol) by centrifugation at 12,000 x g for 10 s, discarding the flow-through after each wash.

For elution, 30-40 µL of elution buffer was added to the column. The column was boiled at 95-100°C for 5 min. After a quick centrifugation at 12,000 x g, the flow-through was collected as the final sample. If the sample was not used immediately, it was aliquoted into the desired volume and stored at -20°C to avoid freeze-thaw cycles. A fresh gel was prepared prior to running the sample. The sample was loaded, and the gel was run for 1 cm on a 10% or 12.5% separating gel. Subsequently, Coomassie Blue staining was performed following the protocol outlined in the next

section. When samples were ready, they were sent to the 4096 Katz Group Centre For Pharmacy and Health Research (KGR) (Edmonton, Alberta, Canada) for analysis.

### 2.17 Coomassie blue staining

The SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) gel was carefully removed from the glass and rinsed once in Mili-Q water using a suitable container with a lid. A Coomassie Stain solution (1 g of Coomassie R250 dye, 100 mL of glacial acetic acid, 400 mL of methanol, and 500 mL of Mili-Q water) was prepared. The gel was immersed in the Coomassie Stain solution, ensuring it was covered by approximately 1/2 inch (~1.5 cm) of the solution. The gel was incubated in the Coomassie Stain solution for 1 h. After the incubation, the Coomassie Stain was poured off (it could be recycled by filtering it for future use). The container was rinsed twice with Mili-Q water or used Destain solution to remove any remaining Coomassie Stain.

For destain, the gel was submerged in the Destain solution (20% methanol, 10% acetic acid). Four Kimwipes, tied in a simple knot, were placed around the gel in the Destain solution. The gel was then incubated in the Destain solution overnight at RT for 1 h on a rocking table. The stained Kimwipes were discarded, and fresh knotted Kimwipes were added. The gel was further incubated overnight on a rocking table. The destaining process could be stopped whenever the desired level of destaining was achieved, and microwaving the container could be employed to expedite the process. Destain solution was recycled by storing it in a sealed container with Kimwipes to remove any remnants of Coomassie Stain.

### 2.18 Western blot (WB)

To make 1.5 mm-thick SDS-PAGE gels, the glass plates were carefully assembled using a casting frame and placed on the cast stand. 7 mL of separating gel solution (please see appendix A 8 for recipe, usually 10-12.5%) was added to the gap between the glass plates. Additionally, 1 mL of isopropanol was added to ensure the separating gel was flattened. The gel was allowed to set at room temperature for ~50 min, after which the isopropanol was poured off. Following this, 3 mL of stacking gel solution (see appendix A 8 for recipe) was added, and a 1.5 mm 10-well (or 15-well) comb was inserted. The gel was left to set for another 50 min.

Once the gels were ready, the Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad, Cat. 1658007FC) was assembled. The cell was filled with 1X running buffer (for 10X running buffer, add 30 g of Tris base, 144 g of glycine and 10 g of SDS to 1 L of Mili-Q water). The 1.5 mm comb was removed, and the input, flow-thru, and elution samples were loaded. A pre-stained protein ladder was used to monitor the target proteins of interest. The gel was run for 1.5 h at 110 volts (V). After electrophoresis, the proteins were transferred to a PVDF membrane. The PVDF membrane (0.2 µm) was incubated in methanol for 15 s, rinsed once in Mili-Q water, and then placed in the prechilled transfer buffer: 100 mL of 10X Tris-glycine (30.3 g of Tris base, 144 g of glycine, and 1 L of Mili-Q water), 200 mL of methanol, and 700 mL of Mili-Q water. The stacking gel was removed, and the sample loading order was labeled by cutting the top right corner of the separating gel. The stack (Bio-Rad, Mini Trans-Blot Cell) was prepared by layering 1 sponge, 2 filter papers, the gel, the PVDF membrane, and again 2 filter papers and 1 sponge on the black side. The protein transferring system was installed, and a cooling unit was inserted into the cell. The system was placed in a container with an ice-water mixture, and the proteins were transferred for 1.5 h at 100 V regardless of the number of gels.

To block the transferred proteins, the PVDF membrane was placed on a rotating platform and incubated with blocking buffer (1X TBS with 0.1% Tween-20 and 5% (w/v) skim milk powder) for 1 h at RT or overnight at 4°C. The blocking buffer was poured off, and appropriate dilutions of the primary antibody (see Table 2.5 for antibodies used in this study) were prepared in blocking buffer. The membrane was incubated with the primary antibody on a rotating platform overnight at 4°C. To conserve the primary antibody, the primary antibody dilution could be stored at -20°C and reused once more. The membrane was washed three times with washing buffer (1X TBS with 0.5% Tween-20) for 15 min each time on a rotating platform. Subsequently, the membrane was incubated with the recommended dilution of conjugated secondary antibody (generally 1:10000) in blocking buffer at room temperature for 1 hour. The membrane was washed three times with washing buffer on a rotating platform for 15-20 min each time.

For signal development, the protocol provided by the kit manufacturer (Amersham ECL Prime Western Blotting Detection Reagent) was followed. An ECL reaction solution was prepared by mixing 200 µL of solution A and 200 µL of solution B for one membrane. The mixture was placed on a piece of parafilm. Using a pair of tweezers, the PVDF membrane was lifted and gently

pressed against a paper towel to remove most of the liquid. The protein side of the membrane was then placed on the ECL mixture for 15 s. The PVDF membrane was carefully removed, and excess ECL solution was quickly blotted with a paper towel. The PVDF membrane was placed in a transparent plastic bag for signal development using a ChemiDoc imaging system (Bio-Rad).

### 2.19 CUT & Tag (Cleavage Under Targets & Tagmentation) protocol

This protocol was modified from "CUT & Tag with *Drosophila* tissues" [89, 90] and "CUTANA CUT & Tag Protocol" provided by EpiCypher (https://www.epicypher.com/resources/protocols/cutana-cut-and-tag-protocol/). For each sample, 30 ring glands or parts of the fat body from 20 larvae were used to prepare the DNA library. Before starting, required reagents (listed in appendix A 9 )were prepared.

Concanavalin A (ConA) beads were gently but thoroughly resuspended. A total of 10 µL of resuspended ConA beads were transferred to a 1.5 mL tube for each planned CUT & Tag reaction to enable batch processing. The tube was placed on a 1.5 mL magnetic separation rack until the slurry cleared, and the supernatant was carefully removed using a pipette. Immediately following, 100 µL of cold Bead Activation Buffer was added to each reaction, and the tube was removed from the magnet and gently mixed using pipetting. The tube was placed back on the magnet until the slurry cleared, and the supernatant was once again removed using a pipette. This wash step was repeated twice, resulting in a total of two washes. The ConA beads were resuspended in 22 µL of cold Bead Activation Buffer per reaction and kept on ice until needed (20 µL activated beads per reaction).

Larvae were dissected using a dissecting microscope equipped with tangential illumination. Healthy third instar larvae (40 h) were collected in a glass dish containing PBT buffer. Any yeast and residual food present on the larvae were washed off. For each genotype, 900 µL of cold PBS were aliquoted into three wells of a dissection plate. The larvae were transferred from PBT to the first well containing cold PBS. The brian-ring gland complex (BRGC) was dissected out and transferred to another well containing 900 µL of cold Wash150 buffer supplemented with the protease inhibitor cocktail. This step minimized contamination with other tissues and reduced protein degradation. Once 30 BRGCs were obtained, the ring glands were dissected and placed in 50 µL of Wash150 buffer in each tube of an 8-tube strip.

To each tube, 20 µL of activated ConA beads were added, and the solution was gently mixed by pipetting. The tubes were incubated at RT for 10 min. All subsequent buffer changes were by grabbing the beads with a magnet, pipetting off the liquid, and replacing it with new buffers. The beads were not resuspended or agitated at any point in the procedure to minimize sample loss. Primary antibody dilutions were prepared in Antibody150-1 Buffer, with 50 µL required for each sample at a typical 1:100 dilution. The sample strip was placed on the low-volume side of the magnet and allowed to bind for 1 min. The buffer was pipetted off, and the strip was removed from the magnet. Next, 50 µL of the antibody dilution was added to each sample, and the beads were gently mixed by pipetting. The samples were incubated without movement at 4°C overnight.

The second antibody dilution was prepared in Antibody150-2 Buffer, with 50 µL needed for each sample at a 1:100 dilution. The sample strip was placed on the low-volume side of the magnet and allowed to bind for 1 minute. The primary antibody solution buffer was pipetted off, and 50 µL of the second antibody dilution was added without resuspending the beads. The strip was removed from the magnet and incubated without movement at RT for 1 h.

The sample strip was placed on the low-volume side of the magnet for 1 min to capture the beads. The second antibody solution was removed, and 200 µL of cold Digitonin150 Buffer was added directly onto the beads of each reaction. The supernatant was pipetted off. This washing step was repeated twice, with the tubes and beads remaining on the magnet throughout the process. After the second wash, the supernatant was removed and discarded. Next, 50 µL of the pAG-Tn5 dilution (50 µL of Digitonin300 Buffer and 1.25 µL of pAG-Tn5 (Epicypher, Cat. # EP151017)) was added, and the PCR strip was moved right and left two times on the magnet to mix the contents. The strip was then removed from the magnet and incubated without movement at RT for 1 h. The beads were washed by placing the strip on the low-volume side of the magnet for 1 min. The tethering solution was removed, and 200 µL of cold Digitonin300 Buffer was added. The beads were not resuspended and incubated on the magnet for 2 minutes. The supernatant was pipetted off and discarded. This wash step was repeated once more, followed by pipetting to remove and discard the supernatant.

For chromatin tagmentation (which combines “DNA fragmentation” and “tag addition”), the tubes were removed from the magnet, and 50 µL of cold Tagmentation Buffer was added to each reaction. The PCR strip was gently moved right and left two times on the magnet to ensure thorough mixing. The 8-strip PCR tubes were then incubated for 1 h at 37ºC in a thermocycler. This step is crucial as it activates the Tn5 enzyme tethered to the antibody-bound chromatin, leading to the tagmentation of the target chromatin. Subsequently, the tubes were returned to the magnet, and the slurry was allowed to clear for 30 s to 2 min. The supernatant was carefully pipetted off and discarded. The tubes were removed from the magnet, and the beads were washed with 50 µL of TAPS-EDTA Buffer by gently moving the tubes right and left two times on the magnet. The tubes were then incubated for 5 min at RT. After incubation, the tubes were returned to the magnet, and the slurry was allowed to clear for 30 s to 2 min. The supernatant was pipetted off and discarded. To quench the tagmentation reaction, 5 µL of SDS Release Buffer (containing 0.1% SDS) was added to each reaction. A quick spin was performed, followed by vortexing the tubes on max speed for 10 s to ensure thorough mixing. A quick spin of the 8-strip PCR tubes in a benchtop microcentrifuge was then performed to collect the beads and buffer. The 8-strip PCR tubes were incubated for 1 h at 58ºC in a thermocycler. This step is critical as it is required to release the tagmented chromatin fragments into solution, for both fixed and unfixed nuclei. After incubation, 15 µL of SDS Quench Buffer (containing 0.67% Triton-X) was added to each reaction. A quick spin was performed, followed by brief vortexing on max speed for 10 s to ensure mixing. This step neutralizes SDS, which potently inhibits PCR.

For library enrichment, 2 µL of 10 µM Universal i5 primer and 2 µL of 10 µM uniquely barcoded i7 primer were aliquoted into each tube of a PCR tube strip, using a different barcode for each sample. The PCR strip was chilled on ice. Then, 25 µL of NEBNext HiFi 2X PCR Master Mix was added to the strip. A quick spin, vortexing, and brief spin were performed to ensure thorough mixing, and the strip was returned to ice. The thermocycler was pre-warmed to 58ºC, then the tube strip was moved into the thermocycler and start the cycling program:

Heated lid
Cycle 1: 58°C for 5 min (gap filling)
Cycle 2: 72°C for 5 min (gap filling)
Cycle 3: 98°C for 45 s

Cycle 4: 98℃ for 15 s

Cycle 5: 60℃ for 10 s

Repeat Cycles 4-5 16 times

72℃ for 1 min and hold at ℃

For library cleanup, the PCR tubes were returned to the magnet, and the slurry was allowed to clear for 30 s to 2 min. Then, 50 µL of the supernatant was carefully pipetted off and transferred to each Eppendorf tube containing 65 µL of Ampure beads. The tubes were vortexed to mix the contents and quickly spun to collect the beads. They were incubated for 5 min at RT. Subsequently, the Eppendorf tubes were placed on the magnet and incubated for 5 min to allow the beads to bind. The supernatant was pipetted off, and the beads were washed with 700 µL of 70% ethanol while still on the magnet. After a 1-min incubation, the ethanol was aspirated off, and the beads were washed again with 700 µL of 70% ethanol. The ethanol was aspirated off, leaving approximately 650 µL. The tubes were quickly spun and returned to the magnet for 2 min. The remaining ethanol was removed, and the beads were resuspended in 25 µL of 10 mM Tris-HCl, pH 8. The tubes were incubated for 5 min, followed by a quick spin. Then, the tubes were returned to the magnet stand and kept for 2 min. The DNA library solution was carefully collected into a new tube and stored at -20°C until further use.

The barcoded libraries were mixed to achieve equal representation as desired, aiming for a final concentration recommended by the sequencer manufacturer. After mixing, an Ampure bead cleanup was performed if needed to remove any residual PCR primers. The barcoded libraries were subjected to paired-end Illumina sequencing following the manufacturer's instructions. PE25 sequencing was considered sufficient for mapping to the *Drosophila* genome. Mapping, spike-in-free normalization, and analysis were described in the next section.

### 2.20 CUT & Tag data analysis

The data analysis was done by following "CUT & Tag Data Processing and Analysis Tutorial using Linux and R commands [91]. Modifications were applied and described as follow. The quality of sequencing data was checked using FastQC (version 0.11.9). Adapter sequences were removed using Trimmomatic (version 0.39), trimmed and paired-end reads were then subjected for alignment to BDGP6 genome via Bowtie2 (version 2.5.1). Duplicate reads were

removed by Picard (version 2.18.29), and normalization was done by ChIPseqSpikeInFree (version 1.2.4) before peak calling via SEACR (version 1.4) [92, 93] . The “top 0.01 peaks” and “control peaks” were called using the stringent mode. Based on the signal intensity, the top 2% of “control peaks” from PG samples and top 50% of “top 0.01 peaks” from fat body samples were picked, followed by peak annotation via Homer (version 4.11). All data visualization figures were generated using R commands. Please see appendix A 10 for detailed codes and procedure.

### 2.21 TurboID protocol

Required reagents were prepared following the recipes in appendix A 13. For each sample, the head of 200 larvae with PG-specific overexpression of TurboID-IRP1A or TurboID-IRP1B were collected. Tissues were lysed in 1-2 mL of lysis buffer with pestle homogenization every 10 min for 1 h. Cell lysate was centrifuged at 4°C, 13,000 rpm for 30 min. Supernatant was kept and centrifuged again for 15 min. The supernatant was transferred to a new 1.5 mL Eppendorf tube and the concentration of protein in each sample was measured using the BCA protein assay. For each sample, 25 µL of streptavidin magnetic beads were used and washed twice with 1 mL of RIPA lysis buffer. The beads were then incubated with 300 µg of protein from each sample, along with an additional 500 µL of RIPA lysis buffer, at 4°C for 2-3 h with rotation. The remaining material from the whole-cell lysates was saved as input for WB analysis and stored at -20°C.

After the enrichment step, the beads were pelleted using a magnetic rack, and the supernatant was collected in fresh microcentrifuge tubes as flow-through for WB analysis. The beads were washed twice with 1 mL of RIPA lysis buffer at RT for 2 min, followed by a wash with 1 mL of 1 M KCl at RT for 2 min, a wash with 1 mL of 0.1 M $Na_2CO_3$ for approximately 10 s, a wash with 1 mL of 2 M urea in 10 mM Tris-HCl (pH 8.0) for ~10 s, and finally, two washes with 1 mL of RIPA lysis buffer at RT for 2 min each. After the final wash, the beads were transferred to fresh tubes using 1 mL of RIPA lysis buffer, and the buffer was removed after beads were pelleted using a magnetic rack.

To elute the enriched material from the beads, each sample was boiled in 60 µL of 3× protein loading buffer supplemented with 2 mM biotin and 0.6 µL of 2-ME at 95 °C for 10 minutes. The beads were then pelleted using a magnetic rack, and the eluate was collected. For WB analysis, the input samples and the flow-through samples were mixed with 6× protein loading buffer. The

samples were boiled at 95 °C for 10 min. The input, flow-through, and eluate samples were loaded onto a 12.5% (vol/vol) SDS-PAGE gel for separation and then transferred to a PVDF membrane. The membranes were blocked with 5% (wt/vol) nonfat dry milk in 1× TBST and gently rocked at RT for 1 h, and subsequently washed three times with 1× TBST for 5 min each to remove any residual blocking agent and unbound proteins.

For the detection of biotinylated proteins, the membranes were incubated with 0.3 μg/mL streptavidin-HRP in 3% (wt/vol) BSA in 1× TBST at RT for 30 min. It is worth noting that nonfat dry milk was excluded during this incubation step because it contains free biotin that could potentially interfere with the binding of streptavidin to biotinylated proteins adsorbed on the membrane. For the detection of fusion constructs expression, the membranes were incubated with mouse anti-V5 (1:10,000) in 3% (wt/vol) BSA in 1× TBST at RT overnight at 4 °C. After the primary antibody incubation, the membranes were washed three times with 1× TBST for 5 min each. Subsequently, the membranes were incubated with a secondary antibody in 3% (wt/vol) BSA in 1× TBST at RT for 30 min. The rest of elution samples were subjected for MS analysis in Alberta Proteomic and Mass Spectrometry Facility.

### 2.22 MS data analysis

SDS-PAGE gel was stained by Coomassie blue and then submitted to Alberta Proteomic and Mass Spectrometry Facility (https://apm.biochem.ualberta.ca/index.html) for MALDI-TOF mass spectrometry analysis. Peptides were identified using SEQUEST search engine with *Drosophila* proteome database from Uniport as reference. Raw data were sent to us after being converted to excel files with features such as Uniport protein identification numbers, scores, PSMs (peptide spectral matches), and so on. To uncover enriched interacting proteins, statistical analysis like two-sample *t*-test was applied using R (see appendix A 13 for detailed codes). Proteins were considered significantly enriched if they had a p-value $\leq 0.05$ and a score higher than 2-fold of that in the control. Enriched target lists were submitted for subcellular localization identification via the FlyBase Batch Download tool (http://flybase.org/batchdownload). The overlaps between treatment groups were visualized by a Venn diagram generated by R.

**Table 2.1 List of fly lines used in this study**

| Line | Source | Purpose |
|---|---|---|
| *Act-Gal4* | BDSC, # 42713 | Whole-body Gal4 driver |
| *Phm-Gal4* | A kind gift from Michael O'Connor's lab | PG-specific Gal4 driver |
| *Lsp2-Gal4* | BDSC, # 6357 | Fat body-specific Gal4 driver |
| $w^{1118}$ | BDSC, # 3605 | As a control in CUT & Tag experiment |
| *UAS-3F-IRP1A*$^{WT}$ | KKJ lab-Nhan Huynh | For CUT & Tag experiment |
| *UAS-3F-IRP1A*$^{C450S}$ | KKJ lab-Nhan Huynh | Provide *IRP1A*$^{C450S}$ cDNA for *TurboID-IRP1A*$^{C450S}$ construction |
| *TurboID-IRP1A* | KKJ lab-Minyi Yan | For TurboID experiment |
| *TurboID-IRP1B* | KKJ lab-Minyi Yan | For TurboID experiment |
| *TurboID-V5-EGFP* | KKJ lab-Wen Liu | As a control in TurboID experiment |
| *TurboID-IRP1A*$^{C450S}$ | KKJ lab-Minyi Yan | For future TurboID experiment |
| *IRP1A*$^{WT.3F}$ | KKJ lab-Minyi Yan | For future survival rate assay |
| *IRP1A*$^{C450S.3F}$ | KKJ lab-Minyi Yan | For future survival rate assay |
| *IRP1A*$^{3R3Q.3F}$ | KKJ lab-Minyi Yan | For future survival rate assay |
| *IRP1A*$^{KO}$ | KKJ lab-Minyi Yan | For future survival rate assay |

**Table 2.2 List of commercial plasmids**

| Plasmid name | Company | Catalogue No. | Purpose |
|---|---|---|---|
| *pAFW* | Addgene | Not available | S2 cell transfection |
| *pAMW* | Addgene | Not available | S2 cell transfection |
| *pAc5-STABLE2-neo* | Addgene | 32426 | S2 cell transfection |
| *pBID-UASC-FG* | Addgene | 35201 | UAS overexpression line construction |
| *pUAS-V5-TurboID-NES* | Addgene | 116904 | TurboID line construction |
| *pCFD5* | Addgene | 73914 | CRISPR line construction |
| *pDsRed-attP* | Addgene | 51019 | CRISPR line construction |

**Table 2.3 List of plasmids generated in this study**

| Plasmid name | Features |
|---|---|
| For S2 cell transfection | |
| *pAFW C-3xFlag* | C-terminal 3X Flag-tagged backbone plasmid |
| *pAFW IPR1A-C-3xFlag* | Expresses a C-terminal 3X Flag-tagged IRP1A in S2 cells |
| *pAMW H2Av-C-6xMyc* | Expresses a C-terminal 6X Myc-tagged H2Av in S2 cells |
| *pAMW N-6xMyc-EGFP* | Expresses an N-terminal 6X Myc-tagged EGFP in S2 cells |
| *pSTABLE2 IRP1A-C 3xFlag-H2Av-C-6xMyc* | Co-expresses a C-terminal 3X Flag-tagged IRP1A and a C-terminal 6X Myc-tagged H2Av, separating by T2A sequence in S2 cells |
| *pSTABLE2-IRP1A-C 3xFlag-N-6xMyc-EGFP* | Co-expresses a C-terminal 3X Flag-tagged IRP1A and an N-terminal 6X Myc-tagged EGFP, separating by T2A sequence in S2 cells |
| For UAS overexpression line construction | |
| *pAFW-IRP1A$^{2R2Q}$* | Introduces 2R2Q (R549Q and R554Q) point mutations into IRP1A |
| *pAFW-IRP1A$^{3R3Q}$* | Provides *IRP1A$^{3R3Q}$* cDNA for overexpression line establishment |
| *pBID-UASC-FG-IRP1A$^{3R3Q}$* | To establish an *IRP1A$^{3R3Q}$* overexpression line |
| *pAFW-IRP1B$^{2R2Q}$* | Introduces 2R2Q (R546Q and R551Q) point mutations into IRP1B |
| *pAFW-IRP1B$^{3R3Q}$* | Provides *IRP1B$^{3R3Q}$* cDNA for overexpression line establishment |
| *pBID-UASC-FG-IRP1B$^{3R3Q}$* | To establish an IRP1B$^{3R3Q}$ overexpression line |
| For TurboID line construction | |

| | |
|---|---|
| *pTurboID-IRP1A* | To establish a V5-tagged TurboID-IRP1A fusion protein line |
| *pTurboID-IRP1A$^{C450S}$* | To establish a V5-tagged TurboID-IRP1A$^{C450S}$ fusion protein line |
| *pTurboID-IRP1B* | To establish a V5-tagged TurboID-IRP1B |
| For CRISPR line construction | |
| *pCFD5-IRP1A-KI* | Expresses two gRNAs for IRP1A CRISPR knock-in line establishment |
| *pDsRed-left arm-IA* | Provides a backbone containing homologous left arm and target sequence of IRP1A for donor plasmid construction |
| *pDsRed-IRP1A$^{WT}$-3F* | Expresses a 3X Flag-tagged wild-type IRP1A donor DNA for CRISPR knock-in line establishment, with DsRed as a screening marker |
| *pDsRed-IRP1A$^{C450S}$-3F* | Expresses a 3X Flag-tagged IRP1A$^{C450S}$ donor DNA containing C450S point mutation for CRISPR knock-in line establishment, with DsRed as a screening marker |
| *pDsRed-IRP1A$^{3R3Q}$-3F* | Expresses a 3X Flag-tagged IRP1A$^{3R3Q}$ donor DNA containing 3-point mutations (R549Q, R554Q, R793Q) for CRISPR knock-in line establishment, with DsRed as a screening marker |
| *pCFD5-IRP1A-KO* | Expresses two gRNAs for IRP1A CRISPR knock-out line establishment |
| *pDsRed-IRP1A-KO* | Expresses an IRP1A$^{KO}$ donor DNA containing a DsRed marker flanked by two homologous arms for CRISPR knock-out line establishment |

**Table 2.4 List of fly lines generated in this study**

| Fly line | Description |
|---|---|
| *UAS-3F-IRP1A$^{3R3Q}$* | Expresses an N-terminal 3xFlag-tagged IRP1A$^{3R3Q}$ containing 3-point mutations (R549Q, R554Q, R793Q) under the control of UAS-Gal4 system |
| *UAS-3F-IRP1B$^{3R3Q}$* | Expresses an N-terminal 3xFlag-tagged IRP1B$^{3R3Q}$ containing 3-point mutations (R546Q, R551Q, R790Q) under the control of UAS-Gal4 system |
| *TurboID-IRP1A$^{WT}$* | Expresses an N-terminal V5-tagged TurboID-IRP1A$^{WT}$ fusion protein under the control of UAS-Gal4 system |
| *TurboID-IRP1A$^{C450S}$* | Expresses an N-terminal V5-tagged TurboID-IRP1A$^{C450S}$ fusion protein containing C450S point mutation under the control of UAS-Gal4 system |
| *TurboID-IRP1B$^{WT}$* | Expresses N-terminal V5-tagged TurboID-IRP1B$^{WT}$ fusion protein line under the control of UAS-Gal4 system |
| *IRP1A$^{WT.3F}$* | Expresses a CRISPR knock-in wild-type IRP1A with a C-terminal 3xFlag tag |
| *IRP1A$^{C450S.3F}$* | Expresses a CRISPR knock-in IRP1A$^{C450S}$ containing a C450S point mutation with a C-terminal 3xFlag tag |
| *IRP1A$^{3R3Q.3F}$* | Expresses a CRISPR knock-in IRP1A$^{3R3Q}$ containing 3-point mutations (R549Q, R554Q, R793Q) with a C-terminal 3xFlag tag |
| *IRP1A$^{KO}$* | IRP1A knock-out by CRISPR |

**Table 2.5 List of antibodies used in this study**

| Antibody | Source | Catalogue No. | Experiments | Dilution ratio |
|---|---|---|---|---|
| monoclonal mouse anti-FLAG | Cell Signaling | 8146S | IF | 1:1000 |
| | | | IP | 1:50 |
| | | | WB | 1:1000 |
| monoclonal rabbit anti-FLAG | Cell Signaling | 14793S | IF | 1:1000 |
| | | | IP | 1:50 |
| | | | WB | 1:1000 |
| monoclonal mouse anti-Myc | Cell Signaling | 2276S | IF | 1:1000 |
| | | | IP | 1:50 |
| | | | WB | 1:1000 |
| monoclonal rabbit anti-Myc | Cell Signaling | 2278S | IF | 1:1000 |
| | | | IP | 1:50 |
| | | | WB | 1:1000 |
| DAPI | Cell Signaling | 4083 | IF | 1:50000 |
| goat anti-mouse IgG H&L Alexa Fluor 488 | Abcam | 150113 | IF | 1:2000 |
| goat anti-mouse IgG H&L Alexa Fluor 555 | Abcam | 150114 | IF | 1:2000 |
| goat anti-rabbit IgG H&L Alexa Fluor 488 | Abcam | 150077 | IF | 1:2000 |
| goat anti-rabbit IgG H&L Alexa Fluor 555 | Abcam | 150078 | IF | 1:2000 |
| goat anti-mouse IgG H&L HRP | Abcam | 97023 | WB | 1:20000 |
| goat anti-rabbit IgG H&L HRP | Abcam | 97051 | WB | 1:20000 |
| monoclonal mouse anti-V5 | Thermo Fisher Scientific | R96025 | IF | 1:500 |
| | | | WB | 1:10000 |
| Streptavidin-HRP | Thermo Fisher Scientific | S911 | WB | 0.3 μg/mL |
| IF: immunofluorescence; IP: immunoprecipitation; WB: western blot. | | | | |

# Chapter 3 IRP1A appears to interact with distinct subsets of genes in the prothoracic gland and the fat body

## 3.1 Introduction

### 3.1.1 Transgenic fly IRPs downregulate iron-related genes

Given the unexpected nuclear localization of fly IRP1s in the PG and fat bodies [19], one would expect that IRPs have hitherto undocumented roles in the nucleus. To explore this, a previous study from our lab examined the transcriptome of IRPs and their variants [19]. Transgenic lines were used for PG-specific overexpression of IRP1A/B and their mutants (namely $IRP1A^{C450S}$ and $IRP1A^{3R3Q}$, $IRP1B^{C447S}$ and $IRP1B^{3R3Q}$), followed by RNA-Seq analysis on dissected ring glands (which harbor the PGs). As previously mentioned, wild-type IRP1A and IRP1B dynamically transitions between the apo- and holo-form as an indicator of intracellular iron concentrations [6], predominantly residing in the nuclei of PG cells (Figure 1.7). On the other hand, predicted apo-variants ($IRP1A^{C450S}$ and $IRP1B^{C447S}$) and the non-RNA-binding variants ($IRP1A^{3R3Q}$ and $IRP1B^{3R3Q}$) primarily localize to the cytosol and nucleus, respectively. Correspondingly, the transcript profiling indicated that overexpression of $IRP1A^{3R3Q}$ resulted in a downregulation of iron- and ecdysone-related genes [19]. Overexpression of wild-type IRP1A exhibited the same trend, albeit with less significance. Conversely, overexpression of $IRP1A^{C450S}$ barely had impact on the transcriptome, consistent with its predicted cytoplasmic role as an RNA-binding protein. A similar behavioral pattern was observed for IRP1B and its cDNA variants, with the overexpression of holo-variant ($IRP1B^{3R3Q}$) and apo-variant ($IRP1B^{C447S}$) exerting the most and least effects on gene transcription, respectively, while the wild-type IRP1B fell in between.

### 3.1.2 IRP1A specifically interacts with histones

The remarkable discovery of nuclear IRP1A/B affecting the transcription of iron-related genes sparked curiosity about the mechanisms by which IRPs accomplishes this task [19]. Considering their nuclear localization, the primary association that comes to mind is their potential role in gene expression. There are several possibilities to consider: a) fly IRPs might act as transcription factors (TFs), by directly binding to DNA [94], b) they could potentially serve as undiscovered co-factors that interact with other TFs or activators, thereby modulating gene expression [95], or c) they may bind to chromatin remodelers to influence gene accessibility and thus initiate or repress gene transcription [96]. Because the nuclear localization of IRPs favors the holo-form, which possesses aconitase activity, it is plausible that d) fly IRPs may function as the

aconitase, participating in the conversion of citrate to succinate. This metabolic pathway produces intermediates like α-ketoglutarate (α-KG), which are implicated in DNA and histone demethylation [97, 98]. Considering the aconitase function and RNA-binding activity of IRP1A are mutually exclusive [6], it is unlikely that IRP1A interacts with RNA directly in the nucleus. Nonetheless, since alternative splicing occurs within the nucleus, it is possible that IRPs e) impact the transcriptome through interactions with the spliceosome [99].

The protein-protein interaction map of IRPs acquired by traditional mass spectrometry provides supporting evidence for the hypothesis that IRP1A may participate in gene regulation through binding to nuclear proteins [19]. In this study from our lab, knock-in tagged IRP1A was reported to interact with a specific subset of histones, including histone H2A, H2B, and H4, and a histone variant H2Av (histone H2A variant) (Figure 3.1). Although mass proteins such as histones are common contaminants in MS [100], these interactions are considered specific by me for several reasons. Firstly, even though histone H2A, H2B, and H4 are abundant proteins in the nucleus due to their essential roles in nucleosome assembly, other nucleosome subunits (e.g., histone H1 and H3) did not show up in the interaction map. Additionally, when knock-in IRP1B was used as bait for MS, only H2A and H4 were identified as interacting proteins (Figure 3.1). Another compelling argument is that, unlike canonical histones, which have more than 20 gene copies in flies, H2Av has only one copy of the gene. As a result, the cellular abundance of H2Av is comparatively low, accounting for only 5-10% of the canonical H2A [101].

Histones are the major components of nucleosome core particles, which consist of eight core histones and are wrapped by 146 bp of DNA [102]. During nucleosome assembly, two H3-H4 dimers first form a symmetrical tetramer, followed by the tethering of two H2A-H2B dimers to each H3-H4 dimer, resulting in a histone octamer. The incredibly long human DNA (spanning over one meter per haploid set) is highly condensed and packed into the compact nucleus by wrapping around histone octamers, and then folding into higher-order structures with the assistance of linker histone H1 and scaffold proteins [102-104]. Efficient gene expression necessitates the access of DNA to transcription factors, which is facilitated by an open chromatin state. Conversely, condensed DNA inhibits gene expression. Therefore, it is possible that IRP1A affects gene expression through modulating DNA accessibility via nucleosome positioning [105]. Alternatively, IRP1A may be recruited to specific modifications found on histone proteins, and

thus form an iron-dependent chromatin reader protein important for the recruitment or displacement of other transcription factors.

Apart from major histones, histone variants also play a crucial role in regulating chromatin dynamics [106]. For example, the *Drosophila* histone H2A variant possesses both human H2A.Z and H2A.X properties, which engages in gene regulation and genome stability, respectively. When two canonical H2A are replaced by H2Av, nucleosomes favorably localize to actively transcribed genes [101]. On the other hand, the recruitment of H2Av is required for heterochromatin establishment via HP1 (heterochromatin protein 1) [107]. If IRP1A is able to influence H2Av replacement or recruitment, either directly or indirectly, it can affect downstream gene transcription.

Taken together, considering these assumptions and evidence, it appears likely that holo-IRP1A acts as a chromatin-binding protein. Therefore, I intended to explore its chromatin profile using the CUT & Tag (Cleavage Under Targets & Tagmentation) technique [108].

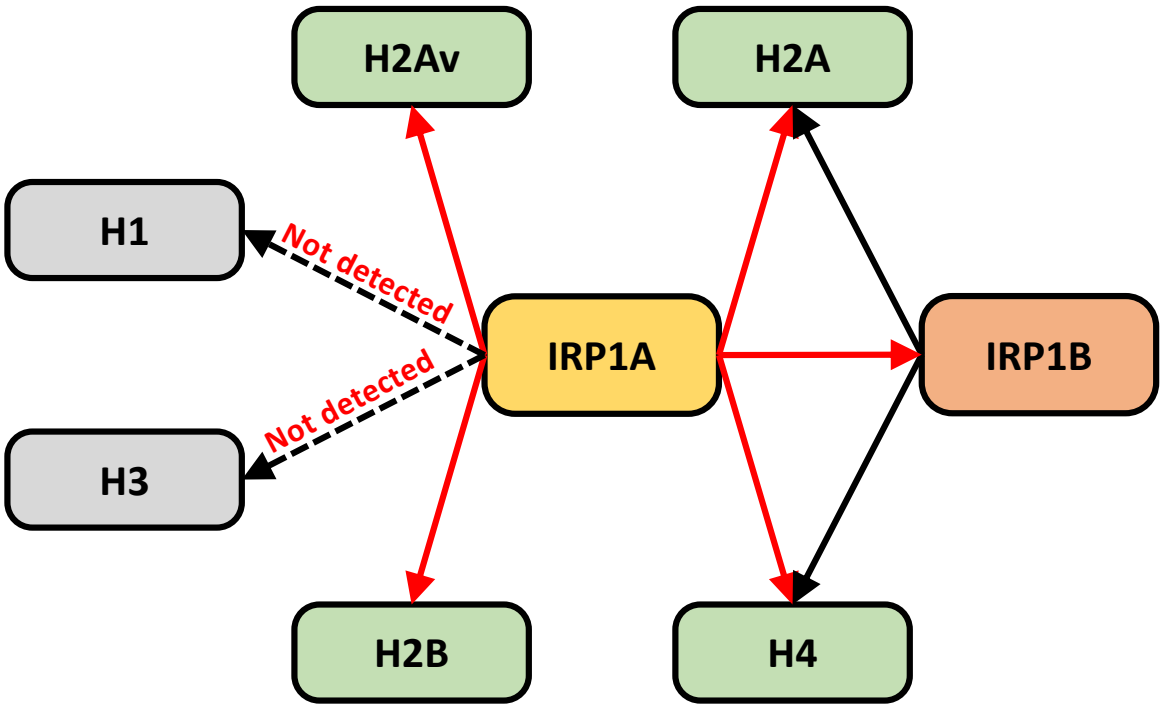



**Figure 3.1 IRPs interaction map obtained by traditional MS**

Core histones, except for histone H1 and H3 (black dashed line), were detected when IRP1A was used as a bait (red line). On the other hand, IRP1B only pulled down histone H2A and H4 (black line). Mass spec was performed on dissected PGs using knock-in Flag-tagged IRP1A or IRP1B as a bait [19].

### 3.1.3 The CUT & Tag approach

ChIP-Seq (chromatin immunoprecipitation followed by sequencing) is a commonly used method to detect DNA-protein interactions and map chromatin factors through the genome. The fundamental concept underlying ChIP-Seq is that cells are first treated with formaldehyde to fix the DNA-protein interaction (Figure 3.2). Subsequently, the chromatin is sheared by sonication and incubated with specific antibodies that recognize the TFs in question. Through immunoprecipitation, the DNA fragments bound by transcription factors are pulled down, ligated with adapters followed by PCR amplification, and then subjected to next-generation sequencing (NGS).

With the rapid development of technology, several new techniques have emerged from the original ChIP-Seq procedure. These may offer substantial improvements, such as an enhanced resolution, a higher signal-to-noise ratio, as well as a lower amount of input material [109]. Among these techniques, the latest one is CUT & Tag, which profiles epigenomics modifications in situ effectively with the help of Tn5 transposase. Similar to ChIP-Seq, specific antibodies are employed to capture DNA-protein complexes in CUT & Tag, allowing for protein-bound DNA fragments to be enriched and sequenced [109] (Figure 3.2). However, instead of crosslinking and shearing chromatin, cells are initially permeabilized and immobilized by ConA beads. Primary antibodies are added to specifically recognize TFs of interest, and then a secondary antibody is introduced to enhance the signals for improved recognition by pA-Tn5 (protein A-Tn5). Tn5 is a transposase derived from *E. coli (Escherichia. coli)*. The fusion with protein A enables Tn5 to be tethered to the secondary antibody. With the addition of magnesium and incubation at 37°C, Tn5 is activated to induce double-stranded breaks in the DNA and simultaneously insert adapters to both ends. The DNA fragments are subsequently extracted, amplified, and sequenced.

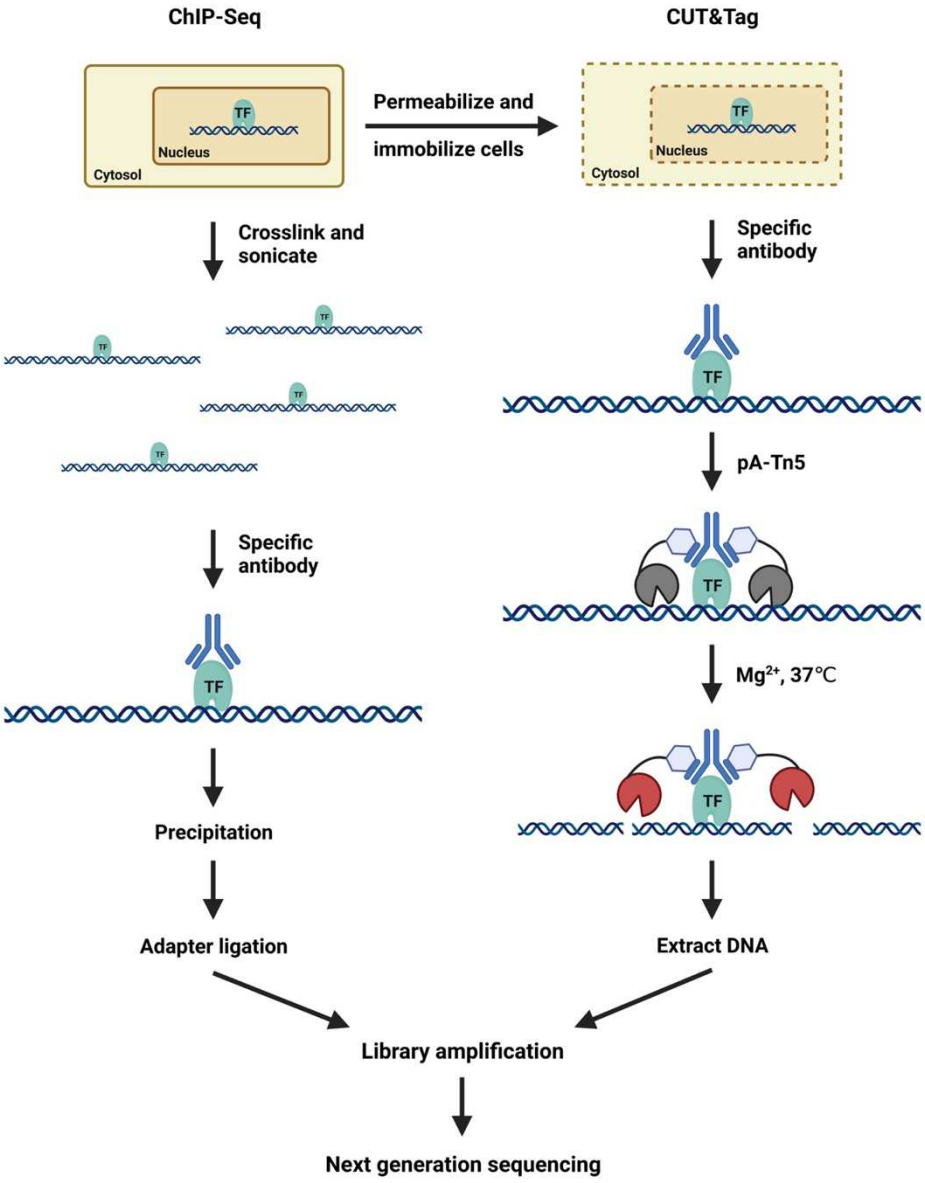



**Figure 3.2 Comparison of ChIP-Seq and CUT & Tag**

In ChIP-Seq, cells are fixed, and the chromatin is sheared to smaller size for downstream antibody incubation and immunoprecipitation. After ligating adapters, DNA fragments are amplified and sequenced. In contrast, cells are first permeabilized and immobilized, followed by antibody incubation *in situ*. Subsequently, pA-Tn5 is added to target DNA fragments bound by TFs via recognizing antibodies. With the addition of magnesium at 37℃, Tn5 is activated to cut and load adapter sequences to DNAs, which is then ready for amplification and sequencing [108, 109].

### 3.1.4 Advantages of CUT & Tag

Compared to traditional ChIP-Seq, CUT & Tag offers several important advantages. For example, one notable limitation of ChIP-Seq is the requirement for a large number of input samples to achieve reliable results, typically involving at least 10,000 cells for chromatin preparation [110]. In contrast, as few as 60 cells are sufficient to perform CUT & Tag successfully [108, 111]. This feature of CUT & Tag makes it possible to obtain chromatin profiles from cell types or tissues that are only available in limited quantities, such as PG cells. The PG is a tiny tissue near the *Drosophila* brain. It is quite time-consuming and laborious to dissect PGs in large quantities. The low input requirement of CUT & Tag allows researchers to gather adequate PG tissues at one time, thereby eliminating the need for freeze-thaw cycles during sample collection and preserving native DNA-protein interactions to a greater extent. Additionally, the background noise in CUT & Tag is exceedingly low owing to the selective binding of Tn5 to TFs. Despite delivering fewer reads, CUT & Tag can provide a higher yield due to its improved signal-to-noise ratio [109].

In contrast to ChIP-Seq, CUT & Tag eliminates the crosslink step and avoids harsh conditions like chromatin fragmentation (Figure 3.2). Consequently, more native epitopes can be exposed and captured via CUT & Tag. Moreover, the intrinsic properties of pA-Tn5 brings a range of advantages to the CUT & Tag approach. For instance, Tn5 is a highly efficient enzyme that can cut and tag target fragments at the same time, which significantly enhances tagmentation efficiency and saves the trouble to ligate adapters separately. In addition, since Tn5 remains tethered to chromatin during incubation, CUT & Tag is preferable for single-cell profiling [108, 109]. Based on these strengths, I employed CUT & Tag to profile the binding patterns of IRP1A throughout the genome.

## 3.2 Results

### 3.2.1 Data quality assessment

CUT & Tag was conducted on IRP1A transgenic tagged line (*UAS-IRP1A-3xFlag*). This line was crossed with tissue-specific (*phm-Gal4*) or whole-body driver (*Act-Gal4*) to express IRP1A in the PG and the fat body, respectively. The $w^{1118}$ strain crossed with either of the two drivers served as corresponding controls. Late-third-instar larvae (40-44 hours old) were dissected

to obtain ring gland (containing PGs) and fat body samples for the experiment. Each replicate was based on 30 ring glands, whereas fat bodies were dissected from 20 larvae to generate DNA libraries. I used two replicates for each condition.

#### 3.2.1.1 Sequencing depths

Following sequencing, the quality of the reads was assessed using FastQC [112]. The FastQC report revealed the presence of adapter sequences, which were subsequently removed using Trimmomatic [113]. The success of the trimming process was confirmed by running FastQC again. The resulting trimmed and paired-end reads from both R1 and R2 were then aligned to the *Drosophila melanogaster* genome BDGP6 using Bowtie2 [114]. The sequencing depth varied among the samples, with the PG samples from the *IRP1A* and $w^{1118}$ lines ranging from 2.9 to 5.6 million reads, while the fat body samples exhibited a higher sequencing depth of 5-13 million reads (Figure 3.3 A) (Table 3.1). This large discrepancy in the sequencing depth between two groups can be attributed to the large difference in cell numbers of two tissues. Compared to the fat body, which spreads throughout the larva's body, the PG is a tiny tissue with much fewer cells, and thus produced a significantly lower number of reads. For *Drosophila*, it is generally recommended to have at least two million reads mapped to the genome [115, 116]. In each replicate, over 75% of the reads successfully aligned to the genome, resulting in more than two million reads eligible for downstream analysis (Figure 3.3 B and C).

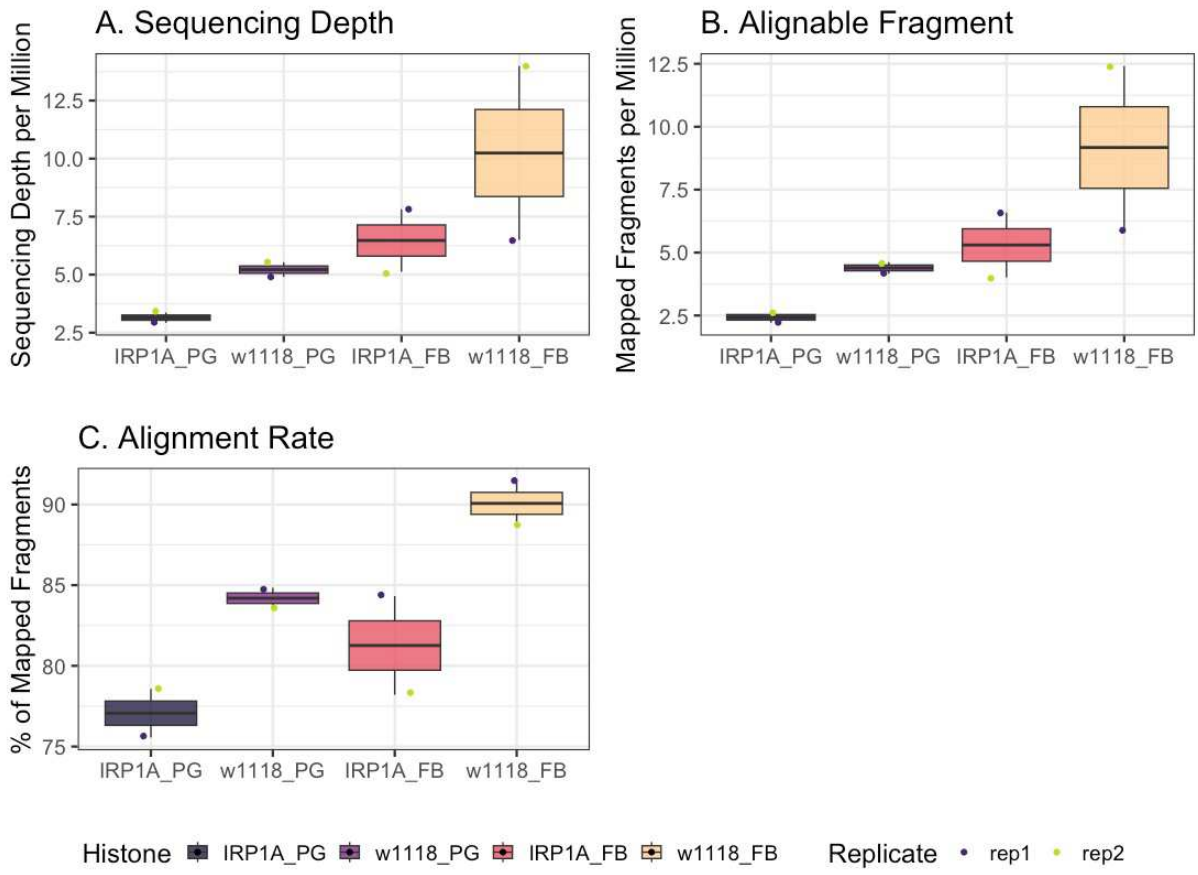



**Figure 3.3 Alignment result**

(A) Sequencing depths of PG and fat body samples. The number of reads obtained from PG samples ranged from 2.9-5.5 million, while fat body samples yielded almost double the reads. (B) Number of mapped fragments. More than 2.5 million reads from PG samples, and 5 million from fat body samples were successfully mapped to the fly genome. (C) Alignment rates. Most of the reads (>75%) in each replicate were alignable. PG: prothoracic gland; FB: fat body; rep: replicate.

#### 3.2.1.2 Removing duplicate reads

Prior to peak calling, it is crucial to remove duplicate reads to prevent their inflation of peak signals and potential masking of true peaks with lower signals, which leads to biases in the results. Duplicate reads are typically generated during the PCR amplification step before sequencing, particularly when there are excessive PCR cycles or insufficient input material. Therefore, removing duplicates is essential in this case to ensure accurate peak calling. However, in high-throughput and genome-wide experiments such as ChIP-Seq and CUT & Tag, instead of providing precise binding loci of transcription factors, they can only offer approximate information or possibilities regarding binding sites. Due to the cutting preference of pA-Tn5, which tends to cut and integrate adapters into easily accessible DNAs, it is possible for the DNA library to contain duplicate reads that share the exact same starting and ending points [91, 108]. In this scenario, these duplicate fragments may represent true fragments and should not be discarded.

To evaluate the extent of duplication in the libraries, duplicate reads were marked, and the duplication rate was calculated using Picard [117]. In each sample, the duplication rate exceeded 60% or even 80% (Figure 3.4 A) (Table 3.1). According to the ChIP-Seq guidelines provided by the ENCODE consortia, the acceptable duplication rate should be below 20% [115]. Due to the low input cell numbers and the fact that IRP1A is not a canonical transcription factor, which is not expected to have a large number of targets in the chromatin, the high duplication numbers observed in my data are more likely a result of technical issues such as PCR over-amplification rather than the Tn5 cutting preference. Hence, duplicates were removed before conducting downstream analysis. Based on the calculation via Picard, the estimated library size and the number of uniquely mapped fragments were around 0.5 million in both PG samples, and ranged from 2 to 4 million in fat body samples (Figure 3.4 B and C) (Table 3.1).

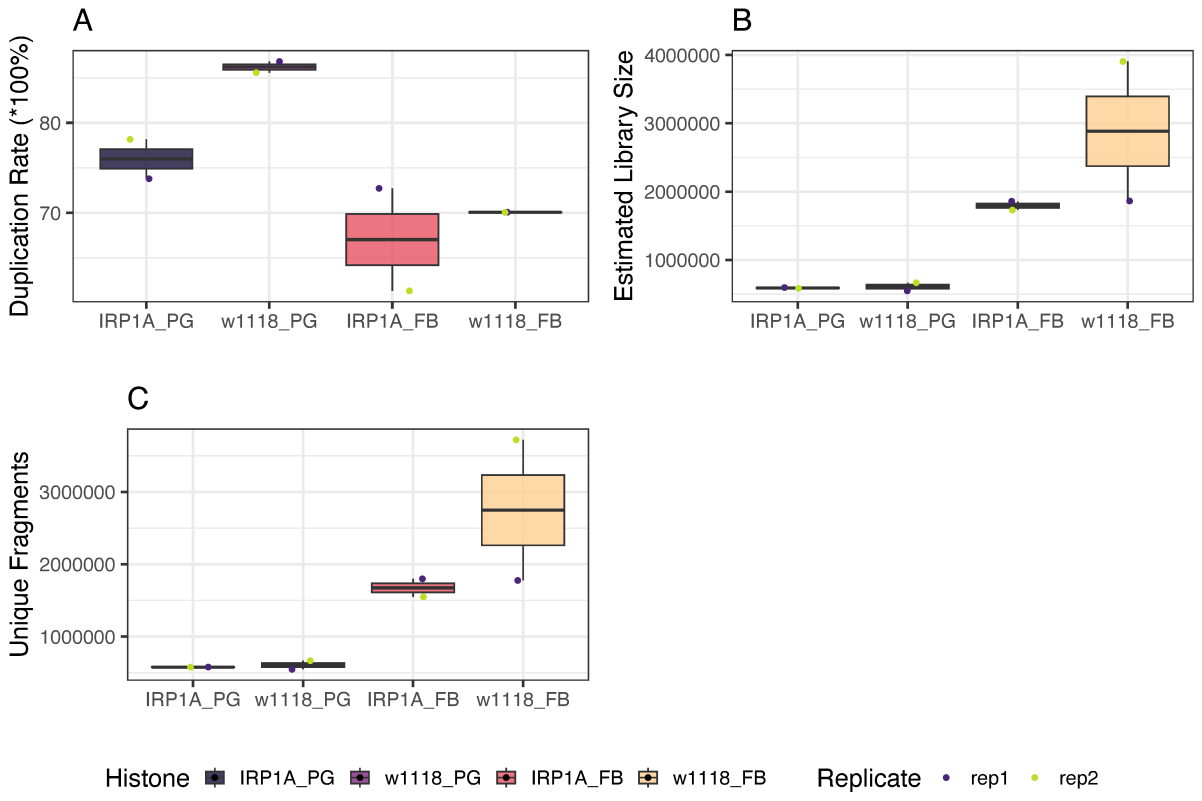



**Figure 3.4 Duplication rates**

(A) Duplication rates. More than 60% of the reads were duplicated in each replicate, which should be discarded. After duplicate removing, (B) The estimated library size and (C) The number of unique fragments dropped below 1 million in PG samples and to 2-4 million in fat body samples. PG: prothoracic gland; FB: fat body; rep: replicate.

#### 3.2.1.3 Assessing fragment sizes

For histone modifications, Tn5 is tethered to nucleosome and induces double-stranded breaks in surrounding DNAs, producing nucleosomal size of fragments (~180, including the 147 bp core DNA and 20-60 bp linker DNA), or multiples of that size [91, 108, 118]. On the other hand, CUT & Tag targeting transcription factors generates variable amounts of shorter fragments. Because core Tagmentation can also occur on DNA wrapped by nucleosomes. The fragment sizes of each replicate were depicted in the left panel of Figure 3.5, indicating consistent sizes among different samples, except for replicate 1 in the $w^{1118}$ fat body sample. The majority of fragments were approximately 110 bp in length, with the remaining fragments varying from 110 to 300 bp. Only the $w^{1118}$_FB_rep1 sample showed a different distribution, with most fragments exhibiting peak lengths of ~180 bp and a smaller peak at 110 bp. A typical hallmark of successful CUT & Tag experiments is the presence of a 10-bp repeating sawtooth pattern in the plot of fragment lengths at single-base-pair resolution [91]. Every replicate (except for $w^{1118}$_FB_rep1) exhibited a repeating 10-bp serration pattern, indicating the success of the CUT & Tag procedure (line chart in the right panel of Figure 3.5). In this regard, $w^{1118}$_FB_rep1 may be an outlier.

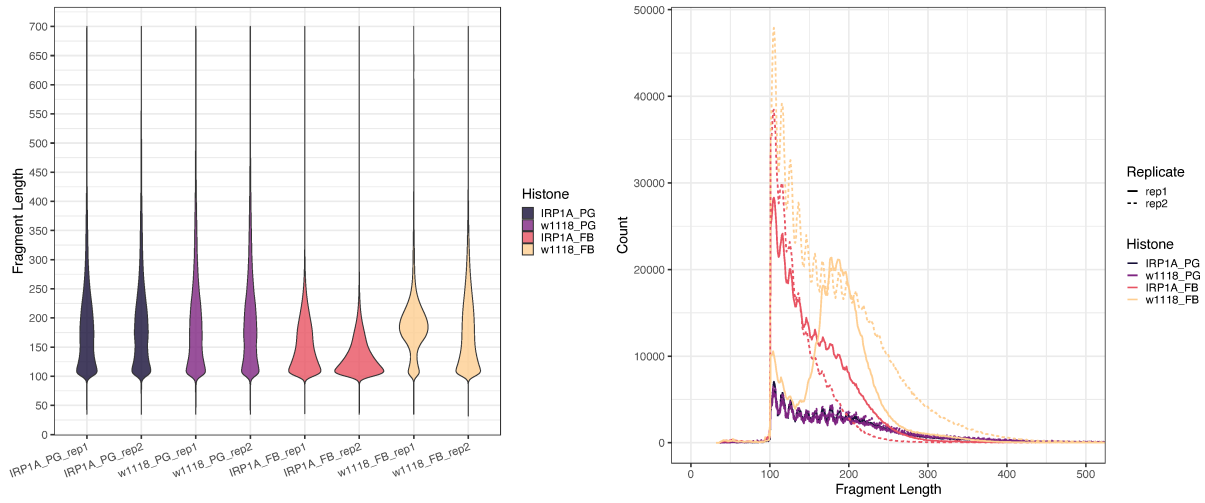



**Figure 3.5 Distribution of fragment lengths**

All replicates, except for $w^{1118}$_FB_rep1, displayed a consistent distribution of fragment lengths ranging from 100 to 300 bp (left panel), and they exhibited a repeating serration pattern with 10-bp intervals (right panel).

**Table 3.1 Alignment summary**

| Sample | Rep | Seq Depth | Mapped Frag Num | Alignment Rate | Dup Rate | Est Library Size | Unique Frag Num | SF |
|---|---|---|---|---|---|---|---|---|
| **IRP1A_PG** | rep1 | 2924791 | 2209908 | 75.56% | 73.80% | 595219 | 579097.552 | 2.48 |
| **IRP1A_PG** | rep2 | 3370054 | 2647969 | 78.57% | 78.17% | 585442 | 578133.72 | 2.65 |
| **w1118_PG** | rep1 | 4902502 | 4159179 | 84.84% | 86.81% | 548773 | 548404.388 | 1.8 |
| **w1118_PG** | rep2 | 5532352 | 4621714 | 83.54% | 85.58% | 667496 | 666640.649 | 1.84 |
| **IRP1A_FB** | rep1 | 7817048 | 6591308 | 84.32% | 72.71% | 1857851 | 1798570.21 | 1.01 |
| **IRP1A_FB** | rep2 | 5127364 | 4009854 | 78.20% | 61.35% | 1731572 | 1549792.53 | 1.02 |
| **w1118_FB** | rep1 | 6490896 | 5934675 | 91.43% | 70.08% | 1863683 | 1775791.26 | 1.1 |
| **w1118_FB** | rep2 | 13991175 | 12411920 | 88.71% | 70.04% | 3904008 | 3718474.7 | 1 |

**Rep**: replicate; **Seq Depth**: sequencing depth; **Mapped Frag Num**: mapped fragment number; **Dup Rate**: duplication rate; **Est Library Size**: estimated library size; **Unique Frag Num**: unique fragment number; **SF**: scaling factor.

#### 3.2.1.4 Assessing the reproducibility rate

The reproducibility rate between two replicates of each sample was assessed and visualized in Figure 3.6. In the PG samples, the reproducibility rates for the experimental group (e.g., IRP1A) and the control group (e.g., $w^{1118}$) were determined to be 0.61 and 0.64, respectively. Similarly, in the FB samples, the reproducibility rates for IRP1A and $w^{1118}$ groups were found to be 0.69 and 0.63, respectively. These rates were comparatively lower than the reproducibility rate stated in the CUT & Tag data analysis tutorial [91], which exceeded 0.84. As a result, peak calling was conducted separately on the two replicates rather than pooling them together.

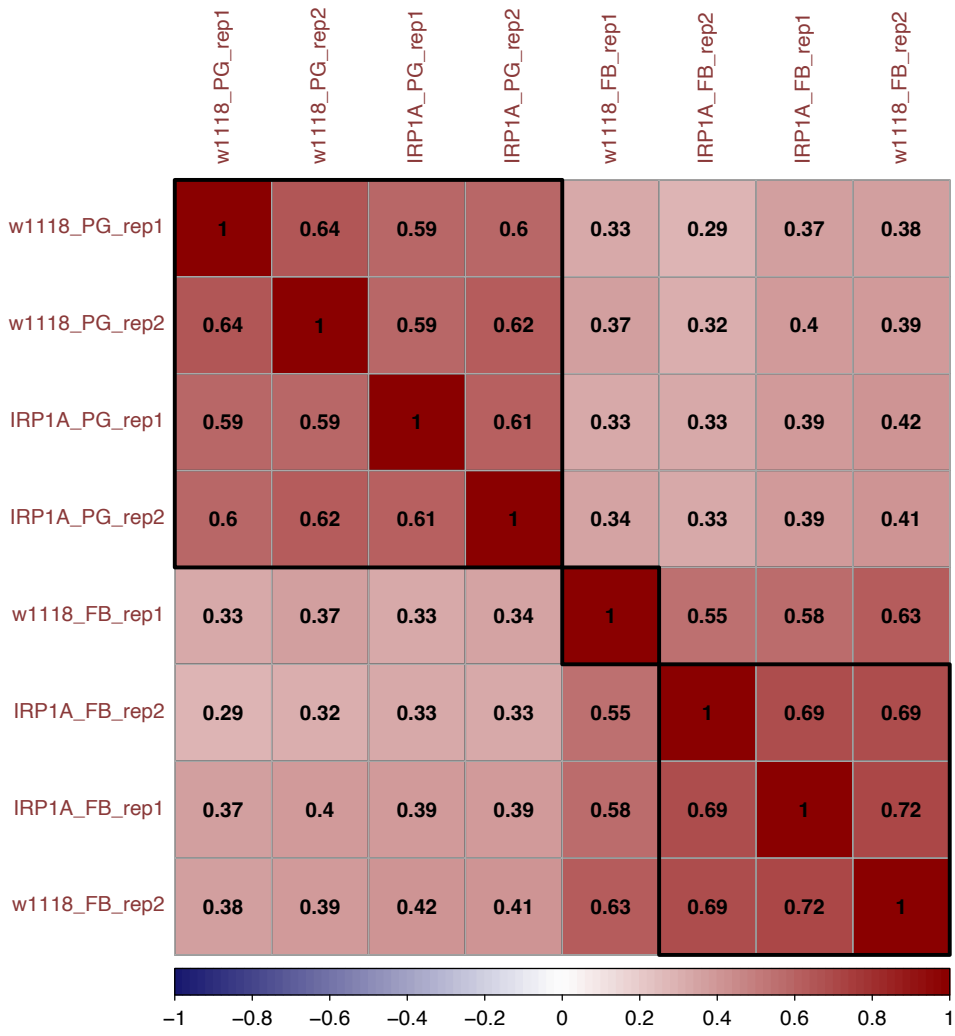



**Figure 3.6 The reproducibility rate between replicates**

### 3.2.2 Peak calling

SEACR (Sparse Enrichment Analysis for CUT&RUN) [93], a newly developed peak calling method specially designed for low background noise experiments such as CUT & RUN and CUT & Tag, was employed for peak calling. The stringent mode was applied to both "control" and "top 0.01" peak types, which meant peaks were called by comparing samples to corresponding controls (e.g., IRP1A_PG_rep1 vs. w1118_PG_rep1) and by calling the top 1% of peaks with the strongest signals, respectively. These two different types of peaks are hereinafter referred to as "control peaks" and "top 0.01 peaks". In comparison to the control, more than 25,000 peaks were called in both PG replicates, whereas only 7 and 16777 peaks were called in two fat body replicates (Table 3.2) (Figure 3.7 A). Alternatively, regarding the top 0.01 peaks, all samples contained approximately 2,000-3,000 targets. It is worth noting that the extremely low number of peaks in IRP1A_FB_rep1 (only 7 peaks) may be attributed to the abnormal control replicate, w1118_FB_rep1. Given this, instead of using the control peaks, the top 1% peaks from the FB samples were subjected for further analysis, while the control peaks from the PG were further analyzed.

Subsequently, the width and reproducibility of peaks were evaluated (Figure 3.7 B and C). In the PG, most control peaks exhibited a width of approximately 400 bp, whereas the width of the top 0.01 peaks was around 1,500 bp. Conversely, in the fat body samples, the top 0.01 peaks were notably wider, spanning approximately 3,000 bp. The reproducibility rate was calculated by dividing the number of overlap targets between rep1 and rep2 by the total number of peaks in either replicate [91], which heavily depended on the peak numbers. For this reason, it was not surprising to observe lower reproducibility rates in the PG replicates (~33% and ~18%) (Figure 3.7 C), considering their great discrepancy in peak numbers, where the peak number in IRP1_PG_rep2 is nearly double than that of IRP1_PG_rep1 (Table 3.2). In the case of the top 0.01 peaks, the reproducibility was much better, ranged from 60% to 70%.

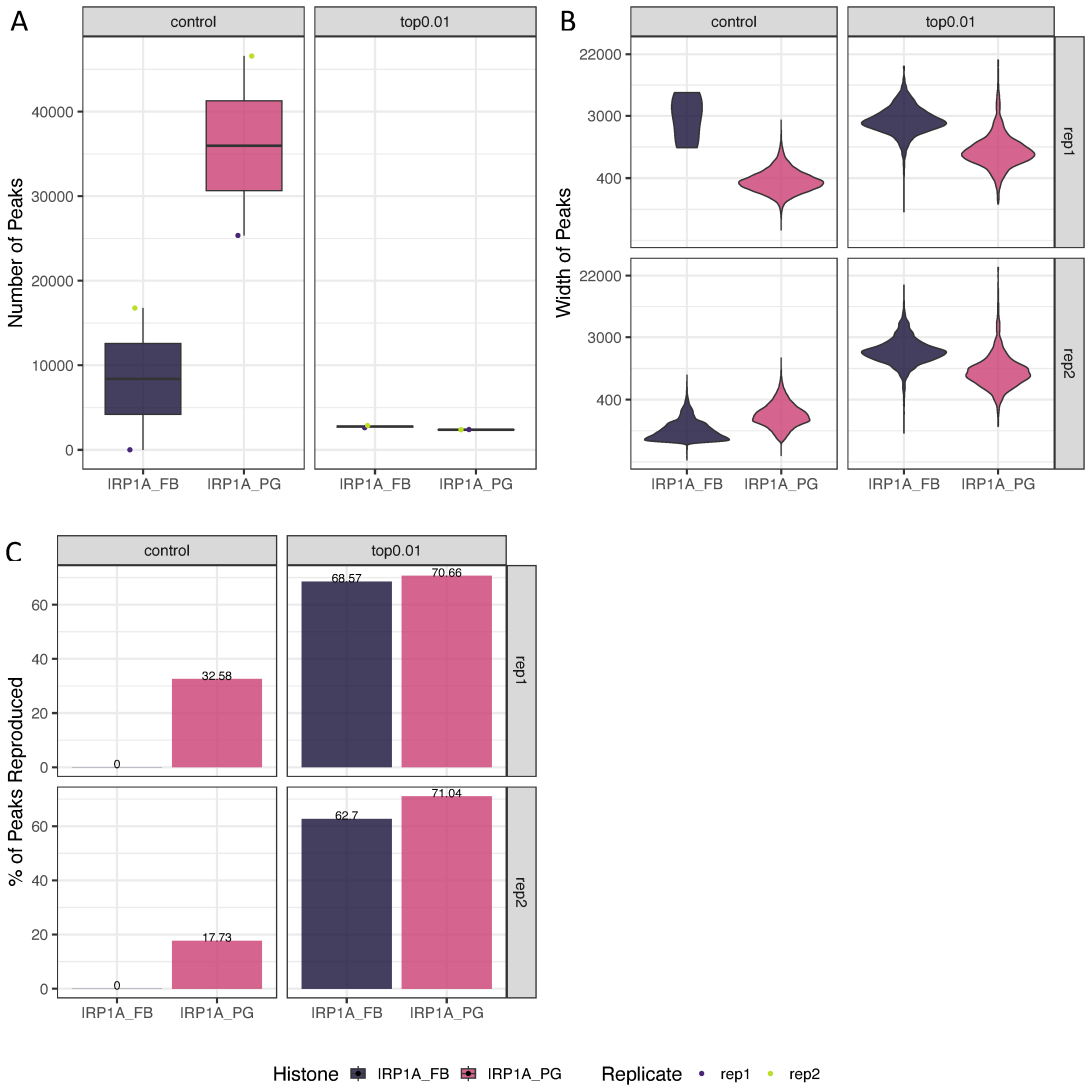



**Figure 3.7 Visualization of the reproducibility rate of peaks between replicates**

(A) Number of peaks. When compared with corresponding control samples, the number of peaks varied greatly in two tissues (~10k in the FB and ~25k in the PG). This can be attributed to the presence of an abnormal control replicate in FB samples, which resulted in only seven peaks were called. (B) Width of peaks. The width of PG peaks (~400 bp for control peaks while ~15,000 bp for top 0.01 peaks) was much smaller than that of FB top 0.01 peaks, which was around 3,000 bp. (C) Reproducibility rates of peaks. The percentages of reproduced peaks in PG controls were low, ranging from 17% to 33%. Regarding the top 0.01 peaks, the reproducibility rate was over 60% in all samples.

**Table 3.2 Summary of the number and reproducibility of peaks**

| Sample | Replicate | Peak Type | Peak Num | Peak Reprod Num | Peak Reprod Rate |
|---|---|---|---|---|---|
| **IRP1A_PG** | rep1 | control | 25357 | 8261 | 32.58 |
| **IRP1A_PG** | rep1 | top 0.01 | 2396 | 1693 | 70.66 |
| **IRP1A_PG** | rep2 | control | 46585 | 8261 | 17.73 |
| **IRP1A_PG** | rep2 | top 0.01 | 2383 | 1693 | 71.04 |
| **IRP1A_FB** | rep1 | control | 7 | 0 | 0 |
| **IRP1A_FB** | rep1 | top 0.01 | 2628 | 1802 | 68.57 |
| **IRP1A_FB** | rep2 | control | 16777 | 0 | 0 |
| **IRP1A_FB** | rep2 | top 0.01 | 2874 | 1802 | 62.70 |

**Num:** number; **Reprod:** reproducibility; **rep**: replicate.

### 3.2.3 IRP1A displayed different binding patterns on the chromatin in PGs and FBs

During development, transcription factors may exhibit oscillating binding activities, which can either enhance or suppress the expression of target genes at specific developmental stages. As a result, this leads to distinct interaction patterns on the chromatin [119, 120]. Similarly, different histone modifications exhibit unique chromatin profiles, where varied peak signals over the genome are shown due to their specific marking preferences. For example, the trimethylation of histone 3 lysine 27 (H3K27me3) modification marks H3 in both the promoter and gene body of the target gene, displaying a dispersed binding pattern across the transcription start site (TSS) to the gene region [121]. Conversely, the trimethylation of histone 3 lysine 4 (H3K4me3) tends to localize to the promoter region, generating a sharp peak signal near the TSS. Therefore, after peak calling analysis, it is important to determine the chromatin regions in which IRP1A is enriched.

By plotting the heatmap of read distribution over the transcription units, distinct chromatin features were observed for the targets of IRP1A in the PG and the fat body. In the PG, IRP1A demonstrated a tendency to bind to the TSS specifically (Figure 3.8). In contrast, in the fat body, IRP1A preferentially resided right after the TSS and within the second half of the gene body (defined as the region between TSS and TES). Typically, canonical transcription factors regulate genes by binding to enhancers or promoter-proximal sequences [122, 123]. When transcription factors interact with promoters near the TSS, their binding pattern should exhibit a peak near the TSS. By contrast, enhancers can be located several hundred kilobases or even megabases away from the TSS [124]. In such cases, when transcription factors bind to enhancers, the binding peak may be distant from the TSS. However, unlike the sharp peak observed in the PG sample, the signal in the FB sample spread out through the gene body. This suggests that in the fat body, instead of interaction with distant enhancers, the binding of IRP1A is more likely associated with chromatin features that extend across the gene body, such as histone modifications. Notably, IRP1A did not show interactions with the transcription end site (TES) in either tissue. To further analyze the distribution of reads within the identified peaks, the fragments were also aligned to the midpoint of the peaks in a heatmap (Figure 3.9). Even though the majority of reads were clustered in the center of the peaks, the signal block in the PG appeared more centered, while in the fat body, it displayed a more dispersed pattern.

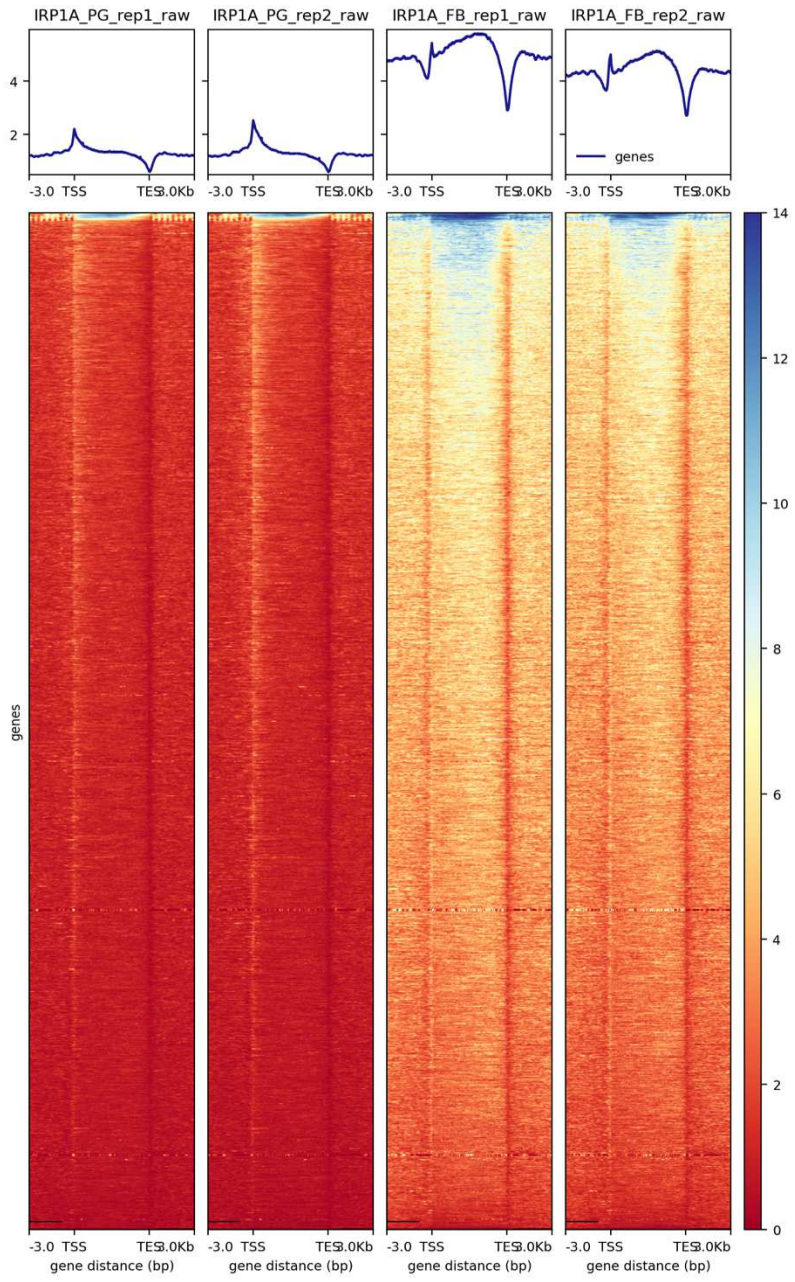



**Figure 3.8 Peaks over transcription units**

IRP1A tended to bind to the TSS in the PG, while in the fat body, it resided right after the TSS and the second half of the gene body. IRP1A did not interact with the TES in neither tissue. The color scale on the Y-axis represents the signal intensity, with higher numbers indicating stronger signals [125]. TSS: transcription start site; TES: transcription end site.

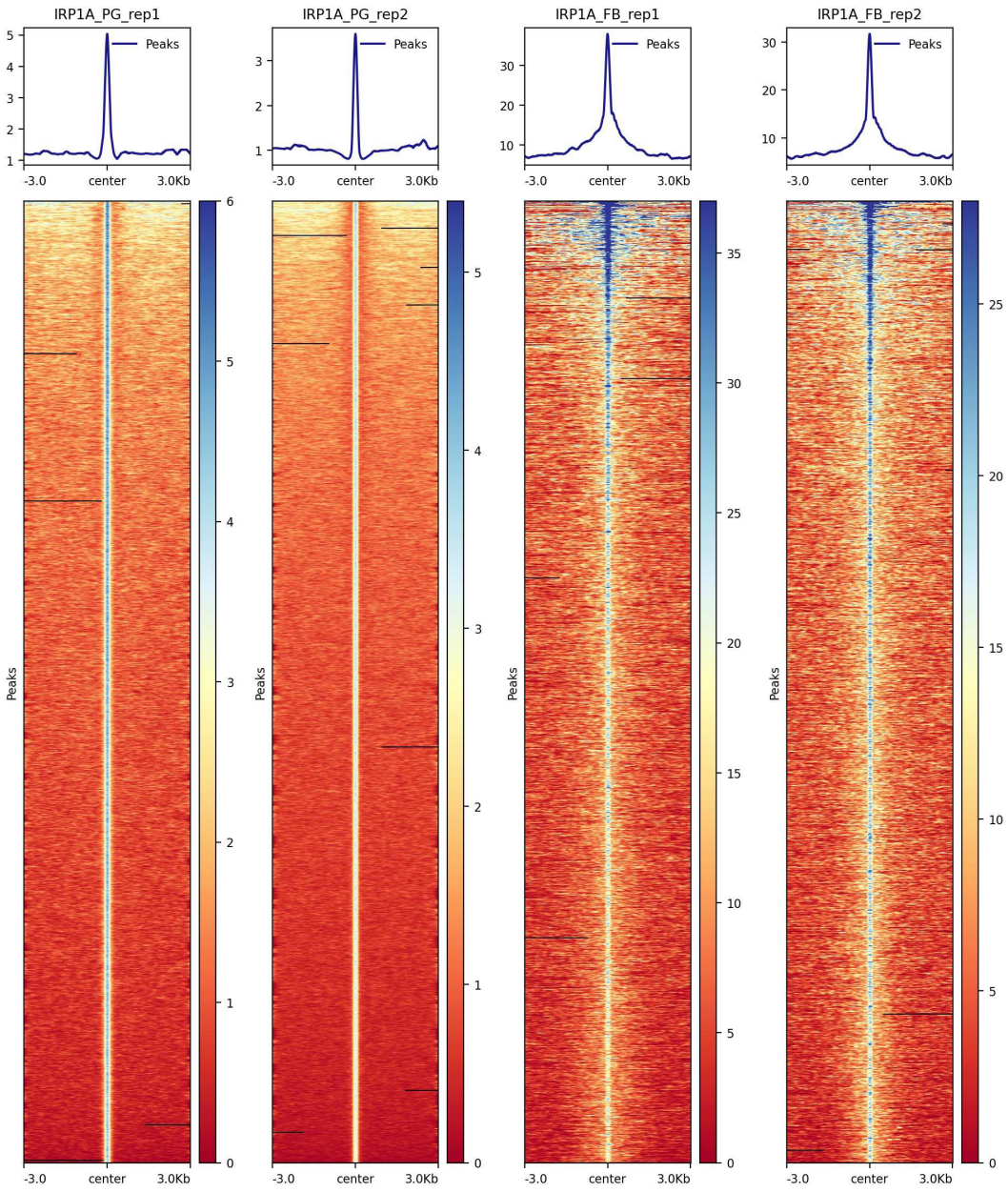



**Figure 3.9 Heatmap of CUT & Tag peaks**

In both PG and fat body, most of the reads were enriched to the midpoint of the peaks. But the signal block in the fat body is more dispersive compared to that in the PG. The color scale on the Y-axis represents the signal intensity, with higher numbers indicating stronger signals [125].

### 3.2.4 IRP1A appeared to bind homeotic genes in the PG

In addition to investigating chromatin binding patterns, I was also interested in identifying the specific genes that are bound by IRP1A. When compared to the control, more than 25,000 peaks were called in both PG replicates (Table 3.2). However, due to the presence of an outlier in the control samples for the fat body, peak calling based on the control background was not feasible. Instead, the top 1% peaks in the fat body were called and used for downstream analysis, resulting in approximately 3,000 bound loci. Peaks were then ranked, based on their signal intensity, retaining the top 2% in the PG and top 50% in the fat body. By finding the overlaps between the replicates, 105 targets in the PG and 588 peaks in the fat body were identified. These candidate targets were then submitted to the DAVID Bioinformatics Resources (https://david.ncifcrf.gov) and PANTHER Classification System (http://pantherdb.org) for gene ontology analysis.

In the PG, target genes were enriched in biological process terms such as "gene regulation" and surprisingly, "segmentation", showing approximately 5-fold and 12-fold enrichment, respectively (Table 3.3). Further analysis of InterPro protein domains confirmed these findings, revealing that many target genes of IRP1A encode DNA-binding proteins that contain zinc finger or helix-turn-helix motifs, as well as homeodomain (HD) proteins. The detailed list of genes enriched in each category can be found in the appendix A 11. Target genes of great interest, including homeobox, segmentation, and zinc finger $C_2H_2$ genes, were listed in Table 3.4. Notably, *Fushi tarazu (ftz)*, a pair-rule gene of the homeotic (Hox) gene family [126], was among the detected genes. The FTZ protein interacts directly with a nuclear receptor called αFTZ-F1 (Fushi tarazu factor 1, α isoform) to regulate the transcription of downstream targets [127]. Specifically, FTZ targets *engrailed (en)* and *ftz* itself, and a more recent finding revealed a new target, *drumstick (drm)* [127-129]. Surprisingly, all these three genes were identified as chromatin targets of IRP1A in the PG. This suggested that in the PG, IRP1A may indirectly downregulate iron- and ecdysone-related genes by modulating the expression of homeotic genes like *ftz* and its targets. An intriguing aspect of this regulation was its potential iron-dependence, as IRP1A enters the nucleus under iron-replete conditions [19].

## Table 3.3 Gene ontology analysis of PG targets

| Category | Term | Num | FE | P-value | FDR |
|---|---|---|---|---|---|
| **Biological process** | Regulation of DNA-templated transcription | 24 | 5.12 | 6.69E-12 | 4.24E-9 |
| | Segmentation | 12 | 12.49 | 2.56E-10 | 7.79E-8 |
| **Cell component** | Nucleus | 29 | 2.44 | 6.03E-7 | 7.71E-4 |
| **Molecular function** | DNA-binding transcription factor activity, RNA Pol II-specific | 21 | 10.25 | 4.88E-16 | 1.46E-12 |
| **InterPro** | Homeodomain | 9 | 13.8 | 2.3E-7 | 3.6E-5 |
| | Zinc finger C2H2-type/integrase DNA-binding domain | 8 | 4.2 | 2.6E-3 | 7.7E-2 |
| | Helix-turn-helix motif | 3 | 36.1 | 2.9E-3 | 7.7E-2 |

**Note:** 105 IRP1A targets genes identified by CUT & Tag in the PG were submitted to PANTHER for biological process, cell component, and molecular function gene ontology analysis, and to DAVID for InterPro protein domain analysis. Num, number of genes that fall in the same category; FE, fold enrichment; FDR, false discovery rate.

## Table 3.4 CUT & Tag target list from PG samples

| PG CUT & Tag target list | |
|---|---|
| **Gene Name** | **Description** |
| ***Fushi tarazu (ftz)*** | Encodes a HOX-like (HOXL) homeobox TF and is a pair-rule segmentation gene expressed in seven stripes in early embryos; the ftz protein interacts with ftz-f1, a nuclear receptor, to directly activate a set of target genes involved in segment development, including segment polarity genes such as *en* and *ftz* itself |
| ***Engrailed (en)*** | Encodes a homeodomain-containing transcription factor that is essential for posterior compartment identity and for compartment boundary formation and maintenance |
| ***HGTX (HGTX)*** | Encodes a homeodomain transcription factor that acts in parallel with the product of *exex* to promote the development and differentiation of motor neurons |
| ***LIM homeobox 1 (Lim1)*** | Encodes a transcription factor protein involved in eye development and leg morphogenesis |
| ***Defective proventriculus (dve)*** | Encodes a transcriptional repressor involved in copper ion import; imaginal disc-derived appendage morphogenesis; and midgut development |
| ***Nubbin (nub)*** | Encodes a POU/homeodomain transcription factor that is expressed in the nascent wing primordium and is required for wing formation |
| ***Scarecrow (scro)*** | An NK-2 homeobox gene that is expressed in specific neuronal clusters of the ventral nerve cord and in the optic lobes of the brain, as well as in the developing pharynx |
| ***Deformed (Dfd)*** | Encodes a homeobox-containing transcription factor mainly involved in proper morphological identity of the maxillary segment and the posterior half of the mandibular segment |
| ***Crocodile (croc)*** | Required for the establishment of head structures; Binds the consensus DNA sequence 5′-[AG]TAAA[TC]A-3′ |
| ***Caudal (cad)*** | Encodes a transcription factor involved in processes such as anterior/posterior patterning formation, organ morphogenesis and innate immune homeostasis |
| ***Drumstick (drm)*** | Encodes a member of the Odd family of zinc finger proteins involved in developmental patterning and cell fate specification |
| ***(A+T)-stretch binding protein (ATbp)*** | May be a transcription factor for genes having (A+T) stretches in their promoter and/or enhancer regions, binds to AT rich DNA |
| ***Lobe (L)*** | Probable transcription factor that can both act as an activator or a repressor; involved in posterior spiracle development |
| ***Sp1 (Sp1)*** | Encodes a member of the Sp-family of C2H2-type zinc finger TFs; involved in ventral thoracic appendage specification, leg growth and in the development of type-II neuroblasts |
| ***Odd skipped (odd)*** | Pair-rule protein that determines both the size and polarity of even-numbered as well as odd-numbered parasegments during embryogenesis; acts primarily as a transcriptional repressor but can also function as a transcriptional activator, depending on the stage of development and spatial restrictions; May function redundantly with odd and drm in leg joint formation during the larval stages, acting downstream of Notch activation |
| ***Earmuff (erm)*** | Encodes a transcriptional repressor that functions as a cell fate regulator by restricting responses to various signaling mechanisms including the Notch pathway during neurogenesis |

| | |
|---|---|
| ***Stripe (sr)*** | Encodes a TF that induces the fate of tendon cells in the embryo as well as in the adult fly |
| ***Chronophage (Cph)*** | Enables RNA polymerase II cis-regulatory region sequence-specific DNA binding activity; involved in negative regulation of peptide hormone secretion; neuroblast fate specification; and positive regulation of gene expression |
| **Note**: gene information is obtained from FlyBase; purple: homeodomain genes; pink: segment specification genes; orange: Zinc finger C2H2-type/integrase DNA-binding domain genes. | |

### 3.2.5 IRP1A interacted with ecdysone-related and hemocyanin genes in the FB

From the FB samples, a total of 588 genes were called based on the signal intensity, representing the top 0.5% of all peaks. Gene ontology analysis using PANTHER and DAVID revealed that a subset of these genes fell in the category transcription regulation (around 4-fold enrichment) and chromatin organization (around 3-fold enrichment) (Table 3.5). Additionally, biological process GO terms "response to ecdysone" and "response to hypoxia" were nearly 5-fold enriched. Furthermore, the InterPro protein domain analysis identified four hemocyanin genes, which represents a remarkable 9-fold enrichment. The detailed list of genes enriched in each category can be found in the appendix A 12.

Table 3.6 presented the targets associated with ecdysone, hypoxia, and hemocyanin genes in the FB. Remarkably, among these targets, nine genes were directly or indirectly related to the ecdysone signaling pathway. Specifically, *ecdysone receptor (EcR)*, which typically heterodimerizes with Usp, was one of the targets. EcR plays a crucial role in the metamorphosis regulation pathway by transducing ecdysone signals to orchestrate the expression of most downstream ecdysone-dependent genes [20]. Additionally, two early genes, *broad (br)* and *ecdysone-induced protein 75B (Eip75B)*, as well as *fat body protein 1 (Fbp1)* are downstream targets of EcR [130-132]. Fbp1 serves as the receptor for larval serum protein 1 (Lsp1), a hemocyanin protein composed of three subunits (Lsp1-α, -β, and -γ) [132]. Moreover, except for responding to heat shock [133] and hypoxia [134], the expression of *Heat shock protein 23 (Hsp23)* can be induced by ecdysone [135]. Furthermore, *Blimp-1* is an ecdysone-inducible gene that encodes a transcription factor controlling the timing of pupation [136]. In conclusion, these findings suggested that in the fat body, IRP1A tended to bind to ecdysone-responsive genes, implying its potential involvement in ecdysone signaling.

## Table 3.5 Gene ontology analysis of FB targets

| Category | Term | Num | FE | P-value | FDR |
|---|---|---|---|---|---|
| **Biological process** | Negative regulation of transcription by RNA Pol II | 22 | 3.98 | 1.31E-7 | 1.11E-5 |
| | Chromatin organization | 19 | 2.66 | 1.77E-4 | 6.73E-3 |
| | Response to ecdysone | 6 | 5.57 | 1.17E-3 | 3.29E-2 |
| | Response to hypoxia | 7 | 4.51 | 1.44E-3 | 3.84E-2 |
| **Cell component** | Brahma complex | 4 | 9.99 | 1.24E-3 | 3.98E-2 |
| | Transcription regulator complex | 16 | 2.88 | 2.53E-4 | 1.16E-2 |
| | Histone deacetylase complex | 6 | 7.05 | 3.87E-4 | 1.65E-2 |
| **Molecular function** | DNA-binding transcription factor activity, RNA polymerase II-specific | 33 | 2.75 | 3.88E-7 | 9.63E-5 |
| | Metal ion binding | 46 | 1.74 | 3.34E-4 | 2.42E-2 |
| **InterPro** | Zinc finger, FYVE/PHD-type | 11 | 3.8 | 5.4E-4 | 6.0E-2 |
| | Hemocyanin | 4 | 9.1 | 7.9E-3 | 4.0E-1 |
| | Steroid hormone receptor | 4 | 5.7 | 3.0E-2 | 9.9E-1 |

**Note:** 588 IRP1A targets genes identified by CUT & Tag in the fat body were submitted to PANTHER for biological process, cell component, and molecular function gene ontology analysis, and to DAVID for InterPro protein domain analysis; Num, number of genes that fall in the same category; FE, fold enrichment; FDR, false discovery rate.

## Table 3.6 CUT & Tag target list of FB samples

| FB CUT & Tag target list | |
|---|---|
| **Gene name** | **Description** |
| ***Ecdysone receptor (EcR)*** | Encodes a protein that interacts with the product of *usp* to form the nuclear ecdysone receptor heterodimer, which modulates, in conjunction with co-activators and co-repressors, the activities of hundreds of genes in a tissue- and stage-specific way |
| ***Broad (br)*** | Broad-complex proteins are required for puffing and transcription of salivary gland late genes during metamorphosis |
| ***Ecdysone-induced protein 75B (Eip75B)*** | Enables DNA binding activity and heme binding activity; involved in ecdysis, chitin-based cuticle, regulation of ecdysteroid metabolic process, and response to ecdysone; controls neuronal remodeling |
| ***Blimp-1 (Blimp-1)*** | Encodes a zinc finger transcription repressor that is induced by 20-hydroxylecdysone and plays important roles for metamorphosis. It also contributes to embryonic trachea formation |
| ***Polyhomeotic proximal (ph-p)*** | Encodes a stoichiometric subunit of the Polycomb repressive complex 1, which silences a variety of genes involved in developmental patterning, regulation of cell proliferation, differentiation, and polarity |
| ***Reptin (rept)*** | Enables transcription corepressor activity; required for instar larval or pupal morphogenesis; proposed core component of the chromatin remodeling Ino80 complex |
| ***Fat body protein 1 (Fbp1)*** | An ecdysteroid-inducible gene encodes the receptor of LSP-1; transcription and translation of Fbp-1 is restricted to the late-third-instar fat body tissue |
| ***Larval serum protein 1 α (Lsp1α) **** | Part of larval serum protein complex, which may serve as a store of amino acids for synthesis of adult proteins |
| ***Larval serum protein 1 β (Lsp1β)*** | Part of larval serum protein complex, which may serve as a store of amino acids for synthesis of adult proteins |
| ***Larval serum protein 1 γ (Lsp1γ)*** | Part of larval serum protein complex, which may serve as a store of amino acids for synthesis of adult proteins |
| ***Prophenoloxidase 3 (PPO3)*** | Expressed in lamellocytes (a type of hemocyte cell involved in encapsulation) and its product involved in the melanization reaction during wasp encapsulation |
| ***Heat shock protein 23 (Hsp23)*** | Encodes a protein involved in protein folding; contributes to the response to cold, heat and hypoxia |
| ***Heat shock protein 70 Ab (Hsp70Ab)*** | Encodes a protein involved in heat shock-mediated polytene chromosome puffing and response to hypoxia |
| ***Heat shock protein 70 Ba (Hsp70Ba)*** | Encodes a protein involved in heat shock-mediated polytene chromosome puffing and response to hypoxia |
| ***Heat shock protein 70 Bb (Hsp70Bb)*** | Encodes a protein involved in heat shock-mediated polytene chromosome puffing and response to hypoxia |
| ***Heat shock protein 70 Bbb (Hsp70Bbb)*** | Encodes a protein involved in heat shock-mediated polytene chromosome puffing and response to hypoxia |
| **Note**: gene information derives from FlyBase; pink: genes response to ecdysone; purple: hemocyanin genes; orange: genes response to hypoxia; *, Lsp1α is detected in the overlap of top 0.01 peaks in the fat body sample. | |

## 3.3 Discussion and future directions

Through the chromatin profile characterized via CUT & Tag, different binding patterns of IRP1A and interactions with different subsets of genes were observed in the PG and the fat body. In the PG, IRP1A demonstrated a tendency to locate near the TSS and associate with homeobox genes, suggesting that it might indirectly downregulate iron- and ecdysone-related genes via the interaction with homeotic genes. By contrast, IRP1A exhibited a dispersed distribution throughout the gene body in the FB, implying its binding might be linked to distributed chromatin features, such as histone modifications. Additionally, a significant number of target genes in the FB were found to be related to ecdysone, indicating that IRP1A may play a role in ecdysone-mediated pathways. The contrasting binding patterns and the enrichment of ecdysone-related genes suggested tissue-specific regulatory roles of IRP1A, highlighting its versatility in both ecdysone synthesis and signal transduction across distinct biological contexts.

### 3.3.1 IRP1A might interact with Hox genes to affect the transcription of iron- and ecdysone-related genes in the PG

The InterPro domain analysis revealed that IRP1A targets in the PG were enriched in Hox genes that encodes HD proteins (Table 3.3). HD proteins are encoded by homeobox genes, which were first discovered through the identification of a shared sequence of ~180 bp in homeotic genes in *Drosophila melanogaster* [137]. The homeodomain is a globular structure that enables the DNA-binding activity of homeobox proteins. All these proteins serve as transcription factors that play critical roles from the earliest developmental stage, responsible for embryonic segmentation, patterning, systemic development and morphogenesis [137, 138]. Given their critical role in cell fate determination and tissue differentiation, homeodomain proteins constitute 15% to 30% of all transcription factors [139]. To date, 103 homeotic genes have been identified in flies [137], accounting for ~0.6% of all *Drosophila* genes (17,728 genes) [140]. Remarkably, among the 105 target genes in the PG, 9 of them were homeotic genes, which is nearly 14-fold higher than the percentage observed in the entire genome (Table 3.3). This suggests that these findings are unlikely false discoveries.

The pair-rule gene *ftz* and its targets *en* and *drm* were identified as target genes of IRP1A in the PG (Table 3.4). During the embryogenesis, the pair-rule genes collaborate with gap and

segment polarity genes to divide the early embryo into 14 identical segments, whose identities are subsequently specified by Hox genes [126, 141]. *Engrailed* is a segment polarity gene responsible for the establishment of segmental border [142], while *drm* is a zinc finger transcription factor involved in hindgut patterning and metamorphosis [143, 144]. Apart from the involvement in determining the parasegment borders in the embryo [142], engrailed is also engaged in the identification of antenna [145], imaginal discs [146], and the genital disc [147]. Hence, the interaction between IRP1 and these homeotic genes may affect the metamorphosis of PG and/or the expression of ecdysone, resulting in the downregulation of iron- and ecdysone-related genes [19]. Because the nuclear localization of IRP1A favors the holo-form [19], which requires iron, this regulation should be iron-dependent.

### 3.3.2 IRP1A may act as an effector protein of ecdysone in the FB during pupation

In the FB, nine genes that are directly or indirectly linked to ecdysone were identified as targets of IRP1A (Table 3.6). Among these genes, EcR and two early genes (e.g., *br* and *Eip75B*) were found. Upon the binding of ecdysone to the EcR, early genes are activated, while the late genes are repressed. As the proteins encoded by the early genes accumulate, they bind to their own promoters to suppress their own expression while activating the late puff genes, thereby facilitating metamorphosis and systemic development [22]. This simple model is known as the "Ashburner model" [148], which straightforwardly explains how ecdysone triggers the gene regulation cascade (Figure 3.10).

Other ecdysone-related genes includes *Lsp1 (Lsp1-α, -β, and -γ)*, which encodes a hemocyanin protein and is exclusively expressed in the fat body during the late-third larval stage [149], demonstrating the specificity of the interaction between IRP1A and its FB targets. It is believed that Lsp1 functions as a storage protein to support nutrition needs during pupariation [150]. Its receptor, Fbp1, exhibits a temporal-specific expression at the same stage as Lsp1 under the regulation of the EcR/USP heterodimer in response to ecdysone [131, 132]. Another gene, *Blimp-1*, responds to ecdysone induction and exhibits active but transient expression during the late larval stage, coinciding with a significant surge of ecdysone. Subsequently, its expression diminishes soon as the ecdysone levels decrease in the early pupal stage [136]. Associated with

βFtz-f1 and Shade, Blimp-1 acts as a precise biological timer in the fat body, controlling the timing of fly pupation [151, 152].

Additionally, *Hsp23*, which encodes a small heat shock protein, can be induced by heat shock [133], hypoxia [134], or ecdysone [135]. The peak expression of *Hsp23* occurs from the late larval stage to pupal period [133]. Induced by ecdysone, EcR binds to the ecdysone response element in *Hsp23*'s promoter in a sequence-dependent manner [135], and its complete ecdysone-responsive expression of relies on the participation of the Broad protein, which is encoded by *br* [153]. Apart from protecting cells from stress damage by heat shock [133] and hypoxia [134], Hsp23 also exhibits a chaperone-like activity that helps maintain unfolded proteins in a refoldable state [154]. Considering its timing of peak expression and its chaperone-like activity, it is possible that Hsp23 plays a role in the regulation of protein denaturation and degradation during the pupariating process in fruit flies [133], where drastic tissue remolding and shaping occur [155].

Taken together, these findings suggest that in the fat body, IRP1A interacts with ecdysone-mediated pupation-related genes, potentially implicating its role as an effector protein of ecdysone and its involvement in metamorphosis, particularly during the larvae-pupation transition.

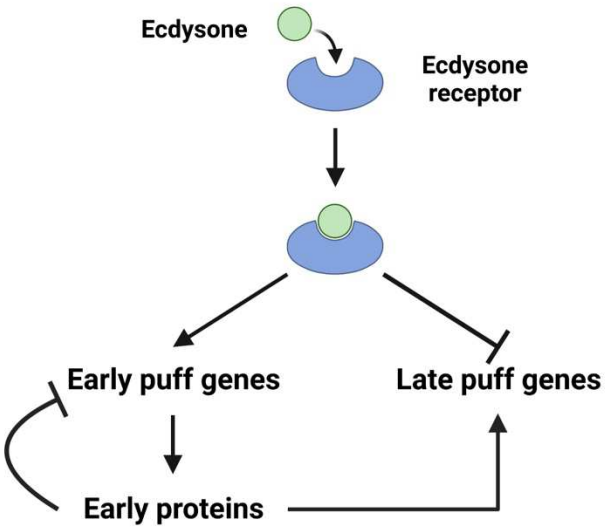



**Figure 3.10 Ashburner model**

By binding to the ecdysone receptor, ecdysone triggers the activation and repression of early and late puff genes, respectively. When early proteins become abundant, they repress their own expression while activating the late puff genes.

### 3.3.3 Future directions

#### 3.3.3.1 Comparing the chromatin profile of fly IRPs

To eliminate any potential confounding effects of overexpressed IRP1A on gene expression, a subsequent CUT & Tag future experiments should be conducted using endogenously tagged-IRP1A, which will be further discussed in Chapter 5. Considering the significant resemblance between IRP1A and IRP1B, including high sequence similarity, shared ability to bind to iron-sulfur clusters, exhibition of aconitase activity, and localization to the nucleus of both the PG and fat body, it is of great value to explore the chromatin profile of IRP1B using CUT & Tag as well. By doing so, the difference of binding patterns/loci between two fly IRPs can be uncovered, thus contributing to a deeper understanding of IRPs.

#### 3.3.3.2 Validating the interaction between IRP1A and its target genes

In order to validate the interaction between IRP1A and its target genes in the PG and the FB, qPCR (quantitative real-time PCR) will be utilized to explore whether IRP1A upregulates or downregulates its targets. Using $w^{1118}$ flies crossed with PG- and FB-specific drivers as controls, the expression levels of IRP1A's target genes in the PG and the FB will be determined under the IRP1A knockdown (e.g., RNAi) context, respectively. As IRP1A functions as the iron sensor in *Drosophila* and its subcellular localization is regulated by iron levels, its regulatory role will be assessed under both iron-deficient and iron-rich conditions using $w^{1118}$ and IRP1A-RNAi lines as well. By comparing with the control (under normal-iron conditions), I anticipate observing opposite regulatory effects when iron is depleted, and an enhanced influence on genes when iron is sufficient.

Considering that IRP1A is not a canonical transcription factor and is unlikely to directly bind to DNA, I hypothesize that the regulation of these genes by IRP1A depends on other DNA/chromatin-binding proteins. Therefore, the protein-protein interaction map of IRP1A was first identified (which will be discussed in Chapter 4). To test this hypothesis, one could perform knockdown experiments (e.g., RNAi) for these potential DNA-binding protein targets characterized in Chapter 4. If the regulation of these genes by IRP1A is dependent on other DNA/chromatin-binding proteins, knocking down these proteins will phenocopy the RNAi of

IRP1A, providing evidence for their involvement in mediating the regulatory effects of IRP1A on its target genes.

#### 3.3.3.3 Is the subcellular location of IRP1A ecdysone-responsive?

In the absence of ecdysone, EcR is found in both the cytoplasm and nucleus [156]. However, upon ecdysone induction, EcR translocates to the nucleus [157]. In the FB, *EcR* (and a couple of other ecdysone-responsive genes) was identified as target genes of IRP1A. Notably, the nuclear localization of IRP1A was discovered in the late L3 stage when there was a surge of ecdysone release. These findings raise questions about whether IRP1A, especially its nuclear localization, responds to ecdysone levels and if it consequently influences the expression of ecdysone- and pupation-related genes in the FB. To investigate this, the subcellular location of IRP1A in the FB could be determined via IF when ecdysone biosynthesis genes are knocked down. In this case, the predominant cytosolic distribution of IRP1A would be expected. Additionally, for validation one would also identify the subcellular location of IRP1A at the onset of the L3 stage when ecdysone levels are low [22]. Because IRP1A enters the PG nucleus as well, these experiments could also be performed using the PG tissue.

#### 3.3.3.4 Are histone modifications involved in the chromatin binding of IRP1A?

In the fat body, the CUT & Tag signal of IRP1A dispersed along the gene body, which implies that the binding of IRP1A may associate with some distributed chromatin features for gene regulation, such as histone modifications. Even though the specific histone modification involved remains unknown, H3 acetylation (e.g., H3K27ac) and methylation (e.g., H3K4me3) are well-defined epigenetic modifications responsible for gene activation and repression, respectively. These might serve as potential starting points to test this hypothesis. To explore this, the binding patterns of H3K27ac and H3K4me3 could be characterized using CUT & Tag with corresponding antibodies. By comparing these chromatin binding profiles obtained under control (e.g., $w^{1118}$ crossed with a FB-specific driver) and IRP1A-RNAi conditions, any significant changes observed could indicate the potential association of IRP1A with these modifications in gene regulation. This comparative analysis will shed light on the possible mechanisms by which IRP1A modulates gene expression through interactions with specific histone modifications in the fat body.

# Chapter 4 The putative mechanism of IRP1A gene regulation

## 4.1 Introduction

Through CUT & Tag, a cluster of homeobox genes involved in gene regulation were identified as chromatin loci to which IRP1A binds in the PG. However, how IRP1A interacts with these target loci remains unclear. Hence, in this chapter, I investigated the protein-protein interaction network of fly IRPs using a more sensitive approach, TurboID proximity labeling [158]. This approach can potentially identify the mechanisms by which IRP1 might affect gene regulation.

### 4.1.1 The regulation of homeotic gene expression

During embryogenesis, early genes like pair-rule and homeotic genes are activated or repressed transiently in response to signals, collectively determining the identities of segments in the developing embryo. Once the segment identities are established, they need to be maintained throughout subsequent stages of development, such as the larval and pupal stages, to ensure the proper body development. Given the fact that cell fate decisions relay on underlying gene expression patterns, this maintenance of segment identities is thus dependent on the maintenance of specific gene expression states [141]. Regulatory proteins known as the Trithorax Group (TrxG) and Polycomb Group (PcG) proteins are responsible for this critical maintenance [159]. Generally, TrxG proteins play a role in gene activation, while PcG complexes repress genes by compacting chromatin, thereby preventing access of transcription factors to DNA recognition sites. In comparison to PcG proteins, the TrxG proteins are more versatile, being able to act as ATP-dependent chromatin remodelers, histone methyltransferases, mediators, and transcription factors [141].

### 4.1.2 Spliceosome assembly and mRNA splicing

As discussed in the introduction of last chapter (3.1.2), the changes in transcript profile caused by nuclear fly IRPs may result from altering gene expression or affecting RNA splicing. In eukaryotes, the initially produced mRNAs by transcription are termed pre-mRNAs (precursor message RNAs), which require intron splicing before nuclear export and translation. RNA splicing is executed via a huge protein complex called the spliceosome, consisting of U1, U2, U4/U6, U5 small nuclear ribonucleoproteins (snRNPs) and other non-snRNP proteins [160]. The initiation of spliceosome assembly depends on the recruitment of U1 to the 5′ splice site (5′ss), resulting in the

E complex (Figure 4.1). Subsequently, the spliceosomal A complex, also known as the pre-spliceosome, is formed when U2 interacts with the branch site (BS). The complex A is bound by the pre-assembled U4/U6.U5 tri-snRNP to generate the pre-catalytic B complex, whose activation relays on the dissociation of U1 and U4 snRNPs. The activated B* complex is capable of catalyzing the ensuing splicing process, where, in the first step, the 2′-OH group of branch site adenosine attacks the 5′ss to produce a "lariat" intermediate. The complex C is formed in this time and facilitates the upcoming intron excision and exon ligation. Spliced pre-mRNAs are exported to the cytoplasm for further modifications and translation, while the spliceosome dissociates and releases snRNPs for another round of RNA splicing [160-162].

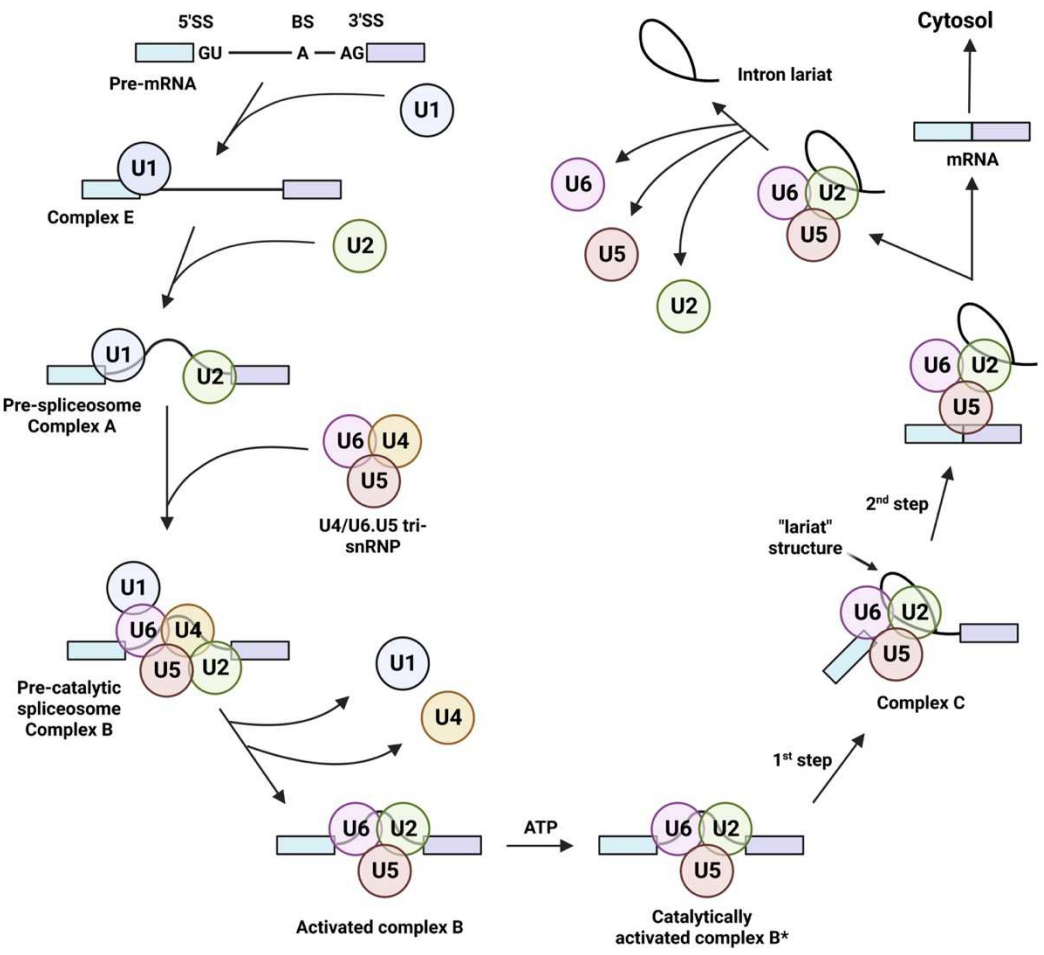



**Figure 4.1 RNA splicing**

RNA splicing starts with the recruitment of U1 snRNP to the 5′ splice site, forming complex E. Subsequently, U2 binds to the branch site, forming complex A (pre-spliceosome), followed by the joining of U4/U6.U5 tri-snRNP to generate the pre-catalytic spliceosome complex B. As the RNA-RNA and RNA-protein interactions rearrange, U1 and U4 dissociate, leading to the activation of complex B, which is further catalytically activated (complex B*). Complex B* catalyzes the first step of RNA splicing, in which the 2′-OH group of branch site adenosine attacks the 5′ss, resulting in the formation of a "lariat" structure and yielding complex C. In the second step, the intron is excised, and exons are ligated by complex C. Following this process, the spliceosome disassembles, and the released snRNPs undergo remodeling to participate in another round of splicing [160, 163]. For simplicity, only snRNPs are shown. 5′SS, 5′ splice site; BS: branch site; 3′SS, 3′ splice site.

### 4.1.3 The advantages of TurboID

TurboID, a recently developed proximity labeling technique, offers a novel approach for studying protein-protein interactions [158]. TurboID is a biotin ligase variant derived from BirA ligase in *E. coli* via directed evolution. It harbors 15 mutations compared to the wild-type BirA. The underlying mechanism of TurboID is straightforward: when TurboID is fused with the target protein (i.e., IRP1A in my case) and exposed to biotin, it can selectively biotinylate all proteins proximal to IRP1A and leave distal proteins unaffected (Figure 4.2) [158]. Those proximal proteins are considered as candidate interacting proteins due to their close physical distance to IRP1A. Following the labeling, biotinylated proteins are immunoprecipitated using streptavidin beads and subsequently identified through mass spectrometry analysis.

In contrast to traditional mass spectrometry, the greatest advantage of TurboID is its high efficiency and sensitivity, allowing researchers to detect more interacting candidates and obtain a more comprehensive interaction map for the protein of interest [164]. Protein- or tag-specific antibodies are used in conventional MS to immunoprecipitate the protein of interest (POI) and thus capture its interacting targets. Hence, the successful performance of this approach heavily relays on the specificity of antibodies, which may not always be ideal or accessible. In contrast, instead of pulling down the POI itself, TurboID immunoprecipitates all the interacting proteins directly, since they were tagged with biotin. The TurboID ligase efficiently adds biotin to all proximal proteins. Coupled with streptavidin beads, which possess very high affinity to biotin (dissociation constant, $K_d = 10^{-14}$ moL/L) [165], more candidates or transient interactions can be detected in TurboID. Furthermore, the inducible expression of TurboID ligase and the efficient boost of biotinylation through biotin allow us to generate spatial- and temporal-specific interaction maps. TurboID can label proteins as early as 10 minutes after biotin addition in cells or 4 hours in larvae and flies, making the identification of interactions within specific time windows or certain developmental stages possible. On the other hand, tissue-specific or even organelle-specific interaction maps can be achieved via utilizing specific promoters [158, 164, 166]. Given these remarkable features, TurboID was employed to explore novel and/or transient interacting candidates of fly IRPs.

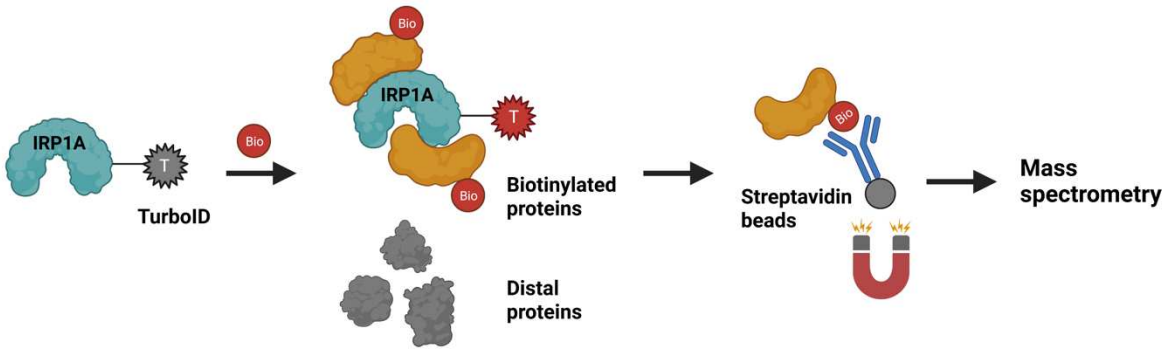



**Figure 4.2 The TurboID procedure**

The protein of interest, in this case, IRP1A, is fused with the biotin ligase TurboID, which is derived from BirA in *E. coli*. In the presence of biotin, the biotinylation is boosted, labelling all the proximal proteins interacting with IRP1A and leaving the distal proteins untouched. Subsequently, biotinylated proteins are pulled down by streptavidin beads and subjected to mass spec for identification [158].

## 4.2 Results

### 4.2.1 Generating of TurboID lines

The generation of TurboID lines were involved in *pUAS-V5-TurboID-NES* plasmid (Addgene, Cat. # 116904). With the NES (nuclear export signal) sequence removed, the V5 tag and TurboID ligase was fused to the N-terminal of IRP1A and IRP1B, respectively (Figure 4.3 A). The fusion protein was expressed in $w^{1118}$ genetic background under the driven of UAS-Gal4 system.

### 4.2.2 Verification of TurboID lines

The homozygous TurboID-IRP1A and -IRP1B lines were crossed with *phm-Gal4* driver for PG-specific expression. Given the cytosolic and nuclear subcellular localization of fly IRPs, TurboID-EGFP with ubiquitous localization in cells served as a control (Figure 4.3 A). Larvae were reared on Nutrifly food to late L3 stage (44 h) for background biotinylation detection. To boost the proximity labeling, larvae were transferred to 100 μM biotin food as 24-hour old third instar larvae for ~6 hours. Late L3 larvae were collected, and their heads removed, and the resulting samples were subjected to immunoprecipitation by streptavidin beads to pull down biotinylated proteins.

To verify the expression of fusion proteins and the activity of TurboID ligase, Western blotting was performed using anti-V5 antibody and streptavidin-HRP, respectively (Figure 4.3 B). Regardless of whether biotin was added or not, all three fusion proteins were expressed with their expected sizes: both TurboID-IRP1A and -IRP1B exhibited a migration rate corresponding to 130 kilodalton (kDa) proteins, whereas TurboID-EGFP migrated faster, at around 70 kDa. Biotinylated proteins were stained via streptavidin-HRP, indicating that the TurboID ligase worked properly after fusing (Figure 4.3 B). With the presence of biotin, the abundance of biotinylated proteins was expected to increase because the biotinylating reaction was enhanced by increased concentrations of the substrate. This was the case for the EGFP control, which showed more and darker bands after the addition of biotin. However, the number and color of bands in group IRP1A and IRP1B were similar with and without biotin.

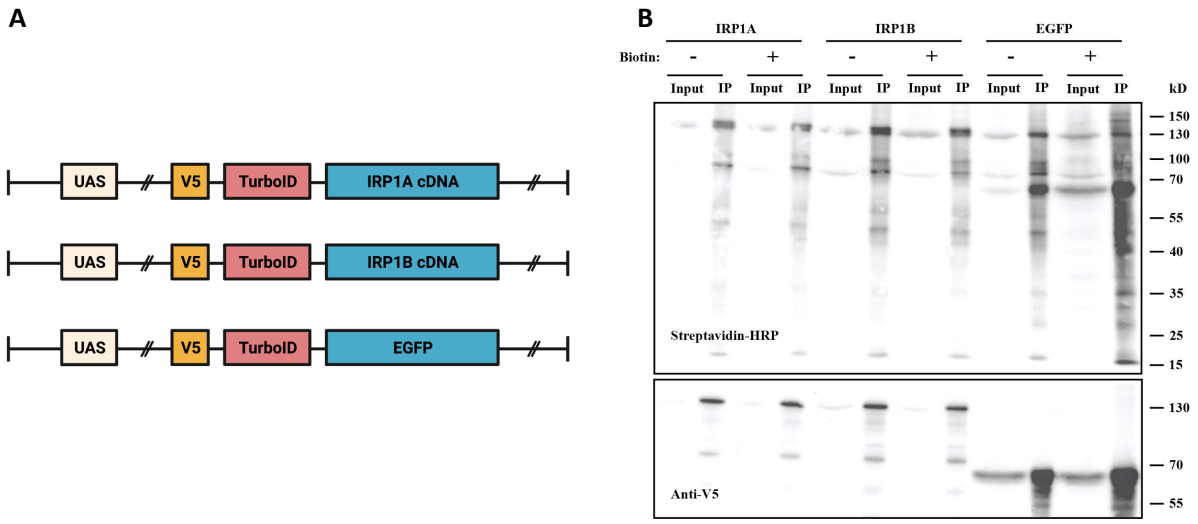



**Figure 4.3 Verification of TurboID lines via Western blotting**

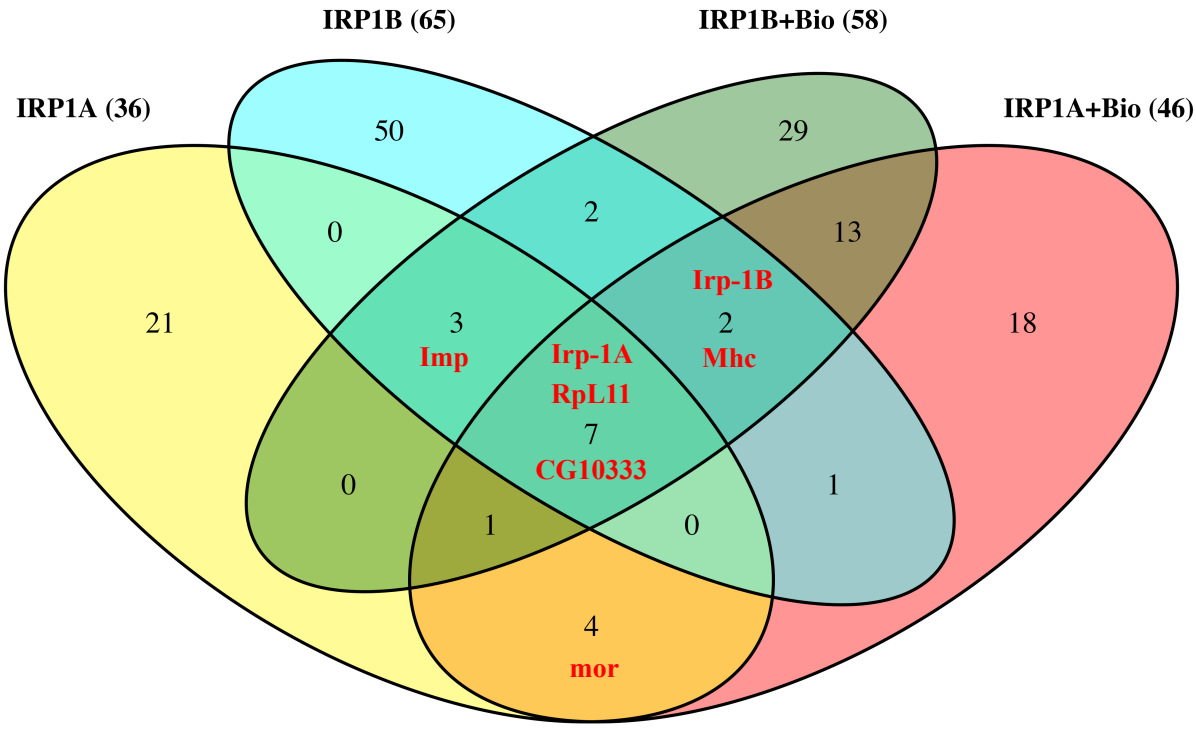
(A) TurboID constructs. The V5 tag and TurboID ligase were fused to the N-terminus of IRP1A, IRP1B, and EGFP, respectively. All TurboID lines are under the control of UAS-Gal4 system. (B) WB analysis of TurboID lines. The input and elution (IP) samples were incubated with streptavidin-HRP (the upper part) and anti-V5 antibody (the bottom part) for biotinylation and fusion protein expression detection, respectively. In the bottom panel, the bands of TurboID-IRP1A and TurboID-IRP1B showed up at ~130 kDa, while TurboID-EGFP is around 70 kDa. Biotinylated proteins were shown in the upper panel. Proteins were biotinylated with or without the addition of biotin in all three groups. But only in the EGFP group, biotinylation was boosted by adding biotin. TurboID lines were crossed with *phm-Gal4*. Larvae were reared in Nutrifly food to wandering L3 (44 h) or transferred to 100 µM biotin food for 20 hours. The head of larvae were homogenized and lysed for WB analysis.

### 4.2.3 Fly IRPs interact with RNA spliceosome subunits

To characterize the protein-protein interactome of fly IRPs in PG, *phm-Gal4* was used for PG-specific expression of TurboID-IRP1A and TurboID-IRP1B. TurboID-EGFP acted as a control. Larvae raised on Nutrifly and biotin food were dissected and subjected to mass spectrometry analysis. Proteins that had a higher score than the control (p-value $\leq 0.05$) or were unique to the treatment group were considered significantly enriched and as interacting candidates.

Through the identification of overlapping proteins across four experimental groups, a total of seven proteins were found to be present in all groups (Figure 4.4). Notably, among these seven targets, IRP1A was present, indicating IRP1B was able to pull down IRP1A. The interaction between IRP1A and IRP1B had been previously reported, albeit with IRP1A serving as the bait [19] (Figure 3.1). The TurboID results further validated this finding. This mutual interaction between fly IRPs indicated that fly IRPs suggests a close relationship between them, implying that they may collaboratively perform certain functions *in vivo*.

Since the primary aim of this study was the nuclear function of IRP1A, I first focused on nuclear interacting targets appeared in the overlapping diagram. Among these proteins, IGF-II mRNA-binding protein (Imp) and CG10333 were detected in both IRP1A and IRP1B groups (Figure 4.4). The Imp and CG10333 proteins are components of the *Drosophila* spliceosome, composing the spliceosomal B and C complexes [161]. Even though spliceosomal proteins are present in abundance in cells to ensure the efficient pre-mRNA splicing and translation, Imp and CG10333 are regarded to interact with fly IRPs specifically due to their absence in the two control groups (namely EGFP and EGFP+Bio; data not shown). Taken together, it can be inferred that fly IRPs may work together to modulate the transcript profile via their interaction with spliceosome (Figure 4.5).




| Symbol | Name | Symbol | Name |
|---|---|---|---|
| **Irp-1A** | Iron regulatory protein 1A | **Imp** | IGF-II mRNA-binding protein |
| **RpL11** | Ribosomal protein L11 | **Irp-1B** | Iron regulatory protein 1B |
| **CG10333** | CG10333 | **Mhc** | Myosin heavy chain |
| **mor** | Moira | | |

**Figure 4.4 Overlap of fly IRPs interacting candidates in the PG**

### 4.2.4 IRP1A appears to bind chromatin remodelers MOR and Psc

Among the overlapping biotin-tagged proteins, a chromatin remodeler Moira (MOR) was identified in both IRP1A groups (with or without biotin) (Figure 4.4). MOR is one of the core components of the Brahma complex that belongs to the SWI/SNF family [167]. SWI/SNF complexes are TrxG proteins with ATP-dependent chromatin remodeling functions, which play a critical role in maintaining the active state of homeotic genes. Being recruited to target genes by transcription activator, the SWI/SNF complex is capable to "slide" the nucleosome along DNA in the presence of ATP, exposing binding sites for general transcription factors and RNA polymerase II (RNA Pol II), thereby facilitating gene expression [141, 168]. Other intriguing interacting candidates that exhibit nuclear localization were listed in Table 4.1, among which Posterior sex combs (Psc) was identified. Psc is part of the Polycomb-repressive complex 1 (PRC1). The subunits of PCR1 and PCR2 are encoded by the *PcG* genes, which are required to maintain the "off" state of Hox genes during development [141]. Upon activation, PCR1 and PCR2 are recruited to Polycomb-response elements (PREs) located in the promoters of Hox genes, thereby blocking transcription by blocking the accessibility of DNAs [168]. To sum up, IRP1A interacts with MOR and Psc in the PG, suggesting its potential role in regulating Hox genes via chromatin remodelers.

## Table 4.1 Interacting candidates of fly IRPs

| Protein Name | Key Role | IRP1A | | IRP1B | |
|---|---|---|---|---|---|
| | Biotin | - | + | - | + |
| **Iron regulatory protein 1A (Irp-1A)** | A cytosolic aconitase that under iron deficiency loses its Fe-S cluster, changes confirmation and binds iron regulatory elements | √ | √ | √ | √ |
| **Iron regulatory protein 1B (Irp-1B)** | A cytosolic aconitase highly homologous to Irp-1A | | √ | √ | √ |
| **Moira (MOR)** | Chromatin remodeler, Brahma complex, SWI/SNF complex, positive regulation of DNA-templated transcription | √ | √ | | |
| **Toutatis (tou)** | Chromatin remodeler, ISWI-type complex, positive regulation by RNA polymerase II | | √ | | |
| **Host cell factor (Hcf)** | ATAC complex; MLL3/4 complex; histone methyltransferase complex; chromatin remodeler; positive regulation of DNA-templated transcription | √ | | | |
| **Posterior sex combs (Psc)** | PRC1 complex, heterochromatin formation; negative regulation of transcription by RNA polymerase II; H2A-K119 mono-ubiquitination | √ | | | |
| **CG10333** | Predicted to enable RNA binding activity; part of catalytic step 2 spliceosome and precatalytic spliceosome; orthologous to human DDX23 (DEAD-box helicase 23) | √ | √ | √ | √ |
| **IGF-II mRNA-binding protein (Imp)** | A protein that regulates the stability, translation and/or transport of its associated mRNAs; part of catalytic step 2 spliceosome and precatalytic spliceosome | √ | | √ | √ |
| **Karyopherin α3 (Kap-α3)** | NLS-bearing protein import into nucleus | √ | | | |
| **Nucleoporin 153kD (Nup153)** | Component of the nuclear pore complex (NPC), essential for the nuclear import of nuclear localization signal (NLS)-containing proteins in an Importin alpha/Importin beta receptor-dependent manner; stimulates transcription by promoting the formation of an open chromatin environment | | √ | | |
| **Megator (Mgtor)** | A nuclear pore protein, during mitosis dynamically reorganizes, together with other nuclear and cytoplasmic proteins to form a gel-like spindle matrix embedding the microtubule spindle apparatus; important for normal cell cycle progression, functioning as a spatial regulator of cell cycle factors | | √ | | |
| **Tudor staphylococcal nuclease (Tudor-SN)** | Endonuclease which shows activity towards both DNA and RNA substrates. Has a role in translation regulation through its association with the RNA-induced silencing complex (RISC) | | √ | | √ |
| **Minichromosome maintenance 5 (Mcm5)** | CMG complex; MCM complex; resolution of meiotic recombination intermediates; endoreduplication; chromosome condensation | √ | | | |
| **Gemini (gem)** | Regulation of transcription by RNA polymerase II | √ | | | |
| **RING-associated factor 2 (RAF2)** | Predicted to enable DNA-binding transcription factor activity, RNA polymerase II-specific; Part of PcG protein complex | | √ | | |
| **scraps (scra)** | A homolog of anilin, a conserved pleckstrin homology domain (PLEKH) containing protein; stabilizes the contractile ring and is required for completion of cytokinesis | √ | | | |

| | | | | | |
|---|---|---|---|---|---|
| **wings up A (wupA)** | Encodes a cytoskeletal protein of the troponin complex of the muscle thin filament, also contributes to non-muscle functions such as apico-basal polarity formation, nuclear division and maintenance of nuclear integrity | | √ | | √ |
| **Myosin heavy chain (Mhc)** | Motor protein that provides the force for muscle contraction through its ATP-dependent interaction with actin filaments | | √ | √ | √ |
| **Ribosomal protein S3 (RpS3)** | Small ribosomal subunit; positive regulation of apoptotic signaling pathway | √ | | | |
| **Ribosomal protein L11 (RpL11)** | Structure constituent of large ribosomal subunit; colocalizes with polytene chromosome puff | √ | √ | √ | √ |
| **Rhino (rhi)** | Chromatin binding; pericentric chromatin; methylated histone binding; positive regulation of piRNA transcription; positive regulation of transcription by RNA polymerase II | | | √ | |
| **winged eye (wge)** | Localizes to the histone locus body and regulates histone expression to ensure genomic stability | | | √ | |
| **Lamin C (LamC)** | A type V intermediate filament that is essential for development; contributes to the shape and structural integrity of the nucleus and plays roles in genome integrity and gene regulation, through contacts made with chromatin | | | | √ |
| **Circadian trip (ctrip)** | Encodes a HECT ubiquitin ligase that regulates the levels and circadian oscillations of the transcriptional activator encoded by *Clk;* negative regulation of histone H2A K63-linked ubiquitination; positive regulation of protein catabolic process | | | √ | |
| **HECT and RLD domain containing protein 2 (HERC2)** | Predicted to enable ubiquitin protein ligase activity; predicted to be located in nucleus and active in cytoplasm and membrane | | | √ | |
| **Hephaestus (heph)** | A nucleocytoplasmic shuttling protein that negatively regulates *osk* mRNA translation; enables mRNA 3′-UTR binding activity and mRNA regulatory element binding translation repressor activity | | | √ | |
| **Caliban (Clbn)** | As part of the ribosome quality control complex (RQC), mediates the extraction of incompletely synthesized nascent chains from stalled ribosomes as well as their ubiquitin-mediated proteasomal degradation. Plays a role in regulating nuclear transport possibly through directly binding to both emb and cargo proteins | | | √ | |
| **belle (bel)** | A DEAD-box RNA helicase that functions both as positive and negative translation regulator | | | √ | |
| **squid (sqd)** | A member of the hnRNPA family of RNA binding proteins; involved in the localization and translational regulation of *grk* mRNA in oogenesis | | | √ | |
| **Polo** | May play a role in regulating both nuclear and cytoplasmic aspects of the mitotic cycle; regulates localization of the augmin complex during mitosis by ensuring its location on mitotic spindles | | | √ | |
| **nudC** | A component of a specialized dynein complex involved in nucleus localization and positive regulation of dendrite morphogenesis | | | √ | |

| | | | | | |
|---|---|---|---|---|---|
| **Cubitus interruptus (ci)** | A Zn-finger family activator, contributes to Hedgehog (Hh) signaling pathway that is involved in pattern formation and growth control | | | √ | |
| **CG10543** | Predicted to enable DNA-binding transcription factor activity, RNA polymerase II-specific; and zinc ion binding activity | | | √ | |
| **Protein tyrosine phosphatase 61F (Ptp61F)** | A non-receptor protein tyrosine phosphatase, acts as a negative regulator of various kinase-dependent signalling pathways, including the JAK-STAT, Insulin-like Receptor, EGFR, and Pvr pathways in a wide variety of contexts | | | √ | √ |
| **Spargel (srl)** | Transcriptional coactivator required for the expression of multiple genes encoding mitochondrial proteins, mediates cell growth and transcription of target genes in response to insulin signalling | | | | √ |
| **Proteasome α1 subunit (Prosα1)** | The proteasome is a multi-catalytic proteinase complex which is characterized by its ability to cleave peptides with Arg, Phe, Tyr, Leu, and Glu adjacent to the leaving group at neutral or slightly basic pH; the proteasome has an ATP-dependent proteolytic activity | | | | √ |
| **14-3-3zeta** | Functions in multiple signaling pathways, most prominently in the Ras/MAPK cascade. It is involved in epithelial cell polarity, development of the eye, embryogenesis and in adult associative learning | | | | √ |

**Note**: Mass spectrometry analysis of PG samples with PG-specific expression of *IRP1A*$^{WT}$ and *IRP1B*$^{WT}$ transgenes in Nutrifly food or biotin food. Nuclear or chromatin-related targets are shown. Proteins related to chromatin remodeling were indicated in orange, while spliceosomal proteins were in green. Protein information is from FlyBase.

## 4.3 Discussion and future directions

### 4.3.1 Biotin did not boost the biotinylation in *TurboID-IRP* lines

Using WB, the biotinylation activity of TurboID was assessed in *TurboID-IRP1A, -IRP1B*, and -*EGFP* lines (Figure 4.3 B). In the EGFP group, biotinylation was enhanced upon the addition of biotin, whereas in the IRP1A and IRP1B groups, the levels of biotinylating remained similar before and after adding biotin. There are several possible explanations for this situation. Since biotinylation was boosted in the EGFP group, the exposure time to biotin (~6 h) should not be the problem. One explanation is that the turnover rate of most IRPs interacting proteins is relatively low. After binding to IRPs, they maintain the interaction or just exist alone for a long period. Hence, although biotin was added, no more proximal proteins were labelled. Another possibility is that the biotinylating reaction was limited by some unknown mechanisms because robust biotinylation, especially nuclear biotinylation, can be detrimental to animals. TurboID ligase is highly active and efficient, it may lead to protein malfunction due to persistent biotinylation and thus cause cell toxicity [158, 166]. Additionally, according to our lab experience, the toxicity of TurboID varied, based on the subcellular location of the fusion proteins. For instance, ubiquitous overexpression of nuclear TurboID fusion protein via whole-body Gal4 driver in flies resulted in an L2/L3 arrest through an unknown mechanism. Addition of biotin worsened the survival rate, which suggests that this phenotype was not caused by biotin depletion via the high activity of TurboID [158]. On the other hand, larvae overexpressing TurboID fusion proteins within the cytosol and/or plasma membrane were able to survive to adulthood.

### 4.3.2 IRP1A may indirectly affect the transcript profile via the interaction with chromatin remodelers

Through MS, chromatin remodelers MOR and Psc were identified as interacting candidates of IRP1A in the PG (Figure 4.4 and Table 4.1). MOR, part of the Brahma complex, is a TrxG protein involved in activating Hox genes [167]. Previous studies have indicated that MOR is required for the regulation of segment gene *engrailed* in imaginal tissues [169]. Additionally, the Brahma complex is ecdysone-inducible [170]. Interestingly, *engrailed*, and a bunch of other homeotic genes, were determined as target genes of IRP1A detected by CUT & Tag in the PG. Therefore, taken together, IRP1A might participate in the regulation of Hox genes via the

interaction with MOR to indirectly alter the transcript profile of iron- and ecdysone-related genes, and this might be a response to ecdysone (Figure 4.5). On the other hand, Psc is one of the components of PRC1, which is required to maintain the "off" state of Hox genes during development [141]. Notably, considering that a subset of Hox genes was identified as chromatin targets of IRP1A, it is plausible that IRP1A also contributes to the maintenance of the repressive state of homeotic genes by interacting with PcG protein Psc, thereby affecting the transcription of genes related to iron and/or ecdysone (Figure 4.5).

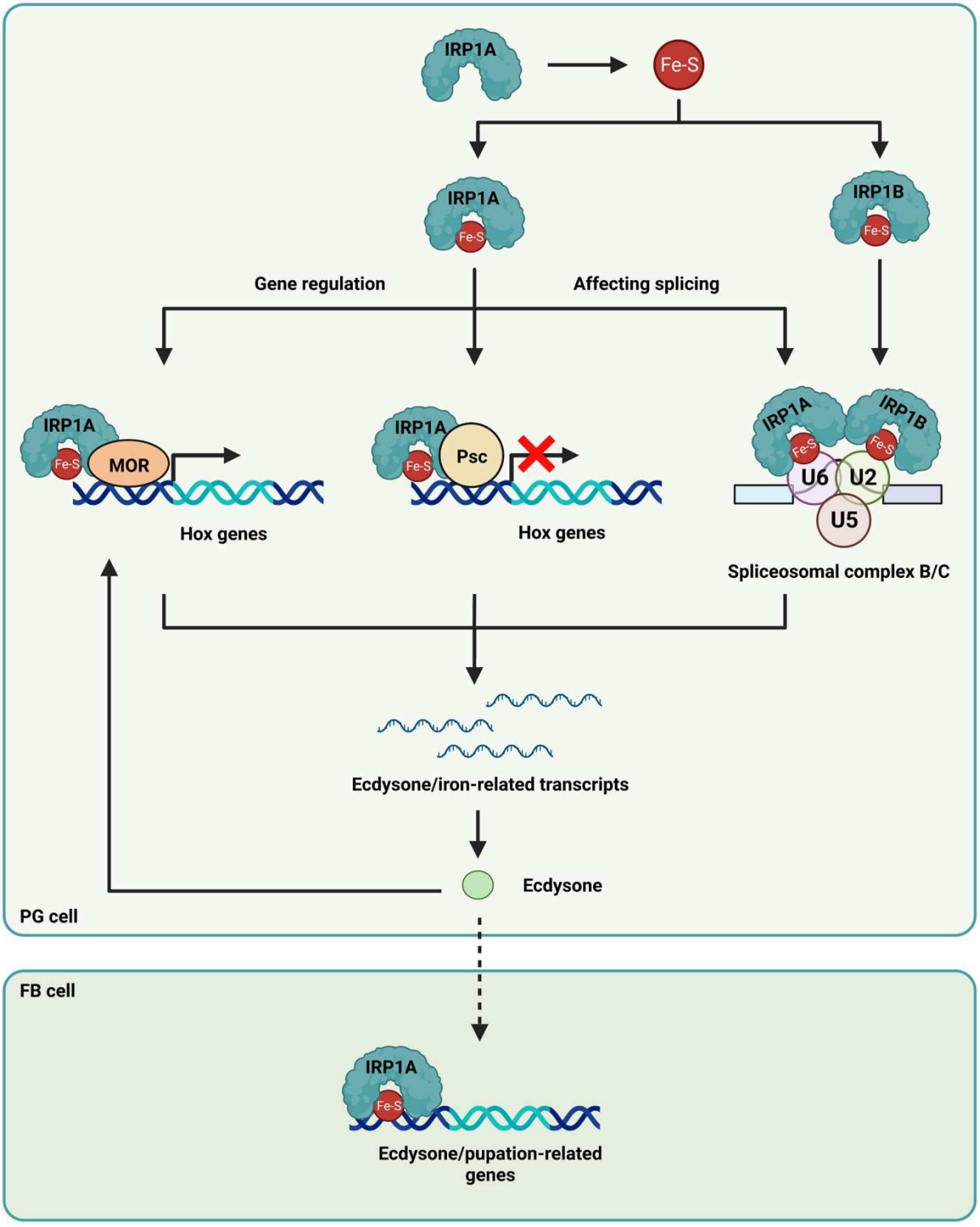



**Figure 4.5 Proposed working model of fly IRPs in the nucleus**

In the PG, when iron is deficient, apo-IRP1A upregulates iron concentrations by binding to mRNAs, which gives rise to the formation of holo-IRPA and holo-IRP1B. The elevated iron levels further facilitate the synthesis of ecdysone as well. The aconitase form of IRP1A translocates to the nucleus in PG cells and exerts its regulatory influence on the transcript of genes related to iron and ecdysone via affecting gene regulation and splicing. Specifically, IRP1A interacts with TrxG proteins MOR, implying its potential role in maintaining the active state of homeotic genes and thus modulating the expression of genes related to iron and ecdysone. Because the Brahma complex, which includes MOR, is inducible by ecdysone, this process may also be regulated by ecdysone. In addition, IRP1A may be involved in the maintenance of repressed state of homeotic genes via its interaction with PcG protein Psc. Moreover, given the intimate relationship between IRP1A and IRP1B, both of which interact with Imp and CG10333 (components of spliceosomal complexes B and C), it is conceivable that they function together to affect RNA splicing, thereby influencing the transcript profile. On the other hand, in the FB, IRP1A may function as an effector protein of ecdysone, potentially participating in the regulation of ecdysone and pupation genes and modulating the larvae-pupa transition.

### 4.3.3 Future directions

In the PG, both IRP1A and IRP1B appeared to interact with the spliceosomal proteins Imp and CG10333, suggesting their potential involvement in RNA splicing and its impact on the transcriptome. Additionally, IRP1A exhibited interactions with chromatin remodelers such as Moira and Psc, implying that IRP1A might play a role in the maintenance of Hox gene states by binding to TrxG and PcG proteins. Having said that, further investigations, such as co-IP, are necessary to validate these physical interactions. Moreover, considering that the nuclear localization of IRP1A was observed in the fat body as well, and chromatin targets in this tissue have been characterized, future experiments should include TurboID using fat body samples. This will help to elucidate the mechanisms by which IRP1A binds to chromatin in the fat body and carries out its regulatory functions. Furthermore, the interactomes of IRP1A variants ($IRP1A^{C450S}$ and $IRP1A^{3R3Q}$) should be characterized using TurboID, in order to confirm the specificity of these interactions. For instance, restricted by their subcellular localization, it is expected that the majority of interacting candidates for $IRP1A^{C450S}$ would be found in the cytosol, while those for $IRP1A^{3R3Q}$ should have the ability to enter the nucleus.

# Chapter 5 Assessing the viability of fly IRP mutants

## 5.1 Introduction

### 5.1.1 PG-specific overexpression of IRP1A and its variants results in low viability

Both the CUT & Tag and TurboID approaches explore IRP1A's nuclear function at the molecular level. However, I would like to investigate its influences on animal viability as well. By doing so, we can obtain a more complete picture about the roles of nuclear IRP1A with observable phenotypes and underlying mechanisms. As previously mentioned, the PG is sensitive to iron levels. A survival rate assay carried out by Song Wang from our lab showed that only 30% of larvae survived to the adulthood when she overexpressed wild-type IRP1A in the PG, indicating that an excess of IRP1A was harmful to flies. When IRP1A mutants, specifically IRP1A$^{C450S}$ and IRP1A$^{3R3Q}$, were overexpressed in the PG, the viabilities dropped even further to 18% and 22%, respectively. This implied that both RNA-binding and aconitase activities of IRP1A are critical for fly survival. The toxicity of these variants may arise from their competition with endogenous IRP1A, thereby impeding the function of normal IRP1A. However, it is worth noting that this experiment was conducted in overexpression lines, where a sheer abundance of proteins was expressed, potentially leading to toxicity due to non-specific functions. Therefore, to obtain more accurate results, a better approach, such as using CRISPR (clustered regularly interspaced short palindromic repeats) knock-in lines, should be employed to measure the survival rate of IRP1A and its mutants.

To mitigate the potential issues caused by overexpression itself and the interference of endogenous IRP1A in transgenic lines, CRISPR knock-in lines were generated. These knock-in lines specifically targeted the *IRP1A* gene and introduced tags (*IRP1A$^{WT.3F}$*) and mutations, resulting in the establishment of *IRP1A$^{C450S.3F}$* and *IRP1A$^{3R3Q.3F}$* mutants. By replacing the native IRP1A with the desired mutants, we aim to gain a better understanding of IRP1A's roles in both the cytosol and nucleus via assessing their impact on survival rates in *Drosophila*. For instance, IRP1A$^{C450S}$ and IRP1A$^{3R3Q}$ are expected to primarily localize to the cytosol and nucleus, respectively. By measuring and comparing the viability of these mutants, we can figure out which function of IRP1A is more important for the fly's survival, the cytosolic or nuclear role. Furthermore, to explore the significance of the conformational interconversion of IRP1A, a heteroallelic combination of IRP1A$^{C450S/3R3Q}$ will be generated to see whether this mutant is viable

or not. These comprehensive approaches allow us to determine the biological significance of cytosolic and nuclear IRP1A in flies.

## 5.2 Results

### 5.2.1 Generation of CRISPR/Cas9 knock-out and knock-in lines

To further explore the physiological functions of IRP1A, I generated a knock-out line, $IRP1A^{KO}$, via CRISPR, whereby the entire *IRP1A* gene was deleted (Figure 5.1). Additionally, in order to compare the difference between cytosolic and nuclear IRP1A with respect to their impacts on fly's survival, CRISPR knock-in lines were introduced to avoid the influence of overexpressing proteins. Epitope tag 3xFlag was added to the C-terminus of wild-type IRP1A by CRISPR, termed $IRP1A^{WT.3F}$ (Figure 5.1). Two mutant lines were constructed by me as well, namely $IRP1A^{C450S.3F}$ and $IRP1A^{3R3Q.3F}$. $IRP1A^{C450S.3F}$ contains a C450S point mutation and is predicted to be mainly RNA-binding and cytosolic because the Fe-S-binding site was mutated. It was tagged with 3xFlag at the C terminus as well. On the other hand, three arginine-to-glutamine point mutations were introduced in $IRP1A^{3R3Q.3F}$ with a C-terminus 3xFlag tag. The RNA-binding ability is predicted to be abolished in this mutant, while allowing the protein to bind Fe-S clusters and thus enter nuclei.

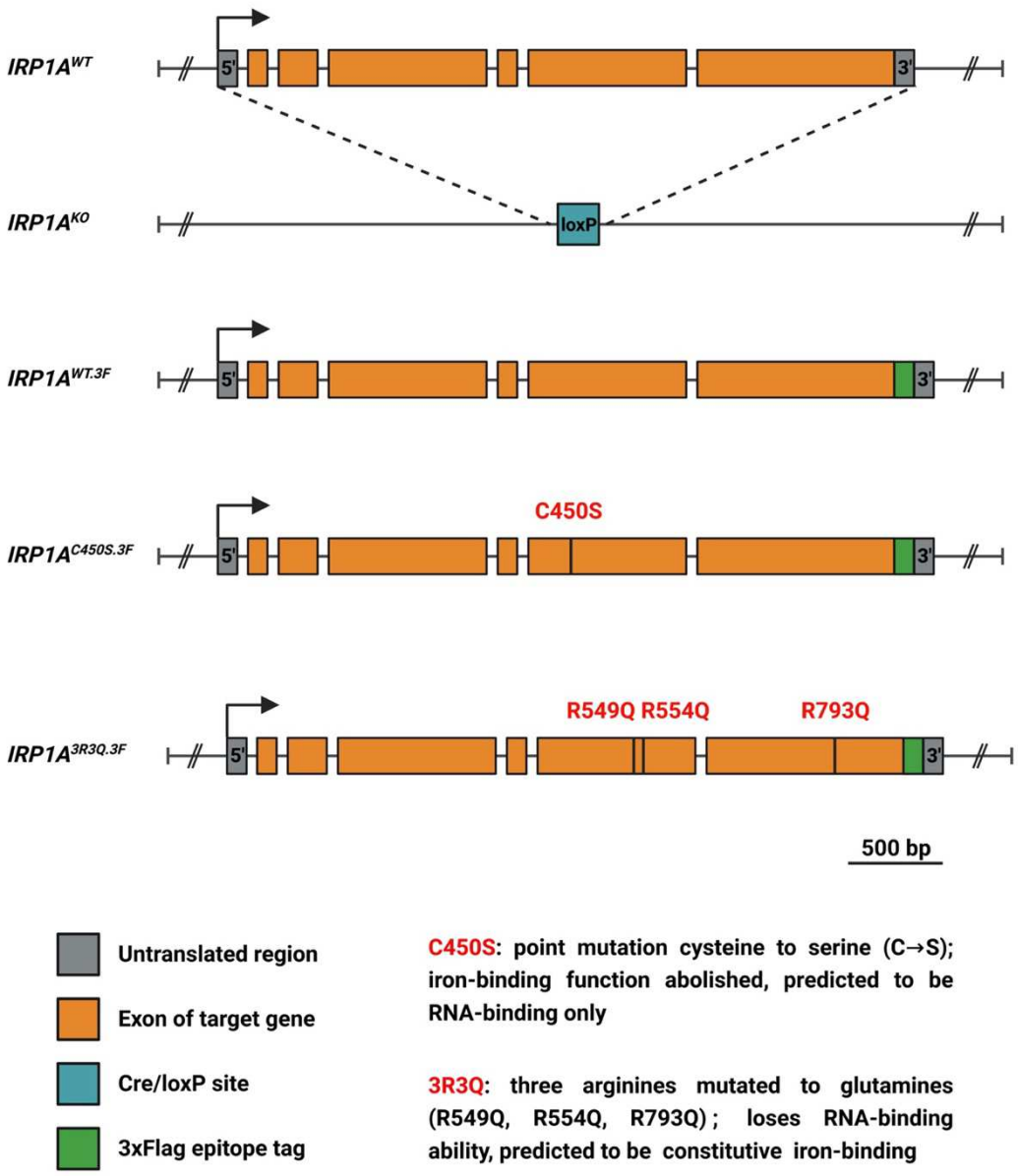



**Figure 5.1 IRP1A knock-out and knock-in lines generated by CRISPR/Cas9**

The complete IRP1A gene region was deleted and replaced by a Cre/loxP site to generate a IRP1A knock-out line, termed *IRP1A*$^{KO}$. Wild-type IRP1A was tagged with 3xFlag at the C-terminus. Two mutant lines, each with C terminal 3xFlag tags, *IRP1A*$^{C450S.3F}$ and *IRP1A*$^{3R3Q.3F}$, carrying the C450S and 3R3Q mutations, respectively. KO: knock-out; WT: wild type; 3F: 3xFlag tag.

## 5.3 Discussion and future directions

Due to time limitation, I did not perform any experiments using CRISPR knock-in or knock-out lines. However, I have verified the lines by Sanger sequencing. In the following, I will discuss some experiments that could be done to further examine IRP1A.

### 5.3.1 Comparing the significance of cytosolic and nuclear IRP1A to fly's survival

The survival rates of $IRP1A^{C450S.3F}$ and $IRP1A^{3R3Q.3F}$ knock-in mutants could be evaluated and compared to determine whether the RNA-binding only form or the iron-binding IRP1A exhibits any detrimental phenotypes, which suggests a role for the holo- and apo-IRP1A, respectively. Because IRP1A is the main iron regulator and its subcellular localization is affected by iron concentrations, these experiments will be performed under low-iron (BPS food), normal-iron (Nutrifly food), and iron-rich (FAC, ferric ammonium citrate food) conditions (Table 5.1). Compared to wild-type IRP1A, which is bifunctional and able to shuttle between cytosol and nucleus, $IRP1A^{C450S.3F}$ and $IRP1A^{3R3Q.3F}$ possess either RNA- or Fe-S-binding function and reside in cytoplasm or nucleus, respectively. Hence, based on the assumption that both cytosolic and nuclear IRP1A are important for fly's survival, it is expected that the survival rates of both mutants will decrease under normal conditions.

In human cells, the RNA-binding mutant $hIRP1^{C437S}$ loses the ability to respond to changes in iron concentrations. It continuously binds to the mRNAs encoding the transferrin receptor and ferritin, leading to increased and decreased expression, respectively [82]. As a consequence, cells take up iron continuously and fail to store excess iron in ferritin, rendering the cells vulnerable to iron overload. Therefore, for $IRP1A^{C450S}$, the equivalent mutant in *Drosophila*, similar phenotypes are expected, with the survival rate being worsened by iron supplementation and rescued by iron depletion. In contrast, because $IRP1A^{3R3Q}$, the holo-form mutant, requires iron to function, the survival rate of this mutant is predicted to be diminished by iron deficiency and rescued by iron addition. To establish reasonable controls, these traits will be observed first in $IRP1A^{WT.3F}$ and $w^{1118}$ lines. If the phenotype of $IRP1A^{WT.3F}$ shows no significant difference compared to that of $w^{1118}$, meaning that the CRISPR/Cas9 editing itself makes no noticeable changes to the animal, $IRP1A^{WT.3F}$ will serve as a control. Otherwise, $w^{1118}$ will be used as a control line.

### 5.3.2 Is the ability to switch from apo- to holo-IRP1A essential?

Since wild-type IRP1A can switch between the apo- and holo-form, we wondered if this interconversion is necessary to flies that are heteroallelic for both knock-ins (*IRP1A*$^{C450S/3R3Q}$). This heteroallelic line should cover both aspects of the protein but separates them into two alleles. IRP1A$^{C450S}$ is predicted to be RNA-binding only, which mimics the apo-form of IRP1A. On the other side, IRP1A$^{3R3Q}$ cannot bind RNA, but can assume the holo-form and is predicted to have aconitase activity. As such, flies expressing both mutated proteins might survive on normal, iron-supplemented, or even iron-depleted food. To test this theory, *IRP1A*$^{C450S.3F}$ and *IRP1A*$^{3R3Q.3F}$ knock-in lines will be crossed to make an *IRP1A*$^{C450S/3R3Q}$ mutant. The survival rate of this mutant line will be measured and compared with *IRP1A*$^{WT.3F}$ (or $w^{1118}$) under BPS, Nutrifly, and FAC food (Table 5.1). If the viability is similar, we would conclude that the conformational change between apo- and holo-IRP1A is not essential for survival. Conversely, if there is a significant difference in the survival rate of two groups, the interconversion may imply an important balance of apo- and holo-IRP1A rather than simply expressing both forms of IRP1A constitutively.

### 5.3.3 Exploring the function of IRP1B in iron metabolism

IRP1A and IRP1B are highly similar in sequence [81], yet their null mutants display significantly different survival rates [19]. *IRP1A*$^{KO}$ is L2 arrested on normal food while *IRP1B*$^{KO}$ manages to survive until adulthood even under iron-depleted conditions [19]. Despite this knowledge, our understanding of IRP1B is limited to its aconitase activity [6]. Therefore, I am wondering if IRP1B plays a role in iron metabolism. To investigate this, we will assess the viability of different IRP1B genotypes, including wild-type IRP1B (*IRP1B*$^{+/+}$), IRP1B single knock-out (*IRP1B*$^{+/-}$), and double knock-out (*IRP1B*$^{-/-}$) under the *IRP1A* null-mutation (*IRP1A*$^{-/-}$) context. We will examine whether there are any differences in their survival rates when exposed to varying iron concentrations (e.g., iron-deficient, normal-iron, and iron-rich) (Table 5.1). The survival rate is expected to be higher when iron is supplemented and lower when iron is depleted. By exploring the viabilities of these mutants under different iron conditions, we aim to gain insights into the potential role of IRP1B in iron metabolism.

**Table 5.1 Proposed survival rate assay to explore the physiological function of IRPs**

| Purpose | Food | Control | Experiment |
|---|---|---|---|
| To identify the importance of cytosolic and nuclear IRP1A | BPS Nutrifly FAC | *IRP1A*$^{WT.3F}$ or *w*$^{1118}$ | *IRP1A*$^{C450S.3F}$ |
| | | | *IRP1A*$^{3R3Q.3F}$ |
| To explore if interconversion of apo- and holo-IRP1A is necessary | | *IRP1A*$^{WT.3F}$ or *w*$^{1118}$ | *IRP1A*$^{C450S/3R3Q}$ |
| To investigate IRP1B's role in iron metabolism | | *IRP1A*$^{WT.3F}$ or *w*$^{1118}$ | *IRP1B*$^{+/+}$/ *IRP1A*$^{KO}$ |
| | | | *IRP1B*$^{+/-}$/ *IRP1A*$^{KO}$ |
| | | | *IRP1B*$^{-/-}$/ *IRP1A*$^{KO}$ |

### 5.3.4 Characterizing chromatin binding loci of endogenous IRP1A

In this research, I conducted CUT & Tag analysis using the IRP1A-overexpression line. Since excessive protein levels could lead to non-specific binding and the overexpression of IRP1A might affect biological processes, potentially influencing the chromatin binding profile, it is crucial to ensure the robustness of the findings. For this purpose, CUT & Tag will be performed in the *IRP1A*$^{WT.3F}$ knock-in line to ensure the identification of bona fide binding loci and eliminate any potential false positives that could arise from gene overexpression.

## References


1. Gorman, M.J., *Iron Homeostasis in Insects.* Annu Rev Entomol, 2023. **68**: p. 51-67.
2. Muckenthaler, M.U., et al., *A Red Carpet for Iron Metabolism.* Cell, 2017. **168**(3): p. 344-361.
3. Rewitz, K.F., et al., *The Halloween genes code for cytochrome P450 enzymes mediating synthesis of the insect moulting hormone.* Biochem Soc Trans, 2006. **34**(Pt 6): p. 1256-60.
4. Crack, J.C., et al., *Iron-sulfur cluster sensor-regulators.* Curr Opin Chem Biol, 2012. **16**(1-2): p. 35-44.
5. Schiffer, L., et al., *Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review.* J Steroid Biochem Mol Biol, 2019. **194**: p. 105439.
6. Lind, M.I., et al., *Of two cytosolic aconitases expressed in Drosophila, only one functions as an iron-regulatory protein.* J Biol Chem, 2006. **281**(27): p. 18707-14.
7. Ward, R.J., et al., *The role of iron in brain ageing and neurodegenerative disorders.* Lancet Neurol, 2014. **13**(10): p. 1045-60.
8. Calap-Quintana, P., et al., *Drosophila melanogaster Models of Metal-Related Human Diseases and Metal Toxicity.* Int J Mol Sci, 2017. **18**(7).
9. Oakley, A.E., et al., *Individual dopaminergic neurons show raised iron levels in Parkinson disease.* Neurology, 2007. **68**(21): p. 1820-5.
10. Zhou, Z.D. and E.K. Tan, *Iron regulatory protein (IRP)-iron responsive element (IRE) signaling pathway in human neurodegenerative diseases.* Mol Neurodegener, 2017. **12**(1): p. 75.
11. Niu, L., et al., *Mutant huntingtin induces iron overload via up-regulating IRP1 in Huntington's disease.* Cell Biosci, 2018. **8**: p. 41.
12. Szlendak, U., K. Bykowska, and A. Lipniacka, *Clinical, Biochemical and Molecular Characteristics of the Main Types of Porphyria.* Adv Clin Exp Med, 2016. **25**(2): p. 361-8.
13. Heinemann, I.U., M. Jahn, and D. Jahn, *The biochemistry of heme biosynthesis.* Arch Biochem Biophys, 2008. **474**(2): p. 238-51.
14. Layer, G., et al., *Structure and function of enzymes in heme biosynthesis.* Protein Sci, 2010. **19**(6): p. 1137-61.
15. Ramanujam, V.S. and K.E. Anderson, *Porphyria Diagnostics-Part 1: A Brief Overview of the Porphyrias.* Curr Protoc Hum Genet, 2015. **86**: p. 17 20 1-17 20 26.
16. Desnick, R.J. and K.H. Astrin, *Congenital erythropoietic porphyria: advances in pathogenesis and treatment.* Br J Haematol, 2002. **117**(4): p. 779-95.
17. Lee, W.-H., W.-C. Tai, and P.-Y. Wu, *Congenital erythropoietic porphyria.* Dermatologica Sinica, 2012. **30**(2): p. 62-65.
18. Nagababu, E. and J.M. Rifkind, *Formation of fluorescent heme degradation products during the oxidation of hemoglobin by hydrogen peroxide.* Biochem Biophys Res Commun, 1998. **247**(3): p. 592-6.
19. Huynh, N., et al., *Glycogen branching enzyme controls cellular iron homeostasis via Iron Regulatory Protein 1 and mitoNEET.* Nat Commun, 2019. **10**(1): p. 5463.

20. Yamanaka, N., K.F. Rewitz, and M.B. O'Connor, *Ecdysone control of developmental transitions: lessons from Drosophila research.* Annu Rev Entomol, 2013. **58**: p. 497-516.
21. Rewitz, K.F., et al., *The insect neuropeptide PTTH activates receptor tyrosine kinase torso to initiate metamorphosis.* Science, 2009. **326**(5958): p. 1403-5.
22. Ou, Q. and K. King-Jones, *What goes up must come down: transcription factors have their say in making ecdysone pulses.* Curr Top Dev Biol, 2013. **103**: p. 35-71.
23. Ou, Q., A. Magico, and K. King-Jones, *Nuclear receptor DHR4 controls the timing of steroid hormone pulses during Drosophila development.* PLoS Biol, 2011. **9**(9): p. e1001160.
24. Kannangara, J.R., C.K. Mirth, and C.G. Warr, *Regulation of ecdysone production in Drosophila by neuropeptides and peptide hormones.* Open Biol, 2021. **11**(2): p. 200373.
25. Huang, X., J.T. Warren, and L.I. Gilbert, *New players in the regulation of ecdysone biosynthesis.* J Genet Genomics, 2008. **35**(1): p. 1-10.
26. Ou, Q., et al., *The Insect Prothoracic Gland as a Model for Steroid Hormone Biosynthesis and Regulation.* Cell Rep, 2016. **16**(1): p. 247-262.
27. Hentze, M.W., et al., *Two to tango: regulation of Mammalian iron metabolism.* Cell, 2010. **142**(1): p. 24-38.
28. Dandekar, T., et al., *Identification of a novel iron-responsive element in murine and human erythroid delta-aminolevulinic acid synthase mRNA.* Embo j, 1991. **10**(7): p. 1903-9.
29. Muckenthaler, M.U., B. Galy, and M.W. Hentze, *Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network.* Annu Rev Nutr, 2008. **28**: p. 197-213.
30. Vashisht, A.A., et al., *Control of iron homeostasis by an iron-regulated ubiquitin ligase.* Science, 2009. **326**(5953): p. 718-21.
31. Kuhn, L.C., *How iron controls iron.* Cell Metab, 2009. **10**(6): p. 439-41.
32. Pantopoulos, K., *Iron metabolism and the IRE/IRP regulatory system: an update.* Ann N Y Acad Sci, 2004. **1012**: p. 1-13.
33. Anderson, C.P., et al., *Mammalian iron metabolism and its control by iron regulatory proteins.* Biochim Biophys Acta, 2012. **1823**(9): p. 1468-83.
34. Eisenstein, R.S., *Iron regulatory proteins and the molecular control of mammalian iron metabolism.* Annu Rev Nutr, 2000. **20**: p. 627-62.
35. Dupuy, J., et al., *Crystal structure of human iron regulatory protein 1 as cytosolic aconitase.* Structure, 2006. **14**(1): p. 129-39.
36. Walden, W.E., et al., *Structure of Dual Function Iron Regulatory Protein 1 Complexed with Ferritin IRE-RNA.* Science, 2006. **314**(5807): p. 1903-1908.
37. Hognon, C., et al., *The Iron Maiden. Cytosolic Aconitase/IRP1 Conformational Transition in the Regulation of Ferritin Translation and Iron Hemostasis.* Biomolecules, 2021. **11**(9).
38. Gray, N.K., et al., *Translational regulation of mammalian and Drosophila citric acid cycle enzymes via iron-responsive elements.* Proc Natl Acad Sci U S A, 1996. **93**(10): p. 4925-30.
39. Dandekar, T. and M.W. Hentze, *Finding the hairpin in the haystack: searching for RNA motifs.* Trends Genet, 1995. **11**(2): p. 45-50.

40. Philpott, C.C., R.D. Klausner, and T.A. Rouault, *The bifunctional iron-responsive element binding protein/cytosolic aconitase: the role of active-site residues in ligand binding and regulation.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 7321-5.
41. Shi, R., et al., *Biogenesis of Iron-Sulfur Clusters and Their Role in DNA Metabolism.* Front Cell Dev Biol, 2021. **9**: p. 735678.
42. Lill, R. and S.A. Freibert, *Mechanisms of Mitochondrial Iron-Sulfur Protein Biogenesis.* Annu Rev Biochem, 2020. **89**: p. 471-499.
43. Kassube, S.A. and N.H. Thoma, *Structural insights into Fe-S protein biogenesis by the CIA targeting complex.* Nat Struct Mol Biol, 2020. **27**(8): p. 735-742.
44. Lill, R., et al., *The role of mitochondria and the CIA machinery in the maturation of cytosolic and nuclear iron-sulfur proteins.* Eur J Cell Biol, 2015. **94**(7-9): p. 280-91.
45. Pondarre, C., et al., *The mitochondrial ATP-binding cassette transporter Abcb7 is essential in mice and participates in cytosolic iron-sulfur cluster biogenesis.* Hum Mol Genet, 2006. **15**(6): p. 953-64.
46. Paul, V.D. and R. Lill, *Biogenesis of cytosolic and nuclear iron-sulfur proteins and their role in genome stability.* Biochim Biophys Acta, 2015. **1853**(6): p. 1528-39.
47. Hausmann, A., et al., *The eukaryotic P loop NTPase Nbp35: an essential component of the cytosolic and nuclear iron-sulfur protein assembly machinery.* Proc Natl Acad Sci U S A, 2005. **102**(9): p. 3266-71.
48. Roy, A., et al., *A novel eukaryotic factor for cytosolic Fe-S cluster assembly.* EMBO J, 2003. **22**(18): p. 4826-35.
49. Stehling, O., et al., *Human Nbp35 is essential for both cytosolic iron-sulfur protein assembly and iron homeostasis.* Mol Cell Biol, 2008. **28**(17): p. 5517-28.
50. Netz, D.J., et al., *A bridging [4Fe-4S] cluster and nucleotide binding are essential for function of the Cfd1-Nbp35 complex as a scaffold in iron-sulfur protein maturation.* J Biol Chem, 2012. **287**(15): p. 12365-78.
51. Banci, L., et al., *Molecular view of an electron transfer process essential for iron-sulfur protein biogenesis.* Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7136-41.
52. Sharma, A.K., et al., *Cytosolic iron-sulfur cluster assembly (CIA) system: factors, mechanism, and relevance to cellular iron regulation.* J Biol Chem, 2010. **285**(35): p. 26745-26751.
53. Seki, M., et al., *IOP1 protein is an external component of the human cytosolic iron-sulfur cluster assembly (CIA) machinery and functions in the MMS19 protein-dependent CIA pathway.* J Biol Chem, 2013. **288**(23): p. 16680-16689.
54. Balk, J., et al., *The essential WD40 protein Cia1 is involved in a late step of cytosolic and nuclear iron-sulfur protein assembly.* Mol Cell Biol, 2005. **25**(24): p. 10833-41.
55. Stehling, O., et al., *MMS19 assembles iron-sulfur proteins required for DNA metabolism and genomic integrity.* Science, 2012. **337**(6091): p. 195-9.
56. Stehling, O., et al., *Human CIA2A-FAM96A and CIA2B-FAM96B integrate iron homeostasis and maturation of different subsets of cytosolic-nuclear iron-sulfur proteins.* Cell Metab, 2013. **18**(2): p. 187-98.
57. Netz, D.J., et al., *Maturation of cytosolic and nuclear iron-sulfur proteins.* Trends Cell Biol, 2014. **24**(5): p. 303-12.
58. Galaris, D., A. Barbouti, and K. Pantopoulos, *Iron homeostasis and oxidative stress: An intimate relationship.* Biochim Biophys Acta Mol Cell Res, 2019. **1866**(12): p. 118535.

59. Wang, J., et al., *Insights on regulation and function of the iron regulatory protein 1 (IRP1).* Hemoglobin, 2008. **32**(1-2): p. 109-15.
60. Cardozo, T. and M. Pagano, *The SCF ubiquitin ligase: insights into a molecular machine.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 739-51.
61. Salahudeen, A.A., et al., *An E3 ligase possessing an iron-responsive hemerythrin domain is a regulator of iron homeostasis.* Science, 2009. **326**(5953): p. 722-6.
62. Hanson, E.S., L.M. Foot, and E.A. Leibold, *Hypoxia post-translationally activates iron-regulatory protein 2.* J Biol Chem, 1999. **274**(8): p. 5047-52.
63. Meyron-Holtz, E.G., M.C. Ghosh, and T.A. Rouault, *Mammalian tissue oxygen levels modulate iron-regulatory protein activities in vivo.* Science, 2004. **306**(5704): p. 2087-90.
64. Bouton, C. and J.C. Drapier, *Iron regulatory proteins as NO signal transducers.* Sci STKE, 2003. **2003**(182): p. pe17.
65. Soum, E. and J.C. Drapier, *Nitric oxide and peroxynitrite promote complete disruption of the [4Fe-4S] cluster of recombinant human iron regulatory protein 1.* J Biol Inorg Chem, 2003. **8**(1-2): p. 226-32.
66. Soum, E., et al., *Peroxynitrite and nitric oxide differently target the iron-sulfur cluster and amino acid residues of human iron regulatory protein 1.* Biochemistry, 2003. **42**(25): p. 7648-54.
67. Haile, D.J., et al., *Cellular regulation of the iron-responsive element binding protein: disassembly of the cubane iron-sulfur cluster results in high-affinity RNA binding.* Proc Natl Acad Sci U S A, 1992. **89**(24): p. 11735-9.
68. Stys, A., et al., *Iron regulatory protein 1 outcompetes iron regulatory protein 2 in regulating cellular iron homeostasis in response to nitric oxide.* J Biol Chem, 2011. **286**(26): p. 22846-54.
69. Pantopoulos, K., G. Weiss, and M.W. Hentze, *Nitric oxide and oxidative stress (H2O2) control mammalian iron metabolism by different pathways.* Mol Cell Biol, 1996. **16**(7): p. 3781-8.
70. Brazzolotto, X., et al., *Human cytoplasmic aconitase (Iron regulatory protein 1) is converted into its [3Fe-4S] form by hydrogen peroxide in vitro but is not activated for iron-responsive element binding.* J Biol Chem, 1999. **274**(31): p. 21625-30.
71. Pantopoulos, K. and M.W. Hentze, *Rapid responses to oxidative stress mediated by iron regulatory protein.* Embo j, 1995. **14**(12): p. 2917-24.
72. Mueller, S., *Iron regulatory protein 1 as a sensor of reactive oxygen species.* Biofactors, 2005. **24**(1-4): p. 171-81.
73. Ferecatu, I., et al., *The diabetes drug target MitoNEET governs a novel trafficking pathway to rebuild an Fe-S cluster into cytosolic aconitase/iron regulatory protein 1.* J Biol Chem, 2014. **289**(41): p. 28070-86.
74. Wu, S., S. Yin, and B. Zhou, *Molecular physiology of iron trafficking in Drosophila melanogaster.* Curr Opin Insect Sci, 2022. **50**: p. 100888.
75. Mandilaras, K., T. Pathmanathan, and F. Missirlis, *Iron absorption in Drosophila melanogaster.* Nutrients, 2013. **5**(5): p. 1622-47.
76. Tang, X. and B. Zhou, *Ferritin is the key to dietary iron absorption and tissue iron detoxification in Drosophila melanogaster.* FASEB J, 2013. **27**(1): p. 288-98.
77. Xiao, G., et al., *Transferrin 1 Functions in Iron Trafficking and Genetically Interacts with Ferritin in Drosophila melanogaster.* Cell Rep, 2019. **26**(3): p. 748-758 e5.

78. Plays, M., S. Muller, and R. Rodriguez, *Chemistry and biology of ferritin.* Metallomics, 2021. **13**(5).
79. Zhang, N., et al., *New Insights into the Role of Ferritin in Iron Homeostasis and Neurodegenerative Diseases.* Mol Neurobiol, 2021. **58**(6): p. 2812-2823.
80. Muckenthaler, M., N.K. Gray, and M.W. Hentze, *IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F.* Mol Cell, 1998. **2**(3): p. 383-8.
81. Muckenthaler, M., et al., *Iron-regulatory protein-1 (IRP-1) is highly conserved in two invertebrate species--characterization of IRP-1 homologues in Drosophila melanogaster and Caenorhabditis elegans.* Eur J Biochem, 1998. **254**(2): p. 230-7.
82. DeRusso, P.A., et al., *Expression of a constitutive mutant of iron regulatory protein 1 abolishes iron homeostasis in mammalian cells.* J Biol Chem, 1995. **270**(26): p. 15451-4.
83. Wang, J. and K. Pantopoulos, *Conditional derepression of ferritin synthesis in cells expressing a constitutive IRP1 mutant.* Mol Cell Biol, 2002. **22**(13): p. 4638-51.
84. Nyayapati, S., et al., *Depletion of cellular iron by bps and ascorbate: effect on toxicity of adriamycin.* Free Radic Biol Med, 1996. **20**(3): p. 319-29.
85. Gu, W., C. Fillebeen, and K. Pantopoulos, *Human IRP1 Translocates to the Nucleus in a Cell-Specific and Iron-Dependent Manner.* Int J Mol Sci, 2022. **23**(18).
86. Groth, A.C., et al., *Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31.* Genetics, 2004. **166**(4): p. 1775-82.
87. Port, F. and S.L. Bullock, *Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs.* Nat Methods, 2016. **13**(10): p. 852-4.
88. Ren, X., et al., *Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9.* Proc Natl Acad Sci U S A, 2013. **110**(47): p. 19012-7.
89. Kami Ahmad, S.H., *CUT&Tag with Drosophila tissues.* protocols.io, 2020.
90. Ahmad, K. and S. Henikoff, *The H3.3K27M oncohistone antagonizes reprogramming in Drosophila.* PLoS Genet, 2021. **17**(7): p. e1009225.
91. Ye Zheng, K.A., Steven Henikoff, *CUT&Tag Data Processing and Analysis Tutorial.* Protocol.io, 2020.
92. Jin, H., et al., *ChIPseqSpikeInFree: a ChIP-seq normalization approach to reveal global changes in histone modifications without spike-in.* Bioinformatics, 2020. **36**(4): p. 1270-1272.
93. Meers, M.P., D. Tenenbaum, and S. Henikoff, *Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling.* Epigenetics Chromatin, 2019. **12**(1): p. 42.
94. Wong, F. and J. Gunawardena, *Gene Regulation in and out of Equilibrium.* Annu Rev Biophys, 2020. **49**: p. 199-226.
95. Kaiser, K. and M. Meisterernst, *The human general co-factors.* Trends Biochem Sci, 1996. **21**(9): p. 342-5.
96. Blossey, R. and H. Schiessel, *The Latest Twists in Chromatin Remodeling.* Biophys J, 2018. **114**(10): p. 2255-2261.
97. Kafkia, E., et al., *Operation of a TCA cycle subnetwork in the mammalian nucleus.* Sci Adv, 2022. **8**(35): p. eabq5206.
98. Xiao, M., et al., *Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors.* Genes Dev, 2012. **26**(12): p. 1326-38.

99. Ule, J. and B.J. Blencowe, *Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution.* Mol Cell, 2019. **76**(2): p. 329-345.
100. Mellacheruvu, D., et al., *The CRAPome: a contaminant repository for affinity purification-mass spectrometry data.* Nat Methods, 2013. **10**(8): p. 730-6.
101. Weber, C.M., J.G. Henikoff, and S. Henikoff, *H2A.Z nucleosomes enriched over active genes are homotypic.* Nat Struct Mol Biol, 2010. **17**(12): p. 1500-7.
102. Akey, C.W. and K. Luger, *Histone chaperones and nucleosome assembly.* Curr Opin Struct Biol, 2003. **13**(1): p. 6-14.
103. Cutter, A.R. and J.J. Hayes, *A brief review of nucleosome structure.* FEBS Lett, 2015. **589**(20 Pt A): p. 2914-22.
104. Horn, P.J. and C.L. Peterson, *Molecular biology. Chromatin higher order folding--wrapping up transcription.* Science, 2002. **297**(5588): p. 1824-7.
105. Kobayashi, W. and H. Kurumizaka, *Structural transition of the nucleosome during chromatin remodeling and transcription.* Curr Opin Struct Biol, 2019. **59**: p. 107-114.
106. Kamakaka, R.T. and S. Biggins, *Histone variants: deviants?* Genes Dev, 2005. **19**(3): p. 295-310.
107. Swaminathan, J., E.M. Baxter, and V.G. Corces, *The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin.* Genes Dev, 2005. **19**(1): p. 65-76.
108. Kaya-Okur, H.S., et al., *CUT&Tag for efficient epigenomic profiling of small samples and single cells.* Nat Commun, 2019. **10**(1): p. 1930.
109. Kaya-Okur, H.S., et al., *Efficient low-cost chromatin profiling with CUT&Tag.* Nat Protoc, 2020. **15**(10): p. 3264-3283.
110. Fosslie, M., et al., *Going low to reach high: Small-scale ChIP-seq maps new terrain.* Wiley Interdiscip Rev Syst Biol Med, 2020. **12**(1): p. e1465.
111. Henikoff, S., et al., *Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation.* Elife, 2020. **9**.
112. Andrews, S., *FastQC: A Quality Control Tool for High Throughput Sequence Data.* 2010.
113. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.
114. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
115. Landt, S.G., et al., *ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia.* Genome Res, 2012. **22**(9): p. 1813-31.
116. Yang, Y., et al., *Leveraging biological replicates to improve analysis in ChIP-seq experiments.* Comput Struct Biotechnol J, 2014. **9**: p. e201401002.
117. *Picard Toolkit.* Broad Institute, GitHub Repository, 2019.
118. Lu, X., et al., *Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure.* Genes Dev, 2009. **23**(4): p. 452-65.
119. Bentsen, M., et al., *ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation.* Nat Commun, 2020. **11**(1): p. 4267.
120. Meers, M.P., D.H. Janssens, and S. Henikoff, *Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded.* Mol Cell, 2019. **75**(3): p. 562-575 e5.

121. Yashar, W.M., et al., *GoPeaks: histone modification peak calling for CUT&Tag.* Genome Biol, 2022. **23**(1): p. 144.
122. Boija, A., et al., *Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.* Cell, 2018. **175**(7): p. 1842-1855 e16.
123. Klemm, S.L., Z. Shipony, and W.J. Greenleaf, *Chromatin accessibility and the regulatory epigenome.* Nat Rev Genet, 2019. **20**(4): p. 207-220.
124. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.
125. Ramírez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.
126. Pick, L., *Hox genes, evo-devo, and the case of the ftz gene.* Chromosoma, 2016. **125**(3): p. 535-51.
127. Guichet, A., et al., *The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors.* Nature, 1997. **385**(6616): p. 548-52.
128. Lawrence, P.A. and L. Pick, *How does the fushi tarazu gene activate engrailed in the Drosophila embryo?* Dev Genet, 1998. **23**(1): p. 28-34.
129. Hou, H.Y., et al., *Stripy Ftz target genes are coordinately regulated by Ftz-F1.* Dev Biol, 2009. **335**(2): p. 442-53.
130. Richards, G., *The Ecdysone Regulatory Cascades in Drosophila*, in *Advances in Developmental Biology (1992)*, P.M. Wassarman, Editor. 1997, Academic Press. p. 81-135.
131. Antoniewski, C., et al., *The ecdysone response enhancer of the Fbp1 gene of Drosophila melanogaster is a direct target for the EcR/USP nuclear receptor.* Mol Cell Biol, 1994. **14**(7): p. 4465-74.
132. Burmester, T., C. Antoniewski, and J.A. Lepesant, *Ecdysone-regulation of synthesis and processing of fat body protein 1, the larval serum protein receptor of Drosophila melanogaster.* Eur J Biochem, 1999. **262**(1): p. 49-55.
133. Arrigo, A.P., *Cellular localization of HSP23 during Drosophila development and following subsequent heat shock.* Dev Biol, 1987. **122**(1): p. 39-48.
134. Azad, P., et al., *Distinct mechanisms underlying tolerance to intermittent and constant hypoxia in Drosophila melanogaster.* PLoS One, 2009. **4**(4): p. e5371.
135. Luo, Y., J. Amin, and R. Voellmy, *Ecdysterone receptor is a sequence-specific transcription factor involved in the developmental regulation of heat shock genes.* Mol Cell Biol, 1991. **11**(7): p. 3660-75.
136. Agawa, Y., et al., *Drosophila Blimp-1 is a transient transcriptional repressor that controls timing of the ecdysone-induced developmental pathway.* Mol Cell Biol, 2007. **27**(24): p. 8739-47.
137. Burglin, T.R. and M. Affolter, *Homeodomain proteins: an update.* Chromosoma, 2016. **125**(3): p. 497-521.
138. Gehring, W.J., M. Affolter, and T. Bürglin, *Homeodomain proteins.* Annu Rev Biochem, 1994. **63**: p. 487-526.
139. de Mendoza, A., et al., *Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages.* Proc Natl Acad Sci U S A, 2013. **110**(50): p. E4858-66.

140. Kaufman, T.C., *A Short History and Description of Drosophila melanogaster Classical Genetics: Chromosome Aberrations, Forward Genetic Screens, and the Nature of Mutations.* Genetics, 2017. **206**(2): p. 665-689.
141. Kingston, R.E. and J.W. Tamkun, *Transcriptional regulation by trithorax-group proteins.* Cold Spring Harb Perspect Biol, 2014. **6**(10): p. a019349.
142. van den Heuvel, M., et al., *Cell patterning in the Drosophila segment: engrailed and wingless antigen distributions in segment polarity mutant embryos.* Dev Suppl, 1993: p. 105-14.
143. Green, R.B., et al., *Drumstick is a zinc finger protein that antagonizes Lines to control patterning and morphogenesis of the Drosophila hindgut.* Development, 2002. **129**(15): p. 3645-56.
144. Uddin, S.N., M. Yano, and R. Murakami, *The drumstick gene acts cell-non-autonomously and triggers specification of the small intestine in the Drosophila hindgut.* Int J Dev Biol, 2011. **55**(10-12): p. 945-52.
145. Morata, G., T. Kornberg, and P.A. Lawrence, *The phenotype of engrailed mutations in the antenna of Drosophila.* Dev Biol, 1983. **99**(1): p. 27-33.
146. Brower, D.L., *Engrailed gene expression in Drosophila imaginal discs.* Embo j, 1986. **5**(10): p. 2649-56.
147. Epper, F. and L. Sánchez, *Effects of engrailed in the genital disc of Drosophila melanogaster.* Dev Biol, 1983. **100**(2): p. 387-98.
148. Ashburner, M. and G. Richards, *Sequential gene activation by ecdysone in polytene chromosomes of Drosophila melanogaster. III. Consequences of ecdysone withdrawal.* Dev Biol, 1976. **54**(2): p. 241-55.
149. Deutsch, J., et al., *Larval fat body-specific gene expression in D. melanogaster.* Dev Genet, 1989. **10**(3): p. 220-31.
150. Telfer, W.H. and J.G. Kunkel, *The function and evolution of insect storage hexamers.* Annu Rev Entomol, 1991. **36**: p. 205-28.
151. Akagi, K. and H. Ueda, *Regulatory mechanisms of ecdysone-inducible Blimp-1 encoding a transcriptional repressor that is important for the prepupal development in Drosophila.* Dev Growth Differ, 2011. **53**(5): p. 697-703.
152. Akagi, K., et al., *A biological timer in the fat body comprising Blimp-1, betaFtz-f1 and Shade regulates pupation timing in Drosophila melanogaster.* Development, 2016. **143**(13): p. 2410-6.
153. Dubrovsky, E.B., G. Dretzen, and E.M. Berger, *The Broad-Complex gene is a tissue-specific modulator of the ecdysone response of the Drosophila hsp23 gene.* Mol Cell Biol, 1996. **16**(11): p. 6542-52.
154. Morrow, G., J.J. Heikkila, and R.M. Tanguay, *Differences in the chaperone-like activities of the four main small heat shock proteins of Drosophila melanogaster.* Cell Stress Chaperones, 2006. **11**(1): p. 51-60.
155. Athilingam, T., et al., *Mechanics of epidermal morphogenesis in the Drosophila pupa.* Semin Cell Dev Biol, 2021. **120**: p. 171-180.
156. Gwóźdź, T., et al., *EcR and Usp, components of the ecdysteroid nuclear receptor complex, exhibit differential distribution of molecular determinants directing subcellular trafficking.* Cell Signal, 2007. **19**(3): p. 490-503.
157. Nieva, C., et al., *Influence of hormone on intracellular localization of the Drosophila melanogaster ecdysteroid receptor (EcR).* Cell Signal, 2007. **19**(12): p. 2582-7.

158. Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID.* Nat Biotechnol, 2018. **36**(9): p. 880-887.
159. Kassis, J.A., J.A. Kennison, and J.W. Tamkun, *Polycomb and Trithorax Group Genes in Drosophila.* Genetics, 2017. **206**(4): p. 1699-1725.
160. Will, C.L. and R. Luhrmann, *Spliceosome structure and function.* Cold Spring Harb Perspect Biol, 2011. **3**(7).
161. Herold, N., et al., *Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes.* Mol Cell Biol, 2009. **29**(1): p. 281-301.
162. Plaschka, C., A.J. Newman, and K. Nagai, *Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast.* Cold Spring Harb Perspect Biol, 2019. **11**(5).
163. Yan, C., R. Wan, and Y. Shi, *Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome.* Cold Spring Harb Perspect Biol, 2019. **11**(1).
164. Cho, K.F., et al., *Proximity labeling in mammalian cells with TurboID and split-TurboID.* Nat Protoc, 2020. **15**(12): p. 3971-3999.
165. Berg Luecke, L. and R.L. Gundry, *Assessment of Streptavidin Bead Binding Capacity to Improve Quality of Streptavidin-based Enrichment Studies.* J Proteome Res, 2021. **20**(2): p. 1153-1164.
166. May, D.G., et al., *Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation.* Cells, 2020. **9**(5).
167. Crosby, M.A., et al., *The trithorax group gene moira encodes a brahma-associated putative chromatin-remodeling factor in Drosophila melanogaster.* Mol Cell Biol, 1999. **19**(2): p. 1159-70.
168. Francis, N.J. and R.E. Kingston, *Mechanisms of transcriptional memory.* Nat Rev Mol Cell Biol, 2001. **2**(6): p. 409-21.
169. Brizuela, B.J. and J.A. Kennison, *The Drosophila homeotic gene moira regulates expression of engrailed and HOM genes in imaginal tissues.* Mech Dev, 1997. **65**(1-2): p. 209-20.
170. Vorobyeva, N.E., et al., *SAYP interacts with DHR3 nuclear receptor and participates in ecdysone-dependent transcription regulation.* Cell Cycle, 2011. **10**(11): p. 1821-7.

# Appendices

## A 1 Homemade Gibson Mastermix

The Homemade Gibson Mastermix used in this study was prepared following https://www.protocols.io/view/homemade-gibson-mastermix-n9xdh7n?step=3.

- **Preparation of 5x isothermal reaction buffer**

Recipe for 4 mL:

| Component | Concentration | Amount | Final concentration |
|---|---|---|---|
| Tris-HCl, pH 7.5 | 1 M | 2 mL | 500 mM |
| $MgCl_2$ | 1 M | 200 µL | 50 mM |
| dATP | 100 mM | 40 µL | 1 mM |
| dCTP | 100 mM | 40 µL | 1 mM |
| dGTP | 100 mM | 40 µL | 1 mM |
| dTTP | 100 mM | 40 µL | 1 mM |
| DTT | 1 M | 200 µL | 50 mM |
| $NAD^+$ | 100 mM | 200 µL | 5 mM |
| PEG-8000 | - | 1 g | 25% |
| $H_2O$ | - | Add ≈1.24 mL $H_2O$ to final volume of 4 mL | |

Tris-HCl, dNTPs, $NAD^+$, $MgCl_2$ and DTT were mixed first, and PEG-8000 was added to the mixture and mixed well. The final volume was adjusted to 4 mL using $H_2O$ and then aliquoted as required, e.g., 100 µL, and stored at -20°C.

- **Preparation of 1.33x Assembly Mastermix**

Recipe for 25 x 15 µL aliquots:

| Component | Concentration | Amount | Final concentration (after adding DNA) |
|---|---|---|---|
| 5x isothermal reaction buffer | 5x | 100 µL | 1x |

| | | | |
|---|---|---|---|
| Taq DNA Ligase | 40 U/µL | 50 µL | 4 U/µL |
| T5 Exonuclease | 1 U/µL | 2 µL | 4 U/mL |
| Q5 Hi-Fi DNA Polymerase | 2 U/µL | 6.25 µL | 25 U/mL |
| $H_2O$ | | 216.75 µL | |
| Total | | 375 µL | |

The mixing procedure was done on ice. 25 x 15 µL aliquots were prepared in PCR tubes and stored at -20℃.

- **Materials**

| Name | Catalogue | Vendor |
|---|---|---|
| beta-Nicotinamide adenine dinucleotide (NAD+) - 0.2 ml | B9007S | New England Biolabs |
| Q5 High-Fidelity DNA Polymerase - 100 units | M0491S | New England Biolabs |
| Taq DNA Ligase - 2,000 units | M0208S | New England Biolabs |
| T5 Exonuclease - 5,000 units | M0363L | New England Biolabs |
| PEG-8000 | | |
| DTT (Dithiothreitol) (> 99% pure) Protease free | DTT | Gold Biotechnology |
| Deoxynucleotide Solution Set - 25 umol of each | N0446S | New England Biolabs |

## A 2 Reagent and medium recipes for ultra-competent cell preparation

- **PIPES (0.5M, pH 6.7)**

15.1 g of PIPES (Sigma-Aldrich P6757, low solubility in acidic solution) was added into 80 mL of Mili-Q water (MBSU, sterile), and the pH was adjusted to 6.7 with 5 M KOH (15 ~ 17.1 mL). The final volume was adjusted to 100 mL using Mili-Q water. The solution was filter-sterile (0.45 μm-pore size) (Sigma-Aldrich F2161), aliquoted and stored at -20℃.

- **Inoue transformation buffer (for 1L)**

| Reagent | Amount | Final concentration |
|---|---|---|
| $MnCl_2.4H_2O$ (Sigma M3634) | 10.88 g | 55 mM |
| $CaCl_2.2H_2O$ (Sigma C3881) | 2.20 g | 15 mM |
| KCl (Sigma P9541) | 18.65 g | 250 mM |
| PIPES (0.5M, pH 6.7) | 20 mL | |
| Mili-Q $H_2O$ | To 1 L | |

The buffer was filter-sterile (0.45 μm-pore size), aliquoted and stored at -20℃.

- **LB (Luria-Bertani) medium**

10 g of tryptone, 5 g of yeast extract, and 10 g of NaCl were added to 950 mL of Mili-Q water and mixed well. The pH was adjusted to 7.0 with 5 N NaOH (~0.2 mL), and then the final volume was adjusted to 1 L using Mili-Q water. The solution was autoclaved for 20 min at 15 psi (1.05 kg/cm2) on liquid cycle.

- **LB agar plates**

The liquid media was prepared according to the provided recipe and 15 g of agar was added to 1 L of media before autoclaving. The media was sterilized through autoclaving for 20 minutes at 15 psi (1.05 kg/cm2) on the liquid cycle. After autoclaving, the medium was gently swirled to ensure proper distribution of the melted agar or agarose throughout the solution. Before adding thermolabile substances, such as antibiotics, the medium was allowed to cool to a temperature

range of 50°C-60°C. Thorough mixing of the medium was achieved by swirling to prevent the formation of air bubbles. The medium was directly poured from the flask onto each 90-mm plate, with an allocation of approximately 30-35 mL of medium per plate. To eliminate air bubbles from the medium in the plate, the surface was flamed using a burner before the agar or agarose solidified. Once the medium had completely solidified, the plates were inverted and stored at 4°C until needed. In this study, ampicillin was used.

- **SOB medium**

To prepare the liquid media, 20 g of tryptone, 5 g of yeast extract, and 0.5 g of NaCl were added to 950 mL of Mili-Q water and shaken until complete dissolution of the solutes. Subsequently, 10 mL of a 250 mM solution of KCl (prepared by dissolving 1.86 g of KCl in 100 mL of deionized H2O) was added. The pH of the medium was adjusted to 7.0 with approximately 0.2 mL of 5 N NaOH, and the volume was adjusted to 1 L with Mili-Q $H_2O$. The solution was sterilized by autoclaving for 20 minutes at 15 psi (1.05 kg/cm$^2$) using the liquid cycle. Immediately prior to use, 5 mL of a sterile 2 M MgCl2 solution (prepared by dissolving 19 g of MgCl2 in 90 mL of deionized H2O and adjusting the volume to 100 mL with deionized H2O, followed by autoclaving for 20 minutes at 15 psi [1.05 kg/cm2] on the liquid cycle) was added to the medium.

- **SOC medium**

SOC medium is identical to SOB medium, except that it contains 20 mM glucose. After autoclaving the SOB medium (1 L), it was cooled to 60°C or below. Following this, 20 mL of a sterile 1 M glucose solution can be added. To prepare the 1 M glucose solution, 18 g of glucose was dissolved in 90 mL of deionized H2O. Once the sugar was fully dissolved, the volume was adjusted to 100 mL with deionized H2O, and the solution was sterilized by passing it through a 0.22-µm filter.

## A 3 List of primers used for S2 cell constructs

| Primer name | Sequence (5′ to 3′) |
|---|---|
| Generation of *pAFW-C-3xFlag* | |
| pAFW C-3xFlag F1 | ACCGGTAAGCTAGCAGGATCTTTGTGAAGGAACC |
| pAFW/pAMW BB R1 | TCAGCTCACTCAAAGGCGGTAATAC |
| pAFW/pAMW BB F2 | CGCCTTTGAGTGAGCTGATACCG |
| pAFW C-3xFlag R2 | CTGCTAGCTTACCGGTGCTTGTCATCGTCATCC |
| Generation of *pAFW IPR1A-C-3xFlag* | |
| pAFW BB F1 | ATGGACTACAAAGACCATGACGGTG |
| pAFW/pAMW BB R2 | GGTGGCCTCTAGACAGGCCTCGATATCTG |
| IRP1A F1-overlap with pAFW | CTGTCTAGAGGCCACCATGGAAATGTCCGGCTCCGGCG CCAAT |
| IRP1A R1-overlap with 3xFlag | CATGGTCTTTGTAGTCCATATCCAGCATTTTGCGTATCA TATAGTTC |
| Generation of *pAMW H2Av-C-6xMyc* | |
| pAMW BB F1 | TTGTACTAGGGAACCTTACTTCTGTGGTGTGACATAATT GG |
| H2Av F1-overlap with pAMW BB | CTGTCTAGAGGCCACCATGGCAATGGCTGGCGGTAAAG CAGGCAAG |
| H2Av R1-overlap with 6xMyc | GGCGGAGCTTCTCATGTAGGCCTGCGACAGAATGACG |
| 6xMyc F1 | ATGAGAAGCTCCGCCACCATG |
| 6xMyc R1-overlap with pAMW F1 | TAAGGTTCCCTAGTACAAACTTGTATACCGGTGATTC |
| Generation of *pAMW N-6xMyc-EGFP* | |
| pAFW/pAMW BB FP | GGAACCTTACTTCTGTGGTGTGACATAATTGG |
| pAMW R2-include 6xMyc | GCCGTACAAACTTGTATACCGGTGATTC |
| EGFP F1-overlap with pAMW, without ATG | GTATACAAGTTTGTACGGCGTGAGCAAGGGCGAGGAG CTG |
| EGFP R1-overlap with pAMW | CACAGAAGTAAGGTTCCTCACGTGGACCGGTGCTTGTA C |
| Generation of *pSTABLE2 IRP1A-C-3xFlag-H2Av-C-6xMyc* | |
| pSTABLE2 F1 | CTCGAGGGATCCAGACATGATAAGATAC |
| pSTABLE2 R1 | GAATTCCACCACACTGGACTAGTAG |

| | |
|---|---|
| IRP1A-3xFlag F1-overlap with pSTABLE2 | TCCAGTGTGGTGGAATTCGCCACCATGGAAATGTCCGGCTCCG |
| IRP1A-3xFlag R1-overlap with T2A | CTTCGACATCTCCGCAAGTCAGTAGGCTGCCGCGTCCTTCCCGGTGCTTGTCATCGTCATCCTTG |
| H2Av-6xMyc F1-overlap with T2A | TTGCGGAGATGTCGAAGAGAACCCTGGCCCTGCCACCATGGCAATGGCTGGCGGTAAAG |
| H2Av-6xMyc R1-overlap with pSTABLE2 | TGTCTGGATCCCTCGAGCTAGTACAAACTTGTATACCGGTGATTC |
| Generation of *pSTABLE2 IRP1A-C-3xFlag-N-6xMyc-EGFP* | |
| 6xMyc-EGFP F1-overlap with T2A | TTGCGGAGATGTCGAAGAGAACCCTGGCCCTAGAAGCTCCGCCACCATGG |
| 6xMyc-EGFP R1-overlap with pSTABLE2 | TGTCTGGATCCCTCGAGTCACGTGGACCGGTGCTTG |

## A 4 List of primers designed for transgenic line construction

| Primer name | Sequence (5′ to 3′) |
|---|---|
| Generation of *pAFW-IRP1A$^{2R2Q}$* | |
| pAc5 BB F | CGTTCAGCTTTCTTGTACAAAGTGGTGACG |
| pAc5 BB R | GGTGGCTTTTTTGTACAAACTTGTATACCGG |
| IRP1A-pAc5 F | TTGTACAAAAAAGCCACCATGTCCGGCTCCGGCGCCAATC |
| IRP1A-2R2Q R1 | TCTGACCCTCGAAATTCTGATTGCCCGACAGG |
| IRP1A-2R2Q F1 | AATTTCGAGGGTCAGATCCATCCCAATACTAG |
| IRP1A-pAc5 R | GTACAAGAAAGCTGAACGCTAATCCAGCATTTTGCGTATC |
| Generation of *pAFW-IRP1A$^{3R3Q}$* | |
| IRP1A-R793Q R1 | GGCCCAATCCTGTGAGCTGCCACTGCCGTAG |
| IRP1A-R793Q F1 | TCACAGGATTGGGCCGCCAAGGGGCCATTTC |
| Generation of *pBID-UASC-FG-IRP1A$^{3R3Q}$* | |
| pBID-FG BB F1 | GACGTAAGCTAGAGGATCTTTG |
| pBID-FG BB R1 | GAAAGAGATAGCGGACGC |
| pBID-FG BB F2 | CTGCGTCCGCTATCTCTTTC |
| pBID-FG BB R2 | GGAGCTCCGAATTCACTGC |
| KZ-3xFlag-pBID F | GTGAATTCGGAGCTCCGCCACCATGGACTACAAAGAC |
| pAc5-pBID R | GATCCTCTAGCTTACGTCCGTCACCACTTTGTACAAGAAAGC |
| Generation of *pAFW-IRP1B$^{2R2Q}$* | |
| IRP1B-pAc5 F | TTGTACAAAAAAGCCACCATGTCAGGCGCCAATCCCTTCGC |
| IRP1B-2R2Q R1 | TCTGACCCTCGAAGTTCTGGTTTCCGGACAAAACTCCAG |
| IRP1B-2R2Q F1 | AACTTCGAGGGTCAGATCCATCCCAACACCAGGGCCAAC |
| IRP1B-pAc5 R | GTACAAGAAAGCTGAACGTTAAGAGAGCATTTTGCGAATCAT GTAGTTG |
| Generation of *pAFW-IRP1B$^{3R3Q}$* | |
| IRP1B-R790Q R1 | CCAATCCTGAGAGCTTCCGCTGCCGTAGTC |
| IRP1B-R790Q F1 | AAGCTCTCAGGATTGGGCCGCCAAGGGTCCCTTCCTG |

## A 5 List of primers designed for TurboID line construction

| Primer name | Sequence (5′ to 3′) |
|---|---|
| Generation of *pTurboID-IRP1A* | |
| Turbo BB F1 | ACCCAGCTTTCTTGTACAAAGTGGTG |
| Turbo BB R1 | GTATTTCCCCTTCGAACATCCCCAC |
| Turbo BB F2 | CCAAAGTCTACTTGTGGGGATGTTCGAAGG |
| Turbo BB R2 | CTGCAGCTTTTCGGCAGACCGCAG |
| IRP1A F-Turbo | TGCCGAAAAGCTGCAGGCCACCATGGAAATGTCCGG |
| IRP1A R-Turbo | GTACAAGAAAGCTGGGTTTAATCCAGCATTTTGCGTATCATATAGTTC |
| Generation of *pTurboID-IRP1B* | |
| IRP1B F-Turbo | TGCCGAAAAGCTGCAGGCCACCATGGAAATGTCAGGCGCCAATCCCTTC |
| IRP1B R-Turbo | GTACAAGAAAGCTGGGTTTAAGAGAGCATTTTGCGAATCATGTAGTTG |

## A 6 List of primers designed for CRISPR line construction

| Primer name | Sequence (5′ to 3′) |
|---|---|
| Generation of *pCFD5-IRP1A-KI* | |
| pCFD5 IA gRNA-1 FP | TTCGATTCCCGGCCGATGCAACTTCGTGACTATTCCCTTGGTTTTAGAGCTAGAAATAGC |
| pCFD5 IA gRNA-2 RP | CTATTTCTAGCTCTAAAACAACGCCTGCTGGTGAATCATTGCACCAGCCGGGAATCGAAC |
| Generation of *pDsRed-left arm-IA* | |
| IA Left arm F | GCTGAAGCAGGTGAATTCCATCGCTGAAGCAGGAAACC |
| IA Left arm R | CGACAAGGGAATAGTCACGAAGTTGTCGAGTGGCC |
| IA-3F donor F | GTGACTATTCCCTTGTCGATCAGGATCCCCAATACACGG |
| IA-3F donor R | CGCTAGCATGCAAGAATTCCATTTTTAGCAGGACTGGTTTGG |
| Generation of *pDsRed-IRP1A$^{WT}$-3F* | |
| pDsRed-IA-left arm BB F | CCGTCAATCGAGTTCAAGGGC |
| IRP1A R1-overlap 3F | CATGGTCTTTGTAGTCCATATCCAGCATTTTGCGTATCATATAGTTC |
| pAFW BB F1 | ATGGACTACAAAGACCATGACGGTG |
| 3xFlag R | CTACCGGTGCTTGTCATCGTCATCC |
| IA-3'UTR F | TGACAAGCACCGGTAGTTAGTGCGTTCGTTGAC |
| DsRed R | AAGAGCCTCGAGCTGCAGAAGGCCTAG |
| IA Right arm F | CAGCTCGAGGCTCTTATTAACGCCTGCTGGTGAATCATCGAC |
| IA Right arm R | CTTGAACTCGATTGACGGGGTGGTGTTGGTGATGTTGC |
| Generation of *pDsRed-IRP1A$^{C450S}$-3F* | |
| pDsRed-IA-left arm BB F | CCGTCAATCGAGTTCAAGGGC |
| pDsRed-IA-left arm BB R | AGGTGATGGCGGCAATCACAAC |
| C450S donor F | ATTGCCGCCATCACCTCGTCCACGAACACTTCGAATCC |
| IA Right arm R | CTTGAACTCGATTGACGGGGTGGTGTTGGTGATGTTGC |
| Generation of *pDsRed-IRP1A$^{3R3Q}$-3F* | |
| IRP1A-2R2Q R1 | TCTGACCCTCGAAATTCTGATTGCCCGACAGG |
| IRP1A-2R2Q F1 | AATTTCGAGGGTCAGATCCATCCCAATACTAG |

| | |
|---|---|
| IRP1A-R793Q R1 | GGCCCAATCCTGTGAGCTGCCACTGCCGTAG |
| IRP1A-R793Q F1 | TCACAGGATTGGGCCGCCAAGGGGCCATTTC |
| Generation of *pCFD5-IRP1A-KO* | |
| IA-KO gRNA-1 FP | TTCGATTCCCGGCCGATGCAAACGAAATTAGCATGTAGGAGT TTTAGAGCTAGAAATAGC |
| Generation of *pDsRed-IRP1A-KO* | |
| IA-KO left arm F | GCTGAAGCAGGTGAATTCACAACTTTCCGCAGTCCAAC |
| IA-KO left arm R | CCGCTAGCATGCAAGAATTCGGGTCCTACATGCTAATTTCG |

## A 7 Reagent recipes for co-IP and MS using S2 cells or *Drosophila* tissues

- **2x Buffer G:**

| Component | Stock concentration | Final concentration | Amount to be added |
|---|---|---|---|
| Na-HEPES, pH 7.5 | 1.0M | 50mM | 5.0mL |
| NaCl | 5.0M | 150mM | 3.0mL |
| EDTA | 0.5M | 1mM | 0.2mL |
| Mili-Q water | | | To 100mL |

Filter-sterilized and storeed at 4°C.

- **1x Lysis Buffer for S2 cells (prepare fresh):**

| Component | Stock concentration | Final concentration | Amount |
|---|---|---|---|
| Tris-HCl pH 7.5 | 1M | 50 mM | 500 µL |
| NaCl | 1 M | 250 mM | 2.5 mL |
| EDTA | 250 mM | 5 mM | 200 µL |
| Triton X-100 | | 1% | 0.1 mL |
| Protease Inhibitor cocktail | 7 X | 1X | 1.42 mL |
| Mili-Q water | | | To 10 mL |

- **1x Lysis Buffer for tissues (prepare fresh):**

| Component | Stock concentration | Final concentration | Amount |
|---|---|---|---|
| 2x Buffer G | 2x | 1x | 10.0 mL |
| Triton X-100 | | 1% | 0.2 mL |
| Glycerol | 50% | 10% | 4.0 mL |
| Protease Inhibitor cocktail | 25x | 1x | 0.8 mL |
| MiliQ water | | | To 20 mL |

- **Wash Buffer 1:**

| Component | Stock concentration | Final concentration | Amount to be added |
|---|---|---|---|
| 2x Buffer G | 2x | 1x | 5.0mL |
| Triton X-100 | | 0.1% | 0.01mL |
| Glycerol | 50% | 5% | 1.0mL |
| Mili-Q water | | | To 10mL |

- **Wash Buffer 2:**

| Component | Stock concentration | Final concentration | Amount to be added |
|---|---|---|---|
| 2x Buffer G | 2x | 1x | 5.0mL |
| Glycerol | 50% | 5% | 1.0mL |
| Mili-Q water | | | To 10mL |

## A 8 SDS-PAGE gel recipe

- **1 M Tris (pH 6.8), 200 mL**

To prepare, 24.22 g of Tris base and 16 mL of concentrated HCl were added to 100 mL of Mili-Q $H_2O$. The pH was adjusted to 6.8 using concentrated HCl and the final volume was brought to 200 mL with Mili-Q $H_2O$. The solution was autoclaved for sterilization and stored at 4°C.

- **1.5 M Tris (pH 8.8), 200 mL**

36.342 g of Tris base and 3 mL of concentrated HCl were added to 100 mL Mili-Q $H_2O$. The pH was adjusted to 8.8 with concentrated HCl and the final volume was brought to 200 mL with Mili-Q $H_2O$. The solution was autoclaved and stored at 4°C.

- **10% SDS**

10 g of SDS was added in 90 mL Mili-Q $H_2O$. The pH was adjusted to 7.2 with concentrated HCl and the final volume was brought to 100 mL with Mili-Q $H_2O$. The solution was stored at room temperature.

- **10% AP solution**

1g of AP was added in 10 mL Mili-Q $H_2O$ and mixed well. The solution was stored at 4°C for up to one month.

- **Prepare SDS-PAGE gel**

| **Separating gel** | **8 mL (one gel)** | | | | |
|---|---|---|---|---|---|
| **Acylamide percentage** | **6%** | **8%** | **10%** | **12%** | **15%** |
| ddH2O | 4.6 mL | 4.2 mL | 3.8 mL | 3.4 mL | 2.8 mL |
| 40% Acrylamide/Bis-acrylamide | 1.2 mL | 1.6 mL | 2 mL | 2.4 mL | 3 mL |
| 1.5 M Tris(pH=8.8) | 2 mL | 2 mL | 2 mL | 2 mL | 2 mL |
| 10% SDS | 80 µL | 80 µL | 80 µL | 80 µL | 80 µL |
| TEMED | 8 µL | 8 µL | 8 µL | 8 µL | 8 µL |

| | | | | | |
|---|---|---|---|---|---|
| 10% AP | 80 µL | 80 µL | 80 µL | 80 µL | 80 µL |

| Stacking gel (4%) | 5 mL |
|---|---|
| Mili-Q water | 3.725 mL |
| 40% Acrylamide/Bis-acrylamide | 0.5 mL |
| 1 M Tris-HCl, pH 6.8 | 0.625 mL |
| 10% SDS | 50 µL |
| TEMED | 5 µL |
| 10% AP | 50 µL |

| Protein size | Gel acrylamide percentage |
|---|---|
| 4–40 kDa | 20% |
| 12–45 kDa | 15% |
| 10–70 kDa | 12.50% |
| 15–100 kDa | 10% |
| 25–200 kDa | 8% |

| Separating gel | 10 mL | | | | |
|---|---|---|---|---|---|
| Acylamide percentage | 6% | 8% | 10% | 12% | 15% |
| Mili-Q water | 5.2 mL | 4.6 mL | 3.8 mL | 3.2 mL | 2.2 mL |
| 30% Acrylamide/Bis-acrylamide | 2 mL | 2.6 mL | 3.4 mL | 4 mL | 5 mL |
| 1.5 M Tris(pH=8.8) | 2.6 mL | 2.6 mL | 2.6 mL | 2.6 mL | 2.6 mL |
| 10% SDS | 100 µL | 100 µL | 100 µL | 100 µL | 100 µL |
| TEMED | 10 µL | 10 µL | 10 µL | 10 µL | 10 µL |
| 10% AP | 100 µL | 100 µL | 100 µL | 100 µL | 100 µL |

| Stacking gel (4%) | 5 mL |
|---|---|
| Mili-Q water | 3.6 mL |
| 30% Acrylamide/Bis-acrylamide | 0.67 mL |
| 1 M Tris-HCl, pH 6.8 | 0.625 mL |
| 10% SDS | 50 µL |
| TEMED | 5 µL |
| 10% AP | 50 µL |

## A 9 Reagent recipes and PCR primers for CUT & Tag

**Before starting the following are used as stock solutions that can be prepared and stored in advance:**

**1. 5% Digitonin/DMSO**

50 mg of digitonin was dissolved in 1 mL DMSO. The solution was stored at RT for up to 1 week or frozen at -20°C. *Caution: Digitonin is toxic, and care should be taken especially when weighing out the powder. Use full PPE including a mask, lab coat and gloves while handling any amount of digitonin. Be aware that DMSO can penetrate through the skin*.

**2. Bead Activation Buffer (50 mL)** [211 µL/reaction]

48.4 mL $H_2O$

1 mL 1 M HEPES-KOH pH7.9 (final 20 mM)

500 µL 1 M KCl, final 10 mM

50 µL 1 M $CaCl_2$ (final 1 mM)

50 µL 1 M $MnCl_2$ (final 1 mM)

*Filter-sterilized and stored at 4 ˚C for up to 6 months.*

**3. PBT buffer (50 mL):** 0.1% (v/v) Tween 20 in 1X PBS

50 µL Tween 20 in 50 mL 1X PBS

**4. 2 M Spermidine (1 mL):** 0.50926 g in 1 mL water

**5. 30% BSA (1 mL):** 0.3 g BSA in 1 mL water

**6. 100 mM $MgCl_2$ (50 mL):** 45 mL $H_2O$ with 5 mL 1 M $MgCl_2$

**7. 1% SDS (10 mL):** 0.1 g SDS in 10 mL water

**8. Wash150 Buffer (50 mL):** use to prepare Digitonin150 Buffer

47.5 mL water

1 mL 1 M HEPES pH 7.5 (final 20 mM)

1.5 mL 5 M NaCl (final 150 mM)

12.5 µL 2 M Spermidine (final 0.5 mM)

1 **large** Roche cOmplete EDTA-free tablet

*Filter-sterilized and stored at 4 ˚C for up to 1 week.*

**9. Digitonin150 Buffer (5 mL)** [450 µL/reaction]

5 mL Wash150 Buffer

50 µL 5% Digitonin/DMSO (final 0.05%)

*Prepared fresh each day and stored at 4 ˚C.*

**10. Antibody150-1 Buffer, with EDTA (1 mL)** [50 µL/reaction]

963 µL Digitonin150 Buffer
4 µL 0.5 M EDTA (final 2 mM)
33 µL 30% BSA (final 1%)
*Prepared fresh each day and stored at 4 ˚C.*

**11. Antibody150-2 Buffer, without EDTA (1 mL)** [50 µL/reaction]

967 µL Digitonin150 Buffer
33 µL 30% BSA (final 1%)
*Prepared fresh each day and stored at 4 ˚C.*

**12. Wash300 Buffer (10 mL):** use to prepare Digitonin300 and Tagmentation Buffers

9.2 mL water
200 µL 1 M HEPES pH 7.5 (final 20 mM)
600 µL 5 M NaCl (final 300 mM)
2.5 µL 2 M Spermidine (final 0.5 mM)
1 mini Roche cOmplete EDTA-free tablet
*Filter-sterilized and stored at 4 ˚C for up to 1 week.*

**13. Digitonin300 Buffer (5 mL)** [450 µL/reaction]

5 mL Wash300 Buffer
50 µL 5% Digitonin/DMSO (final 0.05%)
*Prepared fresh each day and stored at 4 ˚C.*

**14. Tagmentation Buffer (1 mL)** [50 µL/reaction]

900 µL Digitonin300 Buffer

100 µL 100 mM $MgCl_2$ (Final 10 mM)

*Stored at 4 ˚C for up to 1 week.*

15. **TAPS-EDTA Buffer (10 mL)** [50 µL/reaction]

9.5 mL water
500 µL 0.2 M TAPS Buffer, pH8.5 (Final 10 mM)
4 µL 0.5 M EDTA (Final 0.2 mM)
*Stored at room temperature for up to 6 months.*

16. **SDS Release Buffer (1 mL)** [5 µL/reaction]

850 µL water
50 µL 0.2 M TAPS buffer pH8.5 (Final 10 mM)
100 µL 1% SDS (Final 0.1%)
*Stored at room temperature for up to 6 months.*

17. **SDS Quench Buffer (10 mL)** [15 µL/reaction]

10 mL Molecular grade water
67 µL Triton-X 100 (Final 0.67%)
*Stored at room temperature for up to 6 months.*

18. **70% ethanol** [1.4 mL/reaction]**:** Freshly made

19. **10 mM Tris-HCl pH 8 (1 mL)** [25 µL/reaction]**:**

990 µL RNase/DNase free water, 10 µL 1 M Tris-HCl pH 8

***Spermidine** is added to compensate for the removal of Mg2+ from the buffer. Mg2+ can cause DNA degradation and is typically omitted from CUT & Tag and CUT&RUN buffers.
****Digitonin 5% stock solution** should be prepared in DMSO. Aliquots can be stored at -20ºC for 6 months. Rather, digitonin helps prevent the nuclei-conjugated beads from precipitating/clumping and forming a thin film on tubes. Digitonin also makes the protocol compatible with cells, although the use of cells is not recommended.

**Tn5-adapter complex formation (**Epicypher EP151017 is charged with Illumina adapters and ready to be used immediately in CUT & Tag**)**

1. Dilute oligonucleotides in annealing buffer (10mM Tris pH8, 50mM NaCl, 1 mM EDTA)
   - 200 µM of ME-A in annealing buffer;
   - 200 µM of ME-B in annealing buffer;
   - 200 µM of ME-Reverse in annealing buffer

Mosaic end_Adapter A (ME-A): 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′

Mosaic end_Adapter B (ME-B): 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′

Mosaic end_reverse (ME-Reverse): 5′-[phos]CTGTCTCTTATACACATCT-3′

**2.** Pair and anneal the oligos

- Mix equal volume of ME-A and ME-Reverse in a 1.5 mL tube, resulting in 100 μM of annealed product.
- Mix equal volume of ME-B and ME-Reverse in another 1.5 mL tube, resulting in 100 μM of annealed product.
- Place the tubes in a 90–95 °C hot block and leave for 3–5 min, then remove the hot block from the heat source, allowing slow cooling to room temperature (~45 min).
- Mix 16 μL of 100 μM of **ME-A+ME-Reverse**, **ME-B+ME-Reverse**, and 100 μL of 5.5 μM pA-Tn5 fusion protein.
- Incubate the mixture on a rotating platform for 1 h at room temperature and then store at −20 °C for up to 1 year.

**Materials**

**Beads and enzymes**

1. Agencourt Ampure XP beads for PCR DNA purification, Beckman Coulter, Catalog number A63880
2. CUTANA Concanavalin A Conjugated Paramagnetic Beads, 550 μL, MJS BioLynx, Inc. Catalog Code EP211401
3. Cutana pAG-Tn5 ChIC/CUT & Tag, 50 Rxns, MJS BioLynx, Inc. Catalog Code EP151017, CAD 952 (Epicypher EP151017 is charged with Illumina adapters and ready to be used immediately in CUT & Tag)
4. Proteinase K, recombinant, PCR grade, Catalog number: EO0491 thermofisher
5. NEBNext® High-Fidelity 2X PCR Master Mix, M0541S,

**Antibodies**

1. Guinea Pig anti-Rabbit IgG (Heavy & Light Chain) Antibody – Preadsorbed, Catalog No. ABIN101961
2. Anti-Mouse Secondary Antibody for CUTANA ChIC/CUT & Tag Workflows, 50 Reactions, MJS BioLynx, Catalog number EP130048
3. Rabbit IgG Antibody, CUTANA CUT&RUN Negative Control, MJS BioLynx, Catalog number EP130042
4. positive control antibody, Tri-Methyl-Histone H3 (Lys27) (H3K[1]27me3, C36B11) Rabbit mAb, NEB 9733T

**Chemicals**

1. Large, cOmplete™, EDTA-free Protease Inhibitor Cocktail, sigma, 11873580001
2. Magnesium chloride solution, sigma, M1028-100ML, 1.00 M±0.01 M
3. Digitonin, sigma, 300410-250MG
4. 1 M Manganese Chloride (MnCl2), by Sigma Aldrich, Catalog number M1787-10X1ML
5. 1 M Calcium Chloride (CaCl2), sigma, 21115-100ML
6. Spermidine, Sigma Aldrich, Catalog number S2501

[1]

7. phase-lock tubes, Qiagen MaXtract High Density (200 x 1.5 ml) Cat. No. / ID: 129046
8. PCI (Phenol:Chloroform:Isoamyl Alcohol): Thermofisher, Catalog number: 15593031
9. 0.2 M TAPS Buffer, pH8.5, Thermofisher, Catalog number: J63268.AE

**PCR primers**

Synthesize Ad1.1 as the universal i5 primer, synthesize Ad2.1–Ad2.30 unique i7 primers.

| Custom Barcodes **Adapter 1 (index i5)** IDs | **Barcodes** | **Primer Sequence** |
|---|---|---|
| Ad1.noMX | | AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG |
| Ad1.1 | TAGATCGC | AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.2 | CTCTCTAT | AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.3 | TATCCTCT | AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.4 | AGAGTAGA | AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.5 | GTAAGGAG | AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.6 | ACTGCATA | AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.7 | AAGGAGTA | AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.8 | CTAAGCCT | AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.9 | TGGAAATC | AATGATACGGCGACCACCGAGATCTACACTGGAAATCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.10 | AACATGAT | AATGATACGGCGACCACCGAGATCTACACAACATGATTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.11 | TGATGAAA | AATGATACGGCGACCACCGAGATCTACACTGATGAAATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.12 | GTCGGACT | AATGATACGGCGACCACCGAGATCTACACGTCGGACTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.13 | TTTCTAGC | AATGATACGGCGACCACCGAGATCTACACTTTCTAGCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.14 | TAACCAAG | AATGATACGGCGACCACCGAGATCTACACTAACCAAGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.15 | GTGTATCG | AATGATACGGCGACCACCGAGATCTACACGTGTATCGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.16 | TCCATCAA | AATGATACGGCGACCACCGAGATCTACACTCCATCAATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.17 | TTCGTGCA | AATGATACGGCGACCACCGAGATCTACACTTCGTGCATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.18 | AGGTTGCC | AATGATACGGCGACCACCGAGATCTACACAGGTTGCCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.19 | CCTTATGT | AATGATACGGCGACCACCGAGATCTACACCCTTATGTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.20 | CAGCAACG | AATGATACGGCGACCACCGAGATCTACACCAGCAACGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.21 | GGTTCAAT | AATGATACGGCGACCACCGAGATCTACACGGTTCAATTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.22 | ACATTCGT | AATGATACGGCGACCACCGAGATCTACACACATTCGTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.23 | GATTCCCA | AATGATACGGCGACCACCGAGATCTACACGATTCCCATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.24 | CGGACTGC | AATGATACGGCGACCACCGAGATCTACACCGGACTGCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.25 | AGCCGTTC | AATGATACGGCGACCACCGAGATCTACACAGCCGTTCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.26 | ATTGGGTC | AATGATACGGCGACCACCGAGATCTACACATTGGGTCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.27 | TGCATACT | AATGATACGGCGACCACCGAGATCTACACTGCATACTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.28 | GGGCTTGG | AATGATACGGCGACCACCGAGATCTACACGGGCTTGGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.29 | GACGTGGC | AATGATACGGCGACCACCGAGATCTACACGACGTGGCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.30 | GCAAATTT | AATGATACGGCGACCACCGAGATCTACACGCAAATTTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.31 | GCAGCCTC | AATGATACGGCGACCACCGAGATCTACACGCAGCCTCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.32 | TCCGAGTT | AATGATACGGCGACCACCGAGATCTACACTCCGAGTTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.33 | GCATTAAG | AATGATACGGCGACCACCGAGATCTACACGCATTAAGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.34 | ACGATAAC | AATGATACGGCGACCACCGAGATCTACACACGATAACTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.35 | CCTGCGGG | AATGATACGGCGACCACCGAGATCTACACCCTGCGGGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.36 | TGATTGTT | AATGATACGGCGACCACCGAGATCTACACTGATTGTTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.37 | GGCACGGA | AATGATACGGCGACCACCGAGATCTACACGGCACGGATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.38 | GATCATTC | AATGATACGGCGACCACCGAGATCTACACGATCATTCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.39 | ATGGTCAT | AATGATACGGCGACCACCGAGATCTACACATGGTCATTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.40 | CGTACCAA | AATGATACGGCGACCACCGAGATCTACACCGTACCAATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.41 | CCAGTTTA | AATGATACGGCGACCACCGAGATCTACACCCAGTTTATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.42 | ACCGGCCC | AATGATACGGCGACCACCGAGATCTACACACCGGCCCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.43 | CTAGAAGT | AATGATACGGCGACCACCGAGATCTACACCTAGAAGTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.44 | CGCCAGAT | AATGATACGGCGACCACCGAGATCTACACCGCCAGATTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.45 | TCACATGG | AATGATACGGCGACCACCGAGATCTACACTCACATGGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.46 | GAACTCGA | AATGATACGGCGACCACCGAGATCTACACGAACTCGATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.47 | CCACCGTT | AATGATACGGCGACCACCGAGATCTACACCCACCGTTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.48 | TAAGTTAC | AATGATACGGCGACCACCGAGATCTACACTAAGTTACTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.49 | GAGACGTG | AATGATACGGCGACCACCGAGATCTACACGAGACGTGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.50 | TTGCCTAA | AATGATACGGCGACCACCGAGATCTACACTTGCCTAATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.51 | TTAACTTG | AATGATACGGCGACCACCGAGATCTACACTTAACTTGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.52 | CTTTAACA | AATGATACGGCGACCACCGAGATCTACACCTTTAACATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.53 | CGTAGACC | AATGATACGGCGACCACCGAGATCTACACCGTAGACCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.54 | TATTTGCG | AATGATACGGCGACCACCGAGATCTACACTATTTGCGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.55 | ATCCAGGA | AATGATACGGCGACCACCGAGATCTACACATCCAGGATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.56 | TGTTCCTG | AATGATACGGCGACCACCGAGATCTACACTGTTCCTGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.57 | ACGCGCAG | AATGATACGGCGACCACCGAGATCTACACACGCGCAGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.58 | TCTGGCGA | AATGATACGGCGACCACCGAGATCTACACTCTGGCGATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.59 | AATCTACA | AATGATACGGCGACCACCGAGATCTACACAATCTACATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.60 | TACTGACC | AATGATACGGCGACCACCGAGATCTACACTACTGACCTCGTCGGCAGCGTCAGATGTGTAT |

| | | |
|---|---|---|
| Ad1.61 | CGATAGGG | AATGATACGGCGACCACCGAGATCTACACCGATAGGGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.62 | ACTTAGAA | AATGATACGGCGACCACCGAGATCTACACACTTAGAATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.63 | AGAGATCT | AATGATACGGCGACCACCGAGATCTACACAGAGATCTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.64 | GGTGAAGG | AATGATACGGCGACCACCGAGATCTACACGGTGAAGGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.65 | ATCGAATG | AATGATACGGCGACCACCGAGATCTACACATCGAATGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.66 | TCAAGAGC | AATGATACGGCGACCACCGAGATCTACACTCAAGAGCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.67 | GCCCACGT | AATGATACGGCGACCACCGAGATCTACACGCCCACGTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.68 | TGGGCGGT | AATGATACGGCGACCACCGAGATCTACACTGGGCGGTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.69 | CCCTTGGA | AATGATACGGCGACCACCGAGATCTACACCCCTTGGATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.70 | ATTACCGT | AATGATACGGCGACCACCGAGATCTACACATTACCGTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.71 | AGTCCGAG | AATGATACGGCGACCACCGAGATCTACACAGTCCGAGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.72 | ACTTGTTG | AATGATACGGCGACCACCGAGATCTACACACTTGTTGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.73 | GTAATACA | AATGATACGGCGACCACCGAGATCTACACGTAATACATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.74 | GGCGTCTA | AATGATACGGCGACCACCGAGATCTACACGGCGTCTATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.75 | GCGCTGCT | AATGATACGGCGACCACCGAGATCTACACGCGCTGCTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.76 | GTGCCATT | AATGATACGGCGACCACCGAGATCTACACGTGCCATTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.77 | TAGGTATG | AATGATACGGCGACCACCGAGATCTACACTAGGTATGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.78 | AACACCTA | AATGATACGGCGACCACCGAGATCTACACAACACCTATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.79 | CTCCGAAC | AATGATACGGCGACCACCGAGATCTACACCTCCGAACTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.80 | CAACGGCA | AATGATACGGCGACCACCGAGATCTACACCAACGGCATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.81 | CAATGTAG | AATGATACGGCGACCACCGAGATCTACACCAATGTAGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.82 | GGCTACCC | AATGATACGGCGACCACCGAGATCTACACGGCTACCCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.83 | AAAGTCCG | AATGATACGGCGACCACCGAGATCTACACAAAGTCCGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.84 | TTCCGCGG | AATGATACGGCGACCACCGAGATCTACACTTCCGCGGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.85 | AGGCACTT | AATGATACGGCGACCACCGAGATCTACACAGGCACTTTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.86 | CTTCAGTG | AATGATACGGCGACCACCGAGATCTACACCTTCAGTGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.87 | GCCGGTAG | AATGATACGGCGACCACCGAGATCTACACGCCGGTAGTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.88 | TTCAATCC | AATGATACGGCGACCACCGAGATCTACACTTCAATCCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.89 | CCACACAC | AATGATACGGCGACCACCGAGATCTACACCCACACACTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.90 | ATATTATC | AATGATACGGCGACCACCGAGATCTACACATATTATCTCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.91 | CCGAAGCA | AATGATACGGCGACCACCGAGATCTACACCCGAAGCATCGTCGGCAGCGTCAGATGTGTAT |
| Ad1.92 | GTATCGGT | AATGATACGGCGACCACCGAGATCTACACGTATCGGTTCGTCGGCAGCGTCAGATGTGTAT |

| Custom Barcodes **Adapter 2 (index i7)** IDs | **Barcodes** | **Primer Sequence** |
|---|---|---|
| Ad2.1 | TAAGGCGA | CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.2 | CGTACTAG | CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.3 | AGGCAGAA | CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.4 | TCCTGAGC | CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.5 | GGACTCCT | CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.6 | TAGGCATG | CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.7 | CTCTCTAC | CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.8 | CAGAGAGG | CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.9 | GCTACGCT | CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.10 | CGAGGCTG | CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.11 | AAGAGGCA | CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.12 | GTAGAGGA | CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGTG |
| Ad2.13 | TGGATCTG | CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.14 | CCGTTTGT | CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.15 | TGCTGGGT | CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.16 | AGGTTGGG | CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.17 | GTGTGGTG | CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGTG |
| Ad2.18 | TGGGTTTC | CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.19 | TGGTCACA | CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.20 | TTGACCCT | CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.21 | CGCGGACA | CAAGCAGAAGACGGCATACGAGATTGTCCGCGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.22 | TTCCATAT | CAAGCAGAAGACGGCATACGAGATATATGGAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.23 | AATTCGTT | CAAGCAGAAGACGGCATACGAGATAACGAATTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.24 | GGCGTCGA | CAAGCAGAAGACGGCATACGAGATTCGACGCCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.25 | ACAAAGTG | CAAGCAGAAGACGGCATACGAGATCACTTTGTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.26 | TACTTGAA | CAAGCAGAAGACGGCATACGAGATTTCAAGTAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.27 | GTGATAGC | CAAGCAGAAGACGGCATACGAGATGCTATCACGTCTCGTGGGCTCGGAGATGTG |
| Ad2.28 | AGTAGATT | CAAGCAGAAGACGGCATACGAGATAATCTACTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.29 | ATTGCCGG | CAAGCAGAAGACGGCATACGAGATCCGGCAATGTCTCGTGGGCTCGGAGATGTG |
| Ad2.30 | TTGCTAAG | CAAGCAGAAGACGGCATACGAGATCTTAGCAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.31 | ATAAGTTA | CAAGCAGAAGACGGCATACGAGATTAACTTATGTCTCGTGGGCTCGGAGATGTG |
| Ad2.32 | ATCACTCG | CAAGCAGAAGACGGCATACGAGATCGAGTGATGTCTCGTGGGCTCGGAGATGTG |
| Ad2.33 | GTTAACAG | CAAGCAGAAGACGGCATACGAGATCTGTTAACGTCTCGTGGGCTCGGAGATGTG |
| Ad2.34 | AATGGTAG | CAAGCAGAAGACGGCATACGAGATCTACCATTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.35 | GAGCACGT | CAAGCAGAAGACGGCATACGAGATACGTGCTCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.36 | TTTCGTCA | CAAGCAGAAGACGGCATACGAGATTGACGAAAGTCTCGTGGGCTCGGAGATGTG |

| | | |
|---|---|---|
| Ad2.37 | CAAGAATT | CAAGCAGAAGACGGCATACGAGATAATTCTTGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.38 | GAAATGCC | CAAGCAGAAGACGGCATACGAGATGGCATTTCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.39 | AACGCCAT | CAAGCAGAAGACGGCATACGAGATATGGCGTTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.40 | CCTCGCAG | CAAGCAGAAGACGGCATACGAGATCTGCGAGGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.41 | TACACCTC | CAAGCAGAAGACGGCATACGAGATGAGGTGTAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.42 | GGTCATTT | CAAGCAGAAGACGGCATACGAGATAAATGACCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.43 | CAATCTTA | CAAGCAGAAGACGGCATACGAGATTAAGATTGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.44 | TGTGCCTT | CAAGCAGAAGACGGCATACGAGATAAGGCACAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.45 | TCTTATTA | CAAGCAGAAGACGGCATACGAGATTAATAAGAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.46 | GACTTAGT | CAAGCAGAAGACGGCATACGAGATACTAAGTCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.47 | AGACCAGC | CAAGCAGAAGACGGCATACGAGATGCTGGTCTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.48 | AAATACAG | CAAGCAGAAGACGGCATACGAGATCTGTATTTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.49 | TTATGAAA | CAAGCAGAAGACGGCATACGAGATTTTCATAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.50 | CTTGGGTC | CAAGCAGAAGACGGCATACGAGATGACCCAAGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.51 | CCAAATAA | CAAGCAGAAGACGGCATACGAGATTTATTTGGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.52 | GCGTTAAA | CAAGCAGAAGACGGCATACGAGATTTTAACGCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.53 | CATCCTGT | CAAGCAGAAGACGGCATACGAGATACAGGATGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.54 | GGAGTAAG | CAAGCAGAAGACGGCATACGAGATCTTACTCCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.55 | GACGCTCC | CAAGCAGAAGACGGCATACGAGATGGAGCGTCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.56 | TTCGCGGC | CAAGCAGAAGACGGCATACGAGATGCCGCGAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.57 | CGGTTCCC | CAAGCAGAAGACGGCATACGAGATGGGAACCGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.58 | ACCGGCTA | CAAGCAGAAGACGGCATACGAGATTAGCCGGTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.59 | CTCATGGG | CAAGCAGAAGACGGCATACGAGATCCCATGAGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.60 | TTTAATGC | CAAGCAGAAGACGGCATACGAGATGCATTAAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.61 | AAACGGTC | CAAGCAGAAGACGGCATACGAGATGACCGTTTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.62 | GATCCAAA | CAAGCAGAAGACGGCATACGAGATTTTGGATCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.63 | ATGATGAT | CAAGCAGAAGACGGCATACGAGATATCATCATGTCTCGTGGGCTCGGAGATGTG |
| Ad2.64 | CCAACACG | CAAGCAGAAGACGGCATACGAGATCGTGTTGGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.65 | TAACAACA | CAAGCAGAAGACGGCATACGAGATTGTTGTTAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.66 | GGTAAACC | CAAGCAGAAGACGGCATACGAGATGGTTTACCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.67 | CATCGACC | CAAGCAGAAGACGGCATACGAGATGGTCGATGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.68 | ATGGGAAC | CAAGCAGAAGACGGCATACGAGATGTTCCCATGTCTCGTGGGCTCGGAGATGTG |
| Ad2.69 | CGGCCAAT | CAAGCAGAAGACGGCATACGAGATATTGGCCGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.70 | GGGAATGA | CAAGCAGAAGACGGCATACGAGATTCATTCCCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.71 | GTATTCGG | CAAGCAGAAGACGGCATACGAGATCCGAATACGTCTCGTGGGCTCGGAGATGTG |
| Ad2.72 | TCAGCTAT | CAAGCAGAAGACGGCATACGAGATATAGCTGAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.73 | ATTTATCT | CAAGCAGAAGACGGCATACGAGATAGATAAATGTCTCGTGGGCTCGGAGATGTG |
| Ad2.74 | ACAGTTGC | CAAGCAGAAGACGGCATACGAGATGCAACTGTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.75 | CCCGAGAT | CAAGCAGAAGACGGCATACGAGATATCTCGGGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.76 | TAATGTCT | CAAGCAGAAGACGGCATACGAGATAGACATTAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.77 | GCCAATTC | CAAGCAGAAGACGGCATACGAGATGAATTGGCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.78 | CGCCGTGC | CAAGCAGAAGACGGCATACGAGATGCACGGCGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.79 | CTGACCGA | CAAGCAGAAGACGGCATACGAGATTCGGTCAGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.80 | CATTTCGA | CAAGCAGAAGACGGCATACGAGATTCGAAATGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.81 | GCTTGCCA | CAAGCAGAAGACGGCATACGAGATTGGCAAGCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.82 | TTCTACCA | CAAGCAGAAGACGGCATACGAGATTGGTAGAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.83 | ACGTGACG | CAAGCAGAAGACGGCATACGAGATCGTCACGTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.84 | TGTCCGCG | CAAGCAGAAGACGGCATACGAGATCGCGGACAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.85 | TTAAACTT | CAAGCAGAAGACGGCATACGAGATAAGTTTAAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.86 | ACCACAAC | CAAGCAGAAGACGGCATACGAGATGTTGTGGTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.87 | GCCTCTGG | CAAGCAGAAGACGGCATACGAGATCCAGAGGCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.88 | TCGCCCAC | CAAGCAGAAGACGGCATACGAGATGTGGGCGAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.89 | CACTAGGC | CAAGCAGAAGACGGCATACGAGATGCCTAGTGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.90 | TCGAAGCC | CAAGCAGAAGACGGCATACGAGATGGCTTCGAGTCTCGTGGGCTCGGAGATGTG |
| Ad2.91 | GCATGTAC | CAAGCAGAAGACGGCATACGAGATGTACATGCGTCTCGTGGGCTCGGAGATGTG |
| Ad2.92 | GTTCGAGT | CAAGCAGAAGACGGCATACGAGATACTCGAACGTCTCGTGGGCTCGGAGATGTG |
| Ad2.93 | CCGGGCGC | CAAGCAGAAGACGGCATACGAGATGCGCCCGGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.94 | AGATTTAA | CAAGCAGAAGACGGCATACGAGATTTAAATCTGTCTCGTGGGCTCGGAGATGTG |
| Ad2.95 | CACCATTG | CAAGCAGAAGACGGCATACGAGATCAATGGTGGTCTCGTGGGCTCGGAGATGTG |
| Ad2.96 | AATAAGAC | CAAGCAGAAGACGGCATACGAGATGTCTTATTGTCTCGTGGGCTCGGAGATGTG |

## A 10 Codes for CUT & Tag data analysis

CUT & Tag data analysis following Henicoff's lab protocol: https://yezhengstat.github.io/CUTTag_tutorial/#II_Data_Pre-processing.

**1. Data Pre-processing**
**1.1. Quality Control using FastQC [Optional]**
**1.1.1 Obtain FastQC**

```
##== linux command ==##
projPath="/path/to/project/where/data/and/results/are/saved"
projPath="/Users/minyi/Desktop/CUT_Tag"
cd ${projPath}
mkdir -p $projPath/tools
wget -P $projPath/tools
https://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.9.zip
cd $projPath/tools
unzip fastqc_v0.11.9.zip
```

**1.1.2 Run FastQC for quality check**

```
##== linux command ==##
##== Copy raw data files to "fastq/${histName}" folder with a specified
filename ==##
histName=IRP1A_PG_rep1
mkdir -p ${projPath}/fastq/${histName}
cat
${projPath}/Raw_data_CUT_Tag/${histName}/${histName}*_R1_*.fastq.gz >${projPa
th}/fastq/${histName}/${histName}_R1.fastq.gz
cat
${projPath}/Raw_data_CUT_Tag/${histName}/${histName}*_R2_*.fastq.gz >${projPa
th}/fastq/${histName}/${histName}_R2.fastq.gz

##== change "histName" and run the "cat" command again to copy and rename all
samples ==##
histName=IRP1A_PG_rep2
histName=w1118_PG_rep1
histName=w1118_PG_rep2
histName=IRP1A_FB_rep1
histName=IRP1A_FB_rep2
histName=w1118_FB_rep1
histName=w1118_FB_rep2

##== Run FastQC for quality check ==##
histName=IRP1A_PG_rep1
mkdir -p ${projPath}/fastqFileQC/${histName}
```

```
$projPath/tools/FastQC/fastqc -o ${projPath}/fastqFileQC/${histName} -f fastq
${projPath}/fastq/${histName}/${histName}_R1.fastq.gz
$projPath/tools/FastQC/fastqc -o ${projPath}/fastqFileQC/${histName} -f fastq
${projPath}/fastq/${histName}/${histName}_R2.fastq.gz
```

**2. Alignment**

**2.1. Bowtie2 alignment**

```
##== remove adapter sequences before alignment ==##
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
histName=IRP1A_PG_rep1
mkdir -p ${projPath}/TrimmedSeq/${histName}_Trimmed

java -jar ${projPath}/Trimmomatic/trimmomatic-0.39.jar PE
${projPath}/fastq/${histName}/${histName}_R1.fastq.gz
${projPath}/fastq/${histName}/${histName}_R2.fastq.gz
${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R1P.fastq.gz
${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R1U.fastq.gz
${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R2P.fastq.gz
${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R2U.fastq.gz
ILLUMINACLIP:${projPath}/Trimmomatic/adapters/NexteraPE-PE.fa:2:30:10
LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

##== Run FastQC to check the quality of trimmed sequences ==##
histName=IRP1A_PG_rep1
mkdir -p ${projPath}/fastqFileQC_Trimmed/${histName}

$projPath/tools/FastQC/fastqc -o ${projPath}/fastqFileQC_Trimmed/${histName}
-f fastq ${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R1P.fastq.gz
$projPath/tools/FastQC/fastqc -o ${projPath}/fastqFileQC_Trimmed/${histName}
-f fastq ${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R2P.fastq.gz
```

**2.1.1 Alignment to BDGP6**

```
##== Use trimmed, paired-end sequences for alignment ==##
##== Download Bowtie2: https://sourceforge.net/projects/bowtie-
bio/files/bowtie2/2.5.1/ ==##
##== Download Bowtie2 files for Drosophila reference seq: https://bowtie-
bio.sourceforge.net/bowtie2/index.shtml ==##

##== linux command ==##
cores=8
projPath=/Users/minyi/Desktop/CUT_Tag
ref=/Users/minyi/Desktop/CUT_Tag/BDGP6
```

```
histName=IRP1A_PG_rep1

mkdir -p ${projPath}/alignment/sam/bowtie2_summary
mkdir -p ${projPath}/alignment/bam
mkdir -p ${projPath}/alignment/bed
mkdir -p ${projPath}/alignment/bedgraph
export PATH=${projPath}/bowtie2-2.5.1-macos-arm64:$PATH
export BOWTIE2_INDEXES=${projPath}/BDGP6

bowtie2 --end-to-end --very-sensitive --no-mixed --no-discordant --phred33 -I
10 -X 700 -p ${cores} -x BDGP6 -1
${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R1P.fastq.gz -2
${projPath}/TrimmedSeq/${histName}_Trimmed/${histName}_R2P.fastq.gz -S
${projPath}/alignment/sam/${histName}_bowtie2.sam &>
${projPath}/alignment/sam/bowtie2_summary/${histName}_bowtie2.txt
```

**2.2 Report sequencing mapping summary**

**2.2.1 Sequencing depth**

```
##=== R command ===##
## Path to the project and histone list
library(dplyr)
library(stringr)

setwd("/Users/minyi/Desktop/CUT_Tag/SequencingDepth")
projPath = "/Users/minyi/Desktop/CUT_Tag"
sampleList = c("IRP1A_PG_rep1", "IRP1A_PG_rep2", "w1118_PG_rep1",
"w1118_PG_rep2", "IRP1A_FB_rep1", "IRP1A_FB_rep2", "w1118_FB_rep1",
"w1118_FB_rep2")
histList = c("IRP1A_PG", "w1118_PG", "IRP1A_FB", "w1118_FB")

## Collect the alignment results from the bowtie2 alignment summary files ##
alignResult = c()
for(hist in sampleList){
  alignRes = read.table(paste0(projPath, "/alignment/sam/bowtie2_summary/",
hist, "_bowtie2.txt"), header = FALSE, fill = TRUE)
  alignRate = substr(alignRes$V1[6], 1, nchar(as.character(alignRes$V1[6]))-
1)
  histInfo = strsplit(hist, "_")[[1]]
  histone <- paste(histInfo[1], histInfo[2], sep = "_")
  alignResult = data.frame(Histone = histone, Replicate = histInfo[3],
                           SequencingDepth = alignRes$V1[1] %>%
as.character %>% as.numeric,
                           MappedFragNum_BDGP6 = alignRes$V1[4] %>%
as.character %>% as.numeric + alignRes$V1[5] %>% as.character %>% as.numeric,
                           AlignmentRate_BDGP6 = alignRate %>%
as.numeric)  %>% rbind(alignResult, .)
```

```
}
alignResult$Histone = factor(alignResult$Histone, levels = histList)
alignResult <- mutate(alignResult, AlignmentRate_BDGP6 =
paste0(AlignmentRate_BDGP6, "%"))

write.table(alignResult, "alignResult.txt",sep = "\t", col.names = TRUE,
row.names = FALSE)
```

### 2.2.2 Visualizing the sequencing depth and alignment results

```
##== R command ==##
library(ggplot2)
library(viridis)
library(ggpubr)

fig3A = alignResult %>% ggplot(aes(x = Histone, y = SequencingDepth/1000000,
fill = Histone)) +
  geom_boxplot() +
  geom_jitter(aes(color = Replicate), position = position_jitter(0.15)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("Sequencing Depth per Million") +
  xlab("") +
  ggtitle("A. Sequencing Depth")

fig3B = alignResult %>% ggplot(aes(x = Histone, y =
MappedFragNum_BDGP6/1000000, fill = Histone)) +
  geom_boxplot() +
  geom_jitter(aes(color = Replicate), position = position_jitter(0.15)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("Mapped Fragments per Million") +
  xlab("") +
  ggtitle("B. Alignable Fragment")

fig3C = alignResult %>% ggplot(aes(x = Histone, y = AlignmentRate_BDGP6, fill
= Histone)) +
  geom_boxplot() +
  geom_jitter(aes(color = Replicate), position = position_jitter(0.15)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("% of Mapped Fragments") +
```

```
  xlab("") +
  ggtitle("C. Alignment Rate")

ggarrange(fig3A, fig3B, fig3C, ncol = 2, nrow=2, common.legend = TRUE,
legend="bottom")
```

**2.3. Remove duplicates**

```
##== Download and install Picard: https://broadinstitute.github.io/picard/
==##
##== linux command ==##
cd /Users/minyi/Desktop/CUT_Tag
java -jar /Users/minyi/Desktop/CUT_Tag/picard.jar -h

##== linux command ==##
## depending on how you load picard and your server environment, the
picardCMD can be different. Adjust accordingly.
picardCMD="java -jar picard.jar"
projPath=/Users/minyi/Desktop/CUT_Tag
histName=IRP1A_PG_rep1
mkdir -p $projPath/alignment/removeDuplicate/picard_summary

## Sort by coordinate
eval $picardCMD SortSam I=$projPath/alignment/sam/${histName}_bowtie2.sam
O=$projPath/alignment/sam/${histName}_bowtie2.sorted.sam
SORT_ORDER=coordinate

## mark duplicates
eval $picardCMD MarkDuplicates
I=$projPath/alignment/sam/${histName}_bowtie2.sorted.sam
O=$projPath/alignment/removeDuplicate/${histName}_bowtie2.sorted.dupMarked.sa
m
METRICS_FILE=$projPath/alignment/removeDuplicate/picard_summary/${histName}_p
icard.dupMark.txt

## remove duplicates
eval $picardCMD MarkDuplicates
I=$projPath/alignment/sam/${histName}_bowtie2.sorted.sam
O=$projPath/alignment/removeDuplicate/${histName}_bowtie2.sorted.rmDup.sam
REMOVE_DUPLICATES=true ADD_PG_TAG_TO_READS=false
METRICS_FILE=$projPath/alignment/removeDuplicate/picard_summary/${histName}_p
icard.rmDup.txt


## Summarize the duplication information and sequencing depth
##=== R command ===##
## Combine the sequencing depth and ChIPseqSpikeInFree results
```

```
setwd("/Users/minyi/Desktop/CUT_Tag/SequencingDepth")

# Read in the data from each file
SequencingDepth <-
read.table("/Users/minyi/Desktop/CUT_Tag/SequencingDepth/alignResult.txt",
sep = "", header = TRUE)
ChIPseqSpikeInFree_SF <-
read.table("/Users/minyi/Desktop/CUT_Tag/ChIPseqSpikeInFree/test_SF.txt", sep
= "\t", header = TRUE)

# Extract the column(s) you want to add from file2_data
ScalingFactor <- ChIPseqSpikeInFree_SF$SF # Replace "column_name" with the
name of the column(s) you want to add

# Combine the two data frames by column
alignSummary <- cbind(SequencingDepth, ScalingFactor)

# Write the merged data to a new file
write.table(alignSummary, "alignSummary.txt", sep = "\t", col.names = TRUE,
row.names = FALSE)

## Summarize the duplication information from the picard summary outputs.
library(dplyr)
library(stringr)

setwd("/Users/minyi/Desktop/CUT_Tag/alignment/removeDuplicate/picard_summary"
)
projPath = "/Users/minyi/Desktop/CUT_Tag"
sampleList = c("IRP1A_PG_rep1", "IRP1A_PG_rep2", "w1118_PG_rep1",
"w1118_PG_rep2",
               "IRP1A_FB_rep1", "IRP1A_FB_rep2", "w1118_FB_rep1",
"w1118_FB_rep2")
histList = c("IRP1A_PG", "w1118_PG", "IRP1A_FB", "w1118_FB")

dupResult = c()
for(hist in sampleList){
  dupRes = read.table(paste0(projPath,
"/alignment/removeDuplicate/picard_summary/", hist, "_picard.rmDup.txt"),
header = TRUE, fill = TRUE)

  histInfo = strsplit(hist, "_")[[1]]
  histone <- paste(histInfo[1], histInfo[2], sep = "_")
  dupResult = data.frame(Histone = histone, Replicate = histInfo[3],
                         MappedFragNum_BDGP6 =
dupRes$READ_PAIRS_EXAMINED[1] %>% as.character %>% as.numeric,
                         DuplicationRate = dupRes$PERCENT_DUPLICATION[1] %>%
as.character %>% as.numeric * 100,
                         EstimatedLibrarySize =
dupRes$ESTIMATED_LIBRARY_SIZE[1] %>% as.character %>% as.numeric) %>%
```

```
mutate(UniqueFragNum = MappedFragNum_BDGP6 * (1-DuplicationRate/100))  %>%
rbind(dupResult, .)
}
dupResult$Histone = factor(dupResult$Histone, levels = histList)
alignDupSummary = left_join(alignSummary, dupResult, by = c("Histone",
"Replicate", "MappedFragNum_BDGP6")) %>% mutate(DuplicationRate =
paste0(DuplicationRate, "%"))
alignDupSummary

write.table(alignDupSummary, "alignDupSummary.txt", sep = "\t", col.names =
TRUE, row.names = FALSE)



## generate boxplot figure for the  duplication rate
library(ggplot2)
library(viridis)
library(ggpubr)

fig4A = dupResult %>% ggplot(aes(x = Histone, y = DuplicationRate, fill =
Histone)) +
  geom_boxplot() +
  geom_jitter(aes(color = Replicate), position = position_jitter(0.15)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("Duplication Rate (*100%)") +
  xlab("")

fig4B = dupResult %>% ggplot(aes(x = Histone, y = EstimatedLibrarySize, fill
= Histone)) +
  geom_boxplot() +
  geom_jitter(aes(color = Replicate), position = position_jitter(0.15)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("Estimated Library Size") +
  xlab("")

fig4C = dupResult %>% ggplot(aes(x = Histone, y = UniqueFragNum, fill =
Histone)) +
  geom_boxplot() +
  geom_jitter(aes(color = Replicate), position = position_jitter(0.15)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
```

```
  ylab("# of Unique Fragments") +
  xlab("")

ggarrange(fig4A, fig4B, fig4C, ncol = 3, common.legend = TRUE,
legend="bottom")

##== According to the ENCODE ChIP-Seq guideline, the duplication rate should
be <= 20%. The duplication rate of my data > 60%, which might be caused by
PCR amplification. Therefore, duplicates should be removed before downstream
analysis to avoid false positive results in peak calling.
```

**2.4. Assess mapped fragment size distribution**

```
##== Download and install samtools: https://github.com/samtools/samtools ==##

##== linux command ==##
cd /Users/minyi/Desktop/CUT_Tag/samtools-1.17

##== Install Apple's Command Line Developer Tools ==##
xcode-select --install

##== Install samtools ==##
mkdir /Users/minyi/Desktop/CUT_Tag/samtools
./configure --prefix=/Users/minyi/Desktop/CUT_Tag/samtools
make
make install

##== The executable programs will be installed to a bin subdirectory under
your specified prefix, so you may wish to add this directory to your $PATH:
http://www.htslib.org/download/ ==##
export PATH=$PATH:/Users/minyi/Desktop/CUT_Tag/samtools/bin

##== Assess mapped fragment size distribution after removing duplicates==##
##== linux command ==##
mkdir -p $projPath/alignment/sam/fragmentLen
histName=IRP1A_PG_rep1

samtools view -F 0x04
$projPath/alignment/removeDuplicate/${histName}_bowtie2.sorted.rmDup.sam |
awk -F'\t' 'function abs(x){return ((x < 0.0) ? -x : x)} {print abs($9)}' |
sort | uniq -c | awk -v OFS="\t" '{print $2,
$1/2}' >$projPath/alignment/sam/fragmentLen/${histName}_fragmentLen.txt


##=== R command ===##
## Collect the fragment size information
library(dplyr)
```

```
library(stringr)
library(ggplot2)
library(viridis)
library(ggpubr)

setwd("/Users/minyi/Desktop/CUT_Tag/alignment/sam/fragmentLen")
projPath = "/Users/minyi/Desktop/CUT_Tag"
sampleList = c("IRP1A_PG_rep1", "IRP1A_PG_rep2", "w1118_PG_rep1",
"w1118_PG_rep2",
               "IRP1A_FB_rep1", "IRP1A_FB_rep2", "w1118_FB_rep1",
"w1118_FB_rep2")
histList = c("IRP1A_PG", "w1118_PG", "IRP1A_FB", "w1118_FB")

fragLen = c()
for(hist in sampleList){

  histInfo = strsplit(hist, "_")[[1]]
  histone <- paste(histInfo[1], histInfo[2], sep = "_")

  fragLen = read.table(paste0(projPath, "/alignment/sam/fragmentLen/", hist,
"_fragmentLen.txt"), header = FALSE) %>% mutate(fragLen = V1 %>% as.numeric,
fragCount = V2 %>% as.numeric, Weight = as.numeric(V2)/sum(as.numeric(V2)),
Histone = histone, Replicate = histInfo[3], sampleInfo = hist) %>%
rbind(fragLen, .)
}
fragLen$sampleInfo = factor(fragLen$sampleInfo, levels = sampleList)
fragLen$Histone = factor(fragLen$Histone, levels = histList)
## Generate the fragment size density plot (violin plot)
fig5A = fragLen %>% ggplot(aes(x = sampleInfo, y = fragLen, weight = Weight,
fill = Histone)) +
  geom_violin(bw = 5) +
  scale_y_continuous(breaks = seq(0, 800, 50)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 20) +
  ggpubr::rotate_x_text(angle = 20) +
  ylab("Fragment Length") +
  xlab("")

fig5B = fragLen %>% ggplot(aes(x = fragLen, y = fragCount, color = Histone,
group = sampleInfo, linetype = Replicate)) +
  geom_line(linewidth = 1) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9, option =
"magma") +
  theme_bw(base_size = 20) +
  xlab("Fragment Length") +
  ylab("Count") +
  coord_cartesian(xlim = c(0, 500))
```

```
ggarrange(fig5A, fig5B, ncol = 2)
```

**3. Alignment results filtering and file format conversion**
**3.1 File format conversion**

```
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
histName=IRP1A_PG_rep1

##== Download bedtools and static binary files:
https://bedtools.readthedocs.io/en/latest/content/installation.html ==##
mv bedtools.static.binary ${projPath}/bedtools2
chmod a+x ${projPath}/bedtools2

cd ${projPath}/bedtools2
export PATH=$PATH:/Users/minyi/Desktop/CUT_Tag/bedtools2/bin

## Use files after duplicate removal for analysis ##
## file format conversion requires reads are sorted by read names. Sort by
read names using "samtools sort -n" ##
samtools sort -n
$projPath/alignment/removeDuplicate/${histName}_bowtie2.sorted.rmDup.sam -o
$projPath/alignment/removeDuplicate/${histName}_bowtie2.sortedByname.rmDup.sa
m

samtools view -bS -F 0x04
$projPath/alignment/removeDuplicate/${histName}_bowtie2.sortedByname.rmDup.sa
m >$projPath/alignment/bam/${histName}_bowtie2.mapped.bam

## Convert into bed file format ##
bedtools bamtobed -i $projPath/alignment/bam/${histName}_bowtie2.mapped.bam -
bedpe >$projPath/alignment/bed/${histName}_bowtie2.bed

## Keep the read pairs that are on the same chromosome and fragment length
less than 1000bp. ##
awk '$1==$4 && $6-$2 < 1000 {print $0}'
$projPath/alignment/bed/${histName}_bowtie2.bed >$projPath/alignment/bed/${hi
stName}_bowtie2.clean.bed

## Only extract the fragment related columns ##
cut -f 1,2,6 $projPath/alignment/bed/${histName}_bowtie2.clean.bed | sort -
k1,1 -k2,2n -
k3,3n  >$projPath/alignment/bed/${histName}_bowtie2.fragments.bed
```

**3.2 Assess replicate reproducibility**

```
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
histName=IRP1A_PG_rep1

## We use the mid point of each fragment to infer which 500bp bins does this
fragment belong to.
binLen=500
awk -v w=$binLen '{print $1, int(($2 + $3)/(2*w))*w + w/2}'
$projPath/alignment/bed/${histName}_bowtie2.fragments.bed | sort -k1,1V -
k2,2n | uniq -c | awk -v OFS="\t" '{print $2, $3, $1}' |  sort -k1,1V -
k2,2n  >$projPath/alignment/bed/${histName}_bowtie2.fragmentsCount.bin$binLen
.bed


##== R command ==##
library(dplyr)
library(corrplot)

dir.create("/Users/minyi/Desktop/CUT_Tag/Replicate_reproducibility")
setwd("/Users/minyi/Desktop/CUT_Tag/Replicate_reproducibility")
projPath = "/Users/minyi/Desktop/CUT_Tag"
sampleList = c("IRP1A_PG_rep1", "IRP1A_PG_rep2", "w1118_PG_rep1",
"w1118_PG_rep2",
               "IRP1A_FB_rep1", "IRP1A_FB_rep2", "w1118_FB_rep1",
"w1118_FB_rep2")
histList = c("IRP1A_PG", "w1118_PG", "IRP1A_FB", "w1118_FB")

reprod = c()
fragCount = NULL
for(hist in sampleList){

  if(is.null(fragCount)){

    fragCount = read.table(paste0(projPath, "/alignment/bed/", hist,
"_bowtie2.fragmentsCount.bin500.bed"), header = FALSE)
    colnames(fragCount) = c("chrom", "bin", hist)

  }else{

    fragCountTmp = read.table(paste0(projPath, "/alignment/bed/", hist,
"_bowtie2.fragmentsCount.bin500.bed"), header = FALSE)
    colnames(fragCountTmp) = c("chrom", "bin", hist)
    fragCount = full_join(fragCount, fragCountTmp, by = c("chrom", "bin"))

  }
}
```

```
M = cor(fragCount %>% select(-c("chrom", "bin")) %>% log2(), use =
"complete.obs")

corrplot(M, method = "color", outline = T, addgrid.col = "darkgray",
order="hclust", addrect = 3, rect.col = "black", rect.lwd = 3,cl.pos = "b",
tl.col = "indianred4", tl.cex = 1, cl.cex = 1, addCoef.col = "black",
number.digits = 2, number.cex = 1, col =
colorRampPalette(c("midnightblue","white","darkred"))(100))
```

**4. Spike-in calibration: ChIPseqSpikeInFree**

```
##== ChIPseqSpikeInFree calibration ==##

#Jin H, Kasper LH, Larson JD, Wu G, Baker SJ, Zhang J, Fan Y.
ChIPseqSpikeInFree: a ChIP-seq normalization approach to reveal global
changes in histone modifications without spike-in. Bioinformatics. 2020 Feb
15;36(4):1270-1272. doi: 10.1093/bioinformatics/btz720. PMID: 31566663;
PMCID: PMC7523640.
#https://github.com/stjude/ChIPseqSpikeInFree

#Download dm6 chromosome size file as .txt file:
http://hgdownload.soe.ucsc.edu/goldenPath/dm6/bigZips/dm6.chrom.sizes

##== R command ==##
library("ChIPseqSpikeInFree")

if (!requireNamespace("BiocManager", quietly = TRUE))
     install.packages("BiocManager")
BiocManager::install("Rsamtools")
BiocManager::install("GenomicRanges")
BiocManager::install("GenomicAlignments")

install.packages("devtools")
library(devtools)
install_github("stjude/ChIPseqSpikeInFree")
packageVersion('ChIPseqSpikeInFree')

dir.create("/Users/minyi/Desktop/CUT_Tag/ChIPseqSpikeInFree")
setwd("/Users/minyi/Desktop/CUT_Tag/ChIPseqSpikeInFree")
projPath = "/Users/minyi/Desktop/CUT_Tag"

#Generate sample_meta.txt (tab-delimited txt file with header line)

metaFile <- "/Users/minyi/Desktop/CUT_Tag/alignment/bam/Sample_meta.txt"

bams <-
c("/Users/minyi/Desktop/CUT_Tag/alignment/bam/IRP1A_PG_rep1_bowtie2.mapped.ba
```

```
m",
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/IRP1A_PG_rep2_bowtie2.mapped.bam"
,
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/w1118_PG_rep1_bowtie2.mapped.bam"
,
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/w1118_PG_rep2_bowtie2.mapped.bam"
,
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/IRP1A_FB_rep1_bowtie2.mapped.bam"
,
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/IRP1A_FB_rep2_bowtie2.mapped.bam"
,
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/w1118_FB_rep1_bowtie2.mapped.bam"
,
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/w1118_FB_rep2_bowtie2.mapped.bam"
)

ChIPseqSpikeInFree(bamFiles = bams, chromFile =
"/Users/minyi/Desktop/CUT_Tag/alignment/bam/dm6.chrom.sizes.txt", metaFile =
metaFile, prefix = "test", cutoff_QC = 1, maxLastTurn=0.97)
```

**5. Peak calling**

**5.1. SEACR**

```
##== Generate bedgragh file with ChIPseqSpikeInFree scaling factor for SEACR
analysis ==##
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
chromSize=/Users/minyi/Desktop/CUT_Tag/alignment/bam/dm6.chrom.sizes.txt
histName=IRP1A_PG_rep1
mkdir -p $projPath/alignment/bedgraph

PATH=$PATH:/Users/minyi/Desktop/CUT_Tag/samtools/bin
samtools sort $projPath/alignment/bam/${histName}_bowtie2.mapped.bam -o
$projPath/alignment/bam/${histName}_bowtie2.mapped.sorted.bam


##== Generate bedgragh file with ChIPseqSpikeInFree scaling factor ==##
##== If your result from ChipSpikeInFree is not 1.0 then you would insert the
value in place of $scale_factor in the call to Bedtools ==##
export PATH=$PATH:/Users/minyi/Desktop/CUT_Tag/bedtools2/bin

for i in $(seq 2 $(wc -l < "$projPath/ChIPseqSpikeInFree/test_SF.txt")); do
    histName=$(awk -v line=$i 'NR==line{split($1, a, "_"); print
a[1]"_"a[2]"_"a[3]}' "$projPath/ChIPseqSpikeInFree/test_SF.txt") ##== The
split function is used to split the first column of the input file based on
the delimiter "_", and then the first part is extracted as "histName". ==##
```

```
    scale_factor=$(awk -v line=$i 'NR==line{print $7}'
"$projPath/ChIPseqSpikeInFree/test_SF.txt")
    echo "Value $i: $histName $scale_factor"

    bedtools genomecov -scale "$scale_factor" -ibam
"$projPath/alignment/bam/${histName}_bowtie2.mapped.sorted.bam" -bg >
"$projPath/alignment/bedgraph/${histName}_bowtie2.sorted.normalized.bedgraph"
done


##==  Download SEACR: https://github.com/FredHutch/SEACR/releases ==##
##== linux command ==##
seacr=/Users/minyi/Desktop/CUT_Tag/SEACR-1.4-beta.2/SEACR_1.4.sh
projPath=/Users/minyi/Desktop/CUT_Tag
mkdir -p $projPath/peakCalling/SEACR
export PATH=$PATH:/Users/minyi/Desktop/CUT_Tag/bedtools2/bin

##== APG1 vs WPG1, APG2 vs WPG2==##
##== Since we have normalized fragment counts with the ChIPseqSpikeInFree, we
set the normalization option of SEACR to "non". Otherwise, the "norm" is
recommended. ==##
histName=IRP1A_PG_rep1
histControl=w1118_PG_rep1

bash $seacr -b
$projPath/alignment/bedgraph/${histName}_bowtie2.sorted.normalized.bedgraph -
c
$projPath/alignment/bedgraph/${histControl}_bowtie2.sorted.normalized.bedgrap
h -n non -m stringent -o
$projPath/peakCalling/SEACR/${histName}_seacr_control.peaks

bash $seacr -b
$projPath/alignment/bedgraph/${histName}_bowtie2.sorted.normalized.bedgraph -
c 0.01 -n non -m stringent -o
$projPath/peakCalling/SEACR/${histName}_seacr_top0.01.peaks
```

### 5.1.1 Number of peaks called

```
##=== R command ===##
library(dplyr)

setwd("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR")
projPath = "/Users/minyi/Desktop/CUT_Tag"
sampleList = c("IRP1A_PG_rep1", "IRP1A_PG_rep2", "IRP1A_FB_rep1",
"IRP1A_FB_rep2")
histList = c("IRP1A_PG", "IRP1A_FB")
```

```
peakN = c()
peakWidth = c()
peakType = c("control", "top0.01")

for(hist in sampleList) {
  histInfo = strsplit(hist, "_")[[1]]
  histone <- paste(histInfo[1], histInfo[2], sep = "_")

  for(type in peakType) {
      # Read in the file and perform calculations
      peakInfo = read.table(paste0(projPath, "/peakCalling/SEACR/", hist,
"_seacr_", type, ".peaks.stringent.bed"),
                            header = FALSE, fill = TRUE) %>%
        mutate(width = abs(V3 - V2))
      peakN = data.frame(peakN = nrow(peakInfo), peakType = type, Histone =
histone,
                         Replicate = histInfo[3]) %>% rbind(peakN, .)
      peakWidth = data.frame(width = peakInfo$width, peakType = type, Histone
= histone,
                             Replicate = histInfo[3]) %>% rbind(peakWidth, .)

  }
}

peakN <- select(peakN, Histone, Replicate, peakType, peakN)
write.table(peakN, "PeakNumber.txt", sep = "\t", col.names = TRUE, row.names
= FALSE)
```

#### 5.1.2 Reproducibility of the peak across biological replicates

```
##=== R command ===##
## The reproducibility is calculated by:
# peaks overlapping rep1 and rep2/# peaks of rep1 or rep2 * 100
# Therefore, it is sensitive to the total number of peaks called in each
replicate.

library(dplyr)
library(GenomicRanges)

setwd("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR")
projPath = "/Users/minyi/Desktop/CUT_Tag"

histL = c("IRP1A_PG", "IRP1A_FB")
repL = paste0("rep", 1:2)
peakType = c("control", "top0.01")
peakOverlap = c()
for(type in peakType){
```

```
  for(hist in histL){
    overlap.gr = GRanges()
    for(rep in repL){
      peakInfo = read.table(paste0(projPath, "/peakCalling/SEACR/", hist,
"_", rep, "_seacr_", type, ".peaks.stringent.bed"), header = FALSE, fill =
TRUE)
      peakInfo.gr = GRanges(peakInfo$V1, IRanges(start = peakInfo$V2, end =
peakInfo$V3), strand = "*")
      if(length(overlap.gr) >0){
        overlap.gr = overlap.gr[findOverlaps(overlap.gr, peakInfo.gr)@from]
      }else{
        overlap.gr = peakInfo.gr

      }
    }
    peakOverlap = data.frame(peakReprod = length(overlap.gr), Histone = hist,
peakType = type) %>% rbind(peakOverlap, .)
  }
}

peakReprod = left_join(peakN, peakOverlap, by = c("Histone", "peakType")) %>%
mutate(peakReprodRate = peakReprod/peakN * 100)
peakReprod <- select(peakReprod, Histone, Replicate, peakType, peakN,
peakReprodNum = peakReprod, peakReprodRate)

write.table(peakReprod, "peakReproducibility.txt", sep = "\t", row.names =
FALSE)
```

#### 5.1.3 FRagment proportion in Peaks regions (FRiPs)

```
##=== R command ===##
## We calculate the fraction of reads in peaks (FRiPs) as a measure of
signal-to-noise,
## and contrast it to FRiPs in the IgG control dataset for illustration.
## Although sequencing depths for CUT & Tag are typically only 1-5 million
reads,
## the low background of the method results in high FRiP scores.
library(dplyr)
library(GenomicRanges)

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("chromVAR")
library(chromVAR)

setwd("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR")
```

```
projPath = "/Users/minyi/Desktop/CUT_Tag"
histL = c("IRP1A_PG", "IRP1A_FB")
repL = paste0("rep", 1:2)
bamDir = paste0(projPath, "/alignment/bam")
inPeakData = c()
## overlap with bam file to get count
for(hist in histL){
  for(rep in repL){
    peakRes = read.table(paste0(projPath, "/peakCalling/SEACR/", hist, "_",
rep, "_seacr_control.peaks.stringent.bed"), header = FALSE, fill = TRUE)
    peak.gr = GRanges(seqnames = peakRes$V1, IRanges(start = peakRes$V2, end
= peakRes$V3), strand = "*")
    bamFile = paste0(bamDir, "/", hist, "_", rep, "_bowtie2.mapped.bam")
    fragment_counts <- getCounts(bamFile, peak.gr, paired = TRUE, by_rg =
FALSE, format = "bam")
    inPeakN = counts(fragment_counts)[,1] %>% sum
    inPeakData = rbind(inPeakData, data.frame(inPeakN = inPeakN, Histone =
hist, Replicate = rep))
  }
}

## Peak calling was done using bam files after duplicate removal
## calculate FRiP using the "EstimatedLibrarySize" after removing duplicates
instead of "MappedFragNum_BDGP6"
alignResult = read.table(paste0(projPath,
"/alignment/removeDuplicate/picard_summary/alignDupSummary.txt"), header =
TRUE, fill = TRUE)

frip = left_join(inPeakData, alignResult, by = c("Histone", "Replicate")) %>%
mutate(frip = inPeakN/EstimatedLibrarySize * 100)
frip <- select(frip, Histone, Replicate, SequencingDepth,
MappedFragNum_BDGP6, AlignmentRate_BDGP6, EstimatedLibrarySize, FragInPeakNum
= inPeakN, FRiPs = frip)

write.table(frip, "frip.txt", sep = "\t", col.names = TRUE, row.names =
FALSE)
```

**5.1.4 Visualization of peak number, peak width, peak reproducibility and FRiPs.**

```
##=== R command ===##
library(ggplot2)
library(viridis)
library(ggpubr)

setwd("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR")
projPath = "/Users/minyi/Desktop/CUT_Tag"
```

```
fig7A = peakN %>% ggplot(aes(x = Histone, y = peakN, fill = Histone)) +
  geom_boxplot() +
  geom_jitter(aes(color = Replicate), position = position_jitter(0.15)) +
  facet_grid(~peakType) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.55, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("Number of Peaks") +
  xlab("")

fig7B = peakWidth %>% ggplot(aes(x = Histone, y = width, fill = Histone)) +
  geom_violin() +
  facet_grid(Replicate~peakType) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.55, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  scale_y_continuous(trans = "log", breaks = c(400, 3000, 22000)) +
  theme_bw(base_size = 18) +
  ylab("Width of Peaks") +
  xlab("")

fig7C = peakReprod %>% ggplot(aes(x = Histone, y = peakReprodRate, fill =
Histone, label = round(peakReprodRate, 2))) +
  geom_bar(stat = "identity") +
  geom_text(vjust = 0.1) +
  facet_grid(Replicate~peakType) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.55, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("% of Peaks Reproduced") +
  xlab("")

fig7D <- frip %>% ggplot(aes(x = Histone, y = FRiPs, fill = Histone)) +
  geom_boxplot(aes(group = Histone)) +
  geom_jitter(aes(color = Replicate, group = Replicate), position =
position_jitter(0.15)) +
  scale_fill_viridis(discrete = TRUE, begin = 0.1, end = 0.55, option =
"magma", alpha = 0.8) +
  scale_color_viridis(discrete = TRUE, begin = 0.1, end = 0.9) +
  theme_bw(base_size = 18) +
  ylab("% of Fragments in Peaks") +
  xlab("")

ggarrange(fig7A, fig7B, fig7C, fig7D, ncol = 2, nrow=2, common.legend = TRUE,
legend="bottom")
```

**6. Visualization**

**6.1. Browser display of normalized bedgraph files**

```
##== IGV website or desktop application: https://igv.org/app/  ==##
```

**6.2. Heatmap visualization on specific regions**

```
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
mkdir -p $projPath/alignment/bigwig
histName=IRP1A_PG_rep1

##== We have sorted the data before SEACR peak calling. If not, sort it use
samtools: ==##
samtools sort -o $projPath/alignment/bam/${histName}.sorted.bam
$projPath/alignment/bam/${histName}_bowtie2.mapped.bam

##== Download and install deepTools:
https://deeptools.readthedocs.io/en/develop/content/installation.html#id2
==##
##== Download and install Anaconda:
https://docs.anaconda.com/free/anaconda/install/mac-os/ ==##
##== All of the requirements for deepTools can be installed in Anaconda with:
==##
export PATH="/Users/minyi/anaconda3/bin:$PATH"
conda install -c bioconda deeptools
##== Download and install pip: https://pypi.org/project/pip/#files ==##
##== Download and install deepTools using pip ==##
pip install deeptools

##== Generate bigwig files ==##
PATH=$PATH:/Users/minyi/Desktop/CUT_Tag/samtools/bin
samtools index $projPath/alignment/bam/${histName}_bowtie2.mapped.sorted.bam

bamCoverage -b $projPath/alignment/bam/${histName}_bowtie2.mapped.sorted.bam
-o $projPath/alignment/bigwig/${histName}_raw.bw

##== change "histName" and generate bigwig files for all experimental samples
==##
histName=IRP1A_PG_rep2
histName=IRP1A_FB_rep1
histName=IRP1A_FB_rep2
```

### 6.2.1 Heatmap over transcription units

```
##== Download Drosophila dm6 genome BED file from UCSC:
https://genome.ucsc.edu/cgi-bin/hgTables
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
mkdir -p $projPath/data

cores=6
computeMatrix scale-regions -S
$projPath/alignment/bigwig/IRP1A_PG_rep1_raw.bw
$projPath/alignment/bigwig/IRP1A_PG_rep2_raw.bw
$projPath/alignment/bigwig/IRP1A_FB_rep1_raw.bw
$projPath/alignment/bigwig/IRP1A_FB_rep2_raw.bw -R $projPath/data/BDGP6_dm6 -
-beforeRegionStartLength 3000 --regionBodyLength 5000 --
afterRegionStartLength 3000 --skipZeros -o $projPath/data/matrix_gene.mat.gz
-p $cores

plotHeatmap -m $projPath/data/matrix_gene.mat.gz -out
$projPath/data/IRP1A_gene.png --sortUsing sum
```

### 6.2.2. Heatmap on CUT & Tag peaks

```
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
cd=/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR
histName=IRP1A_PG_rep1

cores=6
awk '{split($6, summit, ":"); split(summit[2], region, "-"); print
summit[1]"\t"region[1]"\t"region[2]}'
$projPath/peakCalling/SEACR/${histName}_seacr_control.peaks.stringent.bed >$p
rojPath/peakCalling/SEACR/${histName}_seacr_control.peaks.summitRegion.bed

computeMatrix reference-point -S
$projPath/alignment/bigwig/${histName}_raw.bw -R
$projPath/peakCalling/SEACR/${histName}_seacr_control.peaks.summitRegion.bed
--skipZeros -o $projPath/peakCalling/SEACR/${histName}_SEACR.mat.gz -p $cores
-a 3000 -b 3000 --referencePoint center

plotHeatmap -m $projPath/peakCalling/SEACR/${histName}_SEACR.mat.gz -out
$projPath/peakCalling/SEACR/${histName}_SEACR_heatmap.png --sortUsing sum --
startLabel "Peak Start" --endLabel "Peak End" --xAxisLabel "" --regionsLabel
"Peaks" --samplesLabel "${histName}"

##== change “histName” and generate heatmaps for peak calling files ==##
histName=IRP1A_PG_rep2
```

```
##== Because only 7 peaks are called in IRP1A_FB_rep1 when compared with
control, use "_seacr_top0.01.peaks.stringent.bed" to generate heat map of
peaks for fat body samples
histName=IRP1A_FB_rep1

cores=6
awk '{split($6, summit, ":"); split(summit[2], region, "-"); print
summit[1]"\t"region[1]"\t"region[2]}'
$projPath/peakCalling/SEACR/${histName}_seacr_top0.01.peaks.stringent.bed >$p
rojPath/peakCalling/SEACR/${histName}_seacr_top0.01.peaks.summitRegion.bed

computeMatrix reference-point -S
$projPath/alignment/bigwig/${histName}_raw.bw -R
$projPath/peakCalling/SEACR/${histName}_seacr_top0.01.peaks.summitRegion.bed
--skipZeros -o $projPath/peakCalling/SEACR/${histName}_top0.01_SEACR.mat.gz -
p $cores -a 3000 -b 3000 --referencePoint center

plotHeatmap -m $projPath/peakCalling/SEACR/${histName}_top0.01_SEACR.mat.gz -
out $projPath/peakCalling/SEACR/${histName}_SEACR_top0.01_heatmap.png --
sortUsing sum --startLabel "Peak Start" --endLabel "Peak End" --xAxisLabel ""
--regionsLabel "Peaks" --samplesLabel "${histName}"

##== change "histName" and generate heatmaps for peak calling files ==##
histName=IRP1A_FB_rep2
```

**7. Peak annotation using Homer**

```
## Filter top SEACR peaks ##
##== R command ==##
# Filter the top 2% peaks of "_seacr_control.peaks.stringent.bed" files,
which produces ~500 peaks

dir.create("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR/top0.02peaks_contr
ol")
setwd("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR/top0.02peaks_control")
projPath <- "/Users/minyi/Desktop/CUT_Tag"
sampleList <- c("IRP1A_PG_rep1", "IRP1A_PG_rep2", "IRP1A_FB_rep2")
histList <- c("IRP1A_PG", "IRP1A_FB")

for (hist in sampleList) {
  peakInfo <- read.table(paste0(projPath, "/peakCalling/SEACR/", hist,
"_seacr_control.peaks.stringent.bed"),
                         header = FALSE, fill = TRUE)
    # Sort the data frame based on the total signal in the 4th column in
descending order
```

```
  sorted_data <- peakInfo[order(-peakInfo[, 4]), ]
  # Determine the number of rows to keep (20% of the total rows)
  n_rows_to_keep <- ceiling(0.02 * nrow(sorted_data))
    # Filter the top 0.01% values
  top0.02peaks <- sorted_data[1:n_rows_to_keep, ]
    peakN <- nrow(top0.02peaks)
    write.table(top0.02peaks, paste0(hist, "_", peakN,
"_control_top0.02peaks.bed"), sep = "\t", quote = FALSE, row.names = FALSE)
}


# Filter the top 50% peaks of "_seacr_top0.01.peaks.stringent.bed" files,
which produces ~1000 peaks

dir.create("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR/top0.2peaks_top0.0
1")
setwd("/Users/minyi/Desktop/CUT_Tag/peakCalling/SEACR/top0.2peaks_top0.01")
projPath <- "/Users/minyi/Desktop/CUT_Tag"
sampleList <- c("IRP1A_PG_rep1", "IRP1A_PG_rep2", "IRP1A_FB_rep1",
"IRP1A_FB_rep2")
histList <- c("IRP1A_PG", "IRP1A_FB")

for (hist in sampleList) {
  # Read in the file
  peakInfo <- read.table(paste0(projPath, "/peakCalling/SEACR/", hist,
"_seacr_top0.01.peaks.stringent.bed"),
                         header = FALSE, fill = TRUE)
    # Sort the data frame based on the total signal in the 4th column in
descending order
  sorted_data <- peakInfo[order(-peakInfo[, 4]), ]
    # Determine the number of rows to keep (50% of the total rows)
  n_rows_to_keep <- ceiling(0.5 * nrow(sorted_data))
    # Filter the top 0.01% values
  top0.5peaks <- sorted_data[1:n_rows_to_keep, ]
    peakN <- nrow(top0.5peaks)
  write.table(top0.2peaks, paste0(hist, "_", peakN,
"_top0.01_top0.2peaks.bed"), sep = "\t", quote = FALSE, row.names = FALSE
}


## Peak annotation ##
##== linux command ==##
projPath=/Users/minyi/Desktop/CUT_Tag
export PATH=/Users/minyi/Desktop/CUT_Tag/homer/bin:$PATH

##== Annotate original SEACR "_control" and "_top0.01" files
##== check the chromosome names of SEACR output file using "head" command, if
the chromosome name doesn't have "chr" prefix, add it to each new line by
"sed" command ==##
```

```
mkdir -p $projPath/peakAnnotation/SEACR
mkdir -p $projPath/peakAnnotation/SEACR/InputFiles

for file in $projPath/peakCalling/SEACR/*.peaks.stringent.bed; do
    output_file="$projPath/peakAnnotation/SEACR/InputFiles/$(basename $file)"
    sed 's/^/chr/' $file > $output_file
done


for file in $projPath/peakAnnotation/SEACR/InputFiles/*.peaks.stringent.bed;
do
      output_file="$projPath/peakAnnotation/SEACR/$(basename $file).txt"
      annotatePeaks.pl $file dm6   > $output_file
done
```

## A 11 Genes enriched in GO terms from PG samples

| PG CUT & Tag GO gene list | | |
|---|---|---|
| **Regulation of DNA-templated transcription** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0024244 | drumstick | drm |
| FBgn0288433 | Lobe | L |
| FBgn0262139 | trachealess | trh |
| FBgn0000577 | engrailed | en |
| FBgn0020307 | defective proventriculus | dve |
| FBgn0000251 | caudal | cad |
| FBgn0014143 | crocodile | croc |
| FBgn0001978 | shuttle craft | stc |
| FBgn0020378 | Sp1 | Sp1 |
| FBgn0026411 | LIM homeobox 1 | Lim1 |
| FBgn0003499 | stripe | sr |
| FBgn0002985 | odd skipped | odd |
| FBgn0039946 | (A+T)-stretch binding protein | ATbp |
| FBgn0287186 | scarecrow | scro |
| FBgn0040318 | HGTX | HGTX |
| FBgn0000179 | bifid | bi |
| FBgn0028789 | Dorsocross1 | Doc1 |
| FBgn0000659 | fork head | fkh |
| FBgn0085424 | nubbin | nub |
| FBgn0003117 | pannier | pnr |
| FBgn0031613 | Heterochromatin protein 6 | HP6 |
| FBgn0000439 | Deformed | Dfd |
| FBgn0031375 | earmuff | erm |
| FBgn0001077 | fushi tarazu | ftz |
| | | |
| **Segmentation** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0024244 | drumstick | drm |
| FBgn0015589 | APC-like | Apc |
| FBgn0004956 | unpaired 1 | upd1 |
| FBgn0000577 | engrailed | en |
| FBgn0020307 | defective proventriculus | dve |
| FBgn0000251 | caudal | cad |
| FBgn0014143 | crocodile | croc |
| FBgn0024836 | starry night | stan |
| FBgn0002985 | odd skipped | odd |
| FBgn0003117 | pannier | pnr |
| FBgn0000439 | Deformed | Dfd |
| FBgn0001077 | fushi tarazu | ftz |
| | | |
| **Nucleus** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0063670 | - | CG40228 |
| FBgn0032439 | RNA and export factor binding protein 2 | Ref2 |
| FBgn0033015 | d4 | d4 |
| FBgn0026083 | twenty-four | tyf |
| FBgn0032988 | Tif-IA | Tif-IA |
| FBgn0262139 | trachealess | trh |

| | | |
|---|---|---|
| FBgn0000577 | engrailed | en |
| FBgn0020307 | defective proventriculus | dve |
| FBgn0000251 | caudal | cad |
| FBgn0014143 | crocodile | croc |
| FBgn0001978 | shuttle craft | stc |
| FBgn0020378 | Sp1 | Sp1 |
| FBgn0026411 | LIM homeobox 1 | Lim1 |
| FBgn0002985 | odd skipped | odd |
| FBgn0039972 | Meiosis regulator and mRNA stability factor 1 | Marf1 |
| FBgn0287186 | scarecrow | scro |
| FBgn0040318 | HGTX | HGTX |
| FBgn0000179 | bifid | bi |
| FBgn0022288 | lethal (2) 09851 | l(2)09851 |
| FBgn0020660 | eukaryotic translation initiation factor 4B | eIF4B |
| FBgn0028789 | Dorsocross1 | Doc1 |
| FBgn0000659 | fork head | fkh |
| FBgn0003480 | spindle B | spn-B |
| FBgn0085424 | nubbin | nub |
| FBgn0003117 | pannier | pnr |
| FBgn0031613 | Heterochromatin protein 6 | HP6 |
| FBgn0000439 | Deformed | Dfd |
| FBgn0031375 | earmuff | erm |
| FBgn0001077 | fushi tarazu | ftz |
| | | |
| **DNA-binding transcription factor activity, RNA Pol II-specific** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0262139 | trachealess | trh |
| FBgn0000577 | engrailed | en |
| FBgn0020307 | defective proventriculus | dve |
| FBgn0000251 | caudal | cad |
| FBgn0014143 | crocodile | croc |
| FBgn0001978 | shuttle craft | stc |
| FBgn0020378 | Sp1 | Sp1 |
| FBgn0026411 | LIM homeobox 1 | Lim1 |
| FBgn0003499 | stripe | sr |
| FBgn0002985 | odd skipped | odd |
| FBgn0039946 | (A+T)-stretch binding protein | ATbp |
| FBgn0287186 | scarecrow | scro |
| FBgn0040318 | HGTX | HGTX |
| FBgn0000179 | bifid | bi |
| FBgn0028789 | Dorsocross1 | Doc1 |
| FBgn0000659 | fork head | fkh |
| FBgn0085424 | nubbin | nub |
| FBgn0003117 | pannier | pnr |
| FBgn0000439 | Deformed | Dfd |
| FBgn0031375 | earmuff | erm |
| FBgn0001077 | fushi tarazu | ftz |
| | | |
| **Homeodomain** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0000439 | Deformed | Dfd |
| FBgn0040318 | HGTX | HGTX |
| FBgn0026411 | LIM homeobox 1 | Lim1 |

| | | |
|---|---|---|
| FBgn0000251 | caudal | cad |
| FBgn0020307 | defective proventriculus | dve |
| FBgn0000577 | engrailed | en |
| FBgn0001077 | fushi tarazu | ftz |
| FBgn0085424 | nubbin | nub |
| FBgn0287186 | scarecrow | scro |
| | | |
| **Zinc finger C2H2-type/integrase DNA-binding domain** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0039946 | (A+T)-stretch binding protein | ATbp |
| FBgn0288433 | Lobe | L |
| FBgn0020378 | Sp1 | Sp1 |
| FBgn0024244 | drumstick | drm |
| FBgn0031375 | earmuff | erm |
| FBgn0002985 | odd skipped | odd |
| FBgn0003499 | stripe | sr |
| FBgn0029939 | Chronophage | Cph |
| | | |
| **Helix-turn-helix motif** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0040318 | HGTX | HGTX |
| FBgn0000251 | caudal | cad |
| FBgn0000577 | engrailed | en |
| **Note**: orange: biological process; purple: cell component; blue: molecular function; green: InterPro protein domain. | | |

## A 12 Genes enriched in GO terms from FB samples

| FB CUT & Tag GO gene list | | |
|---|---|---|
| Negative regulation of transcription by RNA polymerase II | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0283451 | broad | br |
| FBgn0037802 | Sirtuin 6 | Sirt6 |
| FBgn0010825 | Grunge | Gug |
| FBgn0000097 | anterior open | aop |
| FBgn0003507 | serpent | srp |
| FBgn0001978 | shuttle craft | stc |
| FBgn0035357 | MEP-1 | MEP-1 |
| FBgn0022764 | Sin3A | Sin3A |
| FBgn0262519 | Mi-2 | Mi-2 |
| FBgn0003396 | schnurri | shn |
| FBgn0024250 | brinker | brk |
| FBgn0014343 | mirror | mirr |
| FBgn0285917 | scribbler | sbb |
| FBgn0033782 | sugarbabe | sug |
| FBgn0000568 | Ecdysone-induced protein 75B | Eip75B |
| FBgn0027951 | Metastasis associated 1-like | MTA1-like |
| FBgn0035625 | Blimp-1 | Blimp-1 |
| FBgn0043364 | cabut | cbt |
| FBgn0000546 | Ecdysone receptor | EcR |
| FBgn0003870 | tramtrack | ttk |
| FBgn0031613 | Heterochromatin protein 6 | HP6 |
| | | |
| **Chromatin organization** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0004861 | polyhomeotic proximal | ph-p |
| FBgn0040071 | taranis | tara |
| FBgn0087008 | enhancer of yellow 3 | e(y)3 |
| FBgn0036398 | upSET | upSET |
| FBgn0002643 | mastermind | mam |
| FBgn0030945 | Inhibitor of growth family, member 3 | Ing3 |
| FBgn0262519 | Mi-2 | Mi-2 |
| FBgn0040075 | reptin | rept |
| FBgn0261885 | osa | osa |
| FBgn0002775 | male-specific lethal 3 | msl-3 |
| FBgn0053862 | His2A:CG33862 | His2A:CG33862 |
| FBgn0053808 | His2A:CG33808 | His2A:CG33808 |
| FBgn0053882 | His2B:CG33882 | His2B:CG33882 |
| FBgn0003862 | trithorax | trx |
| FBgn0260936 | scrawny | scny |
| FBgn0027620 | ATP-dependent chromatin assembly factor large subunit | Acf |
| FBgn0031613 | Heterochromatin protein 6 | HP6 |
| FBgn0000042 | Actin 5C | Act5C |
| FBgn0283657 | Tousled-like kinase | Tlk |
| | | |
| **Response to ecdysone** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0283451 | broad | br |

| | | |
|---|---|---|
| FBgn0004861 | polyhomeotic proximal | ph-p |
| FBgn0040075 | reptin | rept |
| FBgn0000568 | Ecdysone-induced protein 75B | Eip75B |
| FBgn0035625 | Blimp-1 | Blimp-1 |
| FBgn0000546 | Ecdysone receptor | EcR |
| | | |
| **Response to phypoxia** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0038295 | Guanylyl cyclase at 88E | Gyc88E |
| FBgn0001224 | Heat shock protein 23 | Hsp23 |
| FBgn0013277 | Heat shock protein 70 Ba | Hsp70Ba |
| FBgn0013276 | Heat shock protein 70 Ab | Hsp70Ab |
| FBgn0051354 | Heat shock protein 70 Bbb | Hsp70Bbb |
| FBgn0013278 | Heat shock protein 70 Bb | Hsp70Bb |
| FBgn0267487 | Protein tyrosine phosphatase 61F | Ptp61F |
| | | |
| **Brahma complex** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0033015 | d4 | d4 |
| FBgn0087008 | enhancer of yellow 3 | e(y)3 |
| FBgn0261885 | osa | osa |
| FBgn0000042 | Actin 5C | Act5C |
| | | |
| **Transcription regulator complex** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0011836 | TBP-associated factor 2 | Taf2 |
| FBgn0003507 | serpent | srp |
| FBgn0085432 | pangolin | pan |
| FBgn0002643 | mastermind | mam |
| FBgn0284249 | Adh transcription factor 1 | Adf1 |
| FBgn0035357 | MEP-1 | MEP-1 |
| FBgn0022764 | Sin3A | Sin3A |
| FBgn0262519 | Mi-2 | Mi-2 |
| FBgn0003396 | schnurri | shn |
| FBgn0038642 | Mucin 91C | Muc91C |
| FBgn0035145 | Mediator complex subunit 14 | MED14 |
| FBgn0027951 | Metastasis associated 1-like | MTA1-like |
| FBgn0000546 | Ecdysone receptor | EcR |
| FBgn0003870 | tramtrack | ttk |
| FBgn0003415 | skuld | skd |
| FBgn0001297 | kayak | kay |
| | | |
| **Histone deacetylase complex** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0036398 | upSET | upSET |
| FBgn0035357 | MEP-1 | MEP-1 |
| FBgn0022764 | Sin3A | Sin3A |
| FBgn0262519 | Mi-2 | Mi-2 |
| FBgn0040075 | reptin | rept |
| FBgn0027951 | Metastasis associated 1-like | MTA1-like |
| | | |
| **DNA-binding transcription factor activity, RNA polymerase II-specific** | | |
| **Gene ID** | **Name** | **Symbol** |

| | | |
|---|---|---|
| FBgn0005612 | Sox box protein 14 | Sox14 |
| FBgn0000097 | anterior open | aop |
| FBgn0000567 | Ecdysone-induced protein 74EF | Eip74EF |
| FBgn0261647 | AXIN1 up-regulated 1 | Axud1 |
| FBgn0003507 | serpent | srp |
| FBgn0085432 | pangolin | pan |
| FBgn0001978 | shuttle craft | stc |
| FBgn0004914 | Hepatocyte nuclear factor 4 | Hnf4 |
| FBgn0284249 | Adh transcription factor 1 | Adf1 |
| FBgn0000492 | Drop | Dr |
| FBgn0027339 | jim | jim |
| FBgn0283451 | broad | br |
| FBgn0035357 | MEP-1 | MEP-1 |
| FBgn0264005 | H6-like-homeobox | Hmx |
| FBgn0003396 | schnurri | shn |
| FBgn0267978 | apterous | ap |
| FBgn0024250 | brinker | brk |
| FBgn0014343 | mirror | mirr |
| FBgn0287186 | scarecrow | scro |
| FBgn0033010 | Atf6 | Atf6 |
| FBgn0024288 | Sox100B | Sox100B |
| FBgn0033782 | sugarbabe | sug |
| FBgn0000568 | Ecdysone-induced protein 75B | Eip75B |
| FBgn0035625 | Blimp-1 | Blimp-1 |
| FBgn0043364 | cabut | cbt |
| FBgn0016076 | vrille | vri |
| FBgn0000546 | Ecdysone receptor | EcR |
| FBgn0020309 | crooked legs | crol |
| FBgn0022935 | D19A | D19A |
| FBgn0003651 | seven up | svp |
| FBgn0003870 | tramtrack | ttk |
| FBgn0001297 | kayak | kay |
| FBgn0022699 | D19B | D19B |
| | | |
| **Metal ion binding** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0034356 | Phosphoenolpyruvate carboxykinase 2 | Pepck2 |
| FBgn0011836 | TBP-associated factor 2 | Taf2 |
| FBgn0010247 | Poly-(ADP-ribose) polymerase | Parp |
| FBgn0033015 | d4 | d4 |
| FBgn0000100 | Ribosomal protein LP0 | RpLP0 |
| FBgn0051216 | Nicotinamide amidase | Naam |
| FBgn0004861 | polyhomeotic proximal | ph-p |
| FBgn0039848 | Lysyl oxidase-like 1 | Loxl1 |
| FBgn0263601 | mind bomb 1 | mib1 |
| FBgn0003507 | serpent | srp |
| FBgn0013772 | Cytochrome P450 6a8 | Cyp6a8 |
| FBgn0001978 | shuttle craft | stc |
| FBgn0004914 | Hepatocyte nuclear factor 4 | Hnf4 |
| FBgn0033005 | - | CG3107 |
| FBgn0267408 | Aldehyde oxidase 1 | AOX1 |
| FBgn0041337 | Cytochrome P450 309a2 | Cyp309a2 |
| FBgn0013348 | Troponin C at 41C | TpnC41C |

| | | |
|---|---|---|
| FBgn0027580 | Pyruvate carboxylase | PCB |
| FBgn0262519 | Mi-2 | Mi-2 |
| FBgn0003396 | schnurri | shn |
| FBgn0023535 | arginase | arg |
| FBgn0036451 | Helicase with zinc finger | Helz |
| FBgn0053653 | Calcium-dependent secretion activator | Cadps |
| FBgn0035295 | Centrobin | Cnb |
| FBgn0010328 | without children | woc |
| FBgn0004860 | polyhomeotic distal | ph-d |
| FBgn0010300 | brain tumor | brat |
| FBgn0033504 | CAP | CAP |
| FBgn0031450 | Hepatocyte growth factor regulated tyrosine kinase substrate | Hrs |
| FBgn0031107 | HECT and RLD domain containing protein 2 | HERC2 |
| FBgn0260003 | Dystrophin | Dys |
| FBgn0000568 | Ecdysone-induced protein 75B | Eip75B |
| FBgn0038271 | Ubiquinol-cytochrome c reductase core protein 1 | UQCR-C1 |
| FBgn0027951 | Metastasis associated 1-like | MTA1-like |
| FBgn0003862 | trithorax | trx |
| FBgn0000546 | Ecdysone receptor | EcR |
| FBgn0022935 | D19A | D19A |
| FBgn0003651 | seven up | svp |
| FBgn0013733 | short stop | shot |
| FBgn0036999 | iso Glutaminyl cyclase | isoQC |
| FBgn0033868 | Sperm-Leucylaminopeptidase 7 | S-Lap7 |
| FBgn0002566 | light | lt |
| FBgn0022699 | D19B | D19B |
| FBgn0260866 | defense repressor 1 | dnr1 |
| FBgn0086378 | Apoptosis-linked gene-2 | Alg-2 |
| | | |
| Zinc finger, FYVE/PHD-type | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0027620 | ATP-dependent chromatin assembly factor large subunit | Acf |
| FBgn0031450 | Hepatocyte growth factor regulated tyrosine kinase substrate | Hrs |
| FBgn0030945 | Inhibitor of growth family, member 3 | Ing3 |
| FBgn0262519 | Mi-2 | Mi-2 |
| FBgn0039559 | Nuclear receptor binding SET domain protein | NSD |
| FBgn0266756 | bitesize | btsz |
| FBgn0033015 | d4 | d4 |
| FBgn0087008 | enhancer of yellow 3 | e(y)3 |
| FBgn0003862 | trithorax | trx |
| FBgn0039955 | - | CG41099 |
| FBgn0036398 | upSET | upSET |
| | | |
| **Hemocyanin** | | |
| **Gene ID** | **Name** | **Symbol** |
| FBgn0000639 | Fat body protein 1 | Fbp1 |
| FBgn0002563 | Larval serum protein 1 beta | Lsp1beta |
| FBgn0002564 | Larval serum protein 1 gamma | Lsp1gamma |
| FBgn0261363 | Prophenoloxidase 3 | PPO3 |
| | | |
| **Steroid hormone receptor** | | |

| Gene ID | Name | Symbol |
|---|---|---|
| FBgn0000546 | Ecdysone receptor | EcR |
| FBgn0000568 | Ecdysone-induced protein 75B | Eip75B |
| FBgn0004914 | Hepatocyte nuclear factor 4 | Hnf4 |
| FBgn0003651 | seven up | svp |
| **Note**: orange: biological process; purple: cell component; blue: molecular function; green: InterPro protein domain. | | |

## A 13 Reagent recipes for TurboID

**RIPA lysis buffer**

Buffer ingredients (50 mM Tris, 150 mM NaCl, 0.1% (wt/vol) SDS, 0.5% (wt/vol) sodium deoxycholate, 1% (vol/vol) Triton X-100) were combined in Mili-Q water and the pH was adjusted to 7.5. This solution was stored at 4 °C for many months.

**Protein loading buffer (6×)**

Buffer ingredients (0.33 M Tris (pH 6.8), 34% (vol/vol) glycerol, 10% (wt/vol) SDS, 0.09% (wt/vol) DTT, 0.12% (wt/vol) bromophenol blue) were combined in Mili-Q water and the solution was divided in 1-mL aliquots. Aliquots were stored at −20 °C for several months.

**TBST (1×)**

20× TBST was dissolved in Mili-Q water (final concentration of TBST is 0.1% (vol/vol)). This solution was stored at room temperature for several months.

**KCl (1 M)**

KCl was dissolved in Mili-Q water to a final concentration of 1 M. This was stored at room temperature for several months.

**$Na_2CO_3$ (0.1 M)**

Sodium carbonate ($Na_2CO_3$) was dissolved in Mili-Q water to a final concentration of 0.1 M. This solution was made fresh before use.

**Urea (2 M) in 10 mM Tris-HCl (pH 8.0)**

Urea was dissolved in 10 mM Tris-HCl (pH 8.0) to a final concentration of 2 M. This solution was made fresh before use.

## A 14 R codes for MS data analysis.

Identify TurboID target proteins: perform t-test between experimental and control groups, if $p<0.05$, and the score in experimental group is higher than 2 folds of that in controls, that protein is considered as a target candidate.

```
# R commands #
install.packages("readxl")
install.packages("openxlsx")
library(readxl)
library(openxlsx)

setwd("/Users/yanminyi/Desktop/KKJ_Lab/TurboID/MS_TurboID/phm_TurboID/Targ
etList_R")
projPath <-
"/Users/yanminyi/Desktop/KKJ_Lab/TurboID/MS_TurboID/phm_TurboID"
sampleList <- c("IRP1A_Bio", "IRP1A", "IRP1B_Bio", "IRP1B")
controls <- c("EGFP_Bio", "EGFP")

for (sample in sampleList) {
  exp_data <- read_excel(paste0(projPath, "/inputFiles_R/", sample,
".xlsx"))
  exp_proteins <- exp_data$Entry
  exp_scores <- exp_data$Score

  for (control in controls) {
    control_data <- read_excel(paste0(projPath, "/inputFiles_R/", control,
".xlsx"))
    control_proteins <- control_data$Entry
    control_scores <- control_data$Score

    threshold <- 0.05

    target_list <- c()

    common_proteins <- intersect(exp_proteins, control_proteins)
    for (protein in common_proteins) {
        exp_score <- exp_scores[exp_proteins == protein]
        control_score <- control_scores[control_proteins == protein]

        if (length(exp_score) >= 2 & length(control_score) >= 2) {
          result <- t.test(exp_score, control_score)

          if (result$p.value < threshold & exp_score > 2 * control_score)
{
```

```
            target_list <- c(target_list, protein)
          }
        }
      }

      unique_proteins <- setdiff(exp_proteins, control_proteins)
      target_list <- c(target_list, unique_proteins)

      write.table(data.frame(Enrty = target_list), file = paste0(sample,
"_vs_", control, ".txt"),
                  sep = "\t", quote = FALSE, col.names = TRUE, row.names =
FALSE)
   }
}
```