Insights into RNA metabolism in *Drosophila* nervous system development and virus infection



Jeffrey Y Lee

St John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2022

Abstract

Messenger RNA (mRNA) is an integrated component of core cellular function. mRNA holds the capacity to relay genetic information as templates, localise to different spatial compartments within a cell, and bind combinatorially to effector molecules, such as RNA-binding proteins (RBPs). Post-transcriptional control of mRNA metabolism builds on transcriptional cues to provide a highly complex and context-specific network of gene expression patterns. In this thesis, I describe two cases illustrating the unique importance of RNA metabolism in different aspects of biological processes.

Chapter 3 details experiments in which I explored post-transcriptional regulation as a critical player of neurogenesis. In *Drosophila,* a conserved RBP Syncrip (Syp) controls neural stem cell behaviour and progeny fate. I identified the mRNA targets of Syp, and found an enrichment of binding to transcripts encoding transcription factors and RBPs, which are key regulatory nucleic acid binding proteins. Syp targets contained many transcripts that have isoforms with long 3' untranslated regions. This illustrates the role of Syp to influence the expression of key regulatory proteins through *cis*-regulatory elements in their transcripts. I then functionally characterised the post-transcriptional regulation of *prospero* (*pros*) and *e93*, Syp targets as examples of the importance of Syp in regulating various aspects of neural development.

Chapter 4 expands on the breadth of post-transcriptional regulation in the nervous system. In collaboration with Dr Joshua Titlow, Dr Ana Palanca and Maria Kiourlappou, I applied a highly sensitive single-molecule fluorescence *in situ* hybridisation (smFISH) screen to profile transcription, mRNA accumulation and translation of 200 genes in various larval nervous system cell types. In the process, we demonstrate the common lack of correlation (discordance) between the distribution of mRNA and the encoded protein, a manifestation of post-transcriptional control. Our results suggest that mRNA stability may be a common mechanism to achieve cell-type specific protein expression, and that localised mRNAs are frequently observed in synaptic compartments. Our survey provides new insight into gene expression control in the brain and raises the need to improve further our knowledge of mRNA biology for understanding neurodevelopment and neural functions.

Chapter 5 describes the utility of the same single-molecule and high-resolution microscopy methods used in Chapter 4 to dissect RNA biology of virus infection. In response to the current COVID-19 pandemic, I used smFISH to detect and profile replication kinetics of SARS-CoV-2 RNA during the first hours of the infection. This result represents the first detailed molecular characterisation of SARS-CoV-2 RNA dynamics and provides new insight into the variable

susceptibility of individual cells to support virus replication. Furthermore, I show that the B.1.1.7 Variant of Concern (VOC) displays delayed replication kinetics than the ancestral strain, and discuss possible implications for its enhanced dissemination in the human population. Collectively, my thesis emphasises the diverse and unique role of RNA biology in nervous system development and virus infection. I anticipate that this thesis will be a valuable resource for future works to improve our understanding of post-transcriptional control in a diverse range of fields.

Declaration

I hereby declare that this thesis is my own work and the experiments described in the following pages were performed by myself, unless otherwise stated. The experiments were carried out in the Department of Biochemistry, University of Oxford. This thesis has not been submitted for any other degree or professional qualifications.

Jeffrey Y Lee
April 2022

Acknowledgements

My deepest gratitude goes to my supervisor, Ilan Davis, for his unwavering guidance and support throughout my training. Much of this work was only possible because you have created an open environment where enthusiasm for research is always thriving, even during the difficult times of the pandemic.

I am extremely grateful to Alfredo Castello and Jane McKeating for their constant encouragement and a warm invitation to the world of Virology.

I have had the privilege of learning from Darragh Ennis, Joshua Titlow, Mary Kay Thompson, Richard Parton, Francesca Robertson, Aino Jarvelin, Tamsin Samuels, Dalia Gala, Maria Kiourlappou, and Xuan Wang. Thank you for providing much needed positivity, laughter, friendship, and sustenance. I am lucky to have spent my DPhil in such a supportive and intellectually stimulating environment.

To my collaborators, Daniel Agranoff, Peter Wing, Marko Noerenberg, Senko Tsukada, and Matt Stevenson, it has been a pleasure to work towards the understanding of virus replication.

I gratefully acknowledge the Clarendon Fund and St John's College for funding my studies, and special thanks go to Chris Doe, Andrea Brand, James Skeath, and Tzumin Lee for fly stocks and reagents.

Lastly, I would like to thank my family, as without them nothing would have been possible. Thank you for everything.

Contents

Abstract ii

Declaration iv

Acknowledgements v

List of Abbreviations vii

1 Introduction 1

1.1 mRNA metabolism in the *Drosophila* nervous system . . . 2

1.1.1 *Drosophila* neural stem cells as a model of neurogenesis . . . 2

1.1.2 Regulation of neurogenesis . . . 6

1.1.3 Methods for studying RBP targets and their regulations . . . 15

1.1.4 Objectives . . . 18

1.2 RNA metabolism during SARS-CoV-2 infection . . . 19

1.2.1 Lessons from Virology . . . 19

1.2.2 Structure of SARS-CoV-2 genome . . . 20

1.2.3 Infection lifecycle of SARS-CoV-2 . . . 21

1.2.4 Replication compartments of SARS-CoV-2 . . . 24

1.2.5 Methods for studying viral replication dynamics . . . 25

1.2.6 Objectives . . . 26

2 Materials and Methods 28

2.1 Fly stocks and cell culture . . . 28

2.1.1 Fly stocks . . . 28

2.1.2 Cell culture . . . 28

2.1.3 Virus propagation and infection . . . 29

2.2 Microscopy and Imaging . . . 30

2.2.1 Single-molecule fluorescence *in situ* hybridisation (smFISH) . . . 30

2.2.2 smFISH probe design and specificity analysis . . . 32

2.2.3 Immunofluorescence . . . 33

2.2.4 Microscopy . . . 33
2.3 Biochemical techniques . . . 34
2.3.1 iCLIP . . . 34
2.3.2 Western blot . . . 36
2.3.3 RT-qPCR . . . 37
2.4 Image analysis . . . 38
2.4.1 Cell segmentation . . . 38
2.4.2 smFISH quantification . . . 38
2.4.3 Fluorescence intensity quantification . . . 40
2.4.4 Simulating spatial distribution of cells . . . 41
2.4.5 Calculating RNA spatial distribution index . . . 42
2.5 Bioinformatics and statistics . . . 42
2.5.1 iCLIP analysis . . . 42
2.5.2 RNA-sequencing analysis . . . 42
2.5.3 Statistics, data wrangling and visualisation . . . 44
2.6 Other methods . . . 44
2.6.1 Neuroblast clonal analysis . . . 44
2.6.2 Larval locomotion assay . . . 44

3 Characterisation of RNA targets of Syncrip 46
3.1 Introduction . . . 46
3.1.1 Specific aims of this chapter . . . 48
3.1.2 Experimental rationale . . . 48
3.2 Identification of Syp targets . . . 50
3.2.1 Expression of Syp in the developing larvae . . . 50
3.2.2 Adaptation of iCLIP to samples from larval tissues . . . 52
3.2.3 Identification of transcriptome-wide Syp binding sites . . . 59
3.3 General characteristics of Syp targets . . . 65
3.3.1 Syp targets in the brain . . . 65
3.3.2 Syp targets in the whole-larva . . . 73
3.4 Post-transcriptional regulation of Syp-bound transcripts . . . 75
3.4.1 RNA stability regulates cell-type specific expression pattern of Prospero . 75
3.4.2 Syp promotes translation of *e93* for the correct progression of temporal fate patterning . . . 83
3.5 Next steps . . . 89
3.6 Conclusion and discussion . . . 91

4 Investigating the prevalence of post-transcriptional regulation in the developing *Drosophila* nervous system 95

4.1 Introduction . . . 95

4.1.1 Specific aims of this chapter . . . 97

4.1.2 Publication . . . 97

4.2 Investigating gene expression pattern of Syp targets in the larval brain . . . 98

4.2.1 Identification of Syp targets with protein-trap insertions . . . 98

4.2.2 Co-visualisation of protein and RNA using smFISH . . . 99

4.2.3 Prevalence of discordance between RNA and protein expressions of Syp targets . . . 101

4.3 Systematic analysis of post-transcriptional regulation across the nervous system . 103

4.3.1 Selection of 200 YFP-trap collections . . . 103

4.3.2 Systematic image analysis pipeline . . . 105

4.3.3 Post-transcriptional control of NB differentiation . . . 107

4.3.4 Summary of screen results . . . 109

4.4 Conclusion and discussion . . . 112

5 Single-molecule detection of SARS-CoV-2 RNA provides new insights into infection dynamics and variant differences 115

5.1 Introduction . . . 115

5.1.1 Specific aims of this chapter . . . 117

5.1.2 Publication . . . 117

5.2 Visualising SARS-CoV-2 RNA at the single-molecule resolution . . . 118

5.2.1 Sensitive detection of SARS-CoV-2 RNA . . . 118

5.2.2 Specific detection of SARS-CoV-2 RNA . . . 122

5.2.3 SARS-CoV-2 detection in animal model samples . . . 124

5.2.4 dsRNA antibody underestimates SARS-CoV-2 replication . . . 126

5.3 Co-visualising SARS-CoV-2 genomes and sub-genomes . . . 128

5.3.1 Multiplexed smFISH strategy reliably discriminates gRNA and sgRNA . 128

5.3.2 gRNA and sgRNA differentially associate with the viral nucleocapsid protein . . . 130

5.3.3 Negative-strand gRNA requires denaturation for proper detection . . . 131

5.4 Profiling early-stage SARS-CoV-2 replication kinetics . . . 134

5.4.1 Developing an analysis pipeline to automate quantification of smFISH datasets . . . 134

5.4.2 SARS-CoV-2 replication is heterogeneous . . . 136

5.4.3 Relative level of gRNA/sgRNA changes with the infection progression . 141
5.4.4 SARS-CoV-2 gRNA is highly stable 142
5.5 Investigating the emergence of super-permissive cells 145
5.5.1 Super-permissive cells are commonly observed in multiple cell lines . . . 145
5.5.2 Spatial distribution of super-permissive cells 146
5.6 Comparative replication kinetics of the SARS-CoV-2 strains 149
5.6.1 B.1.1.7 replicates slower than the ancestral VIC strain 149
5.6.2 Super-permissive B.1.1.7 cells become clustered at 24 hpi 153
5.6.3 B.1.1.7 phenotype is also observed in A549 cells 155
5.6.4 Orthogonal RNA-seq confirms smFISH results 157
5.7 Application of smFISH in Hepatitis B virus infection 159
5.8 Conclusion and discussion 161
5.8.1 New insight into SARS-CoV-2 early replication events 161
5.8.2 Differential replication kinetics of SARS-CoV-2 variants 163

6 Discussion **165**
6.1 Post-transcriptional regulation of *Drosophila* neurogenesis 165
6.1.1 Relevance to mammalian neurogenesis 167
6.1.2 Future perspectives 168
6.2 RNA replication dynamics of SARS-CoV-2 infection 169
6.2.1 Can smFISH be utilised in clinical settings? 171
6.3 Conclusion 172

Appendices

A List of smFISH probe sequences **175**

B Syp targets in the brain **179**

C Syp targets in the whole larva **181**

D Summary of smFISH protein and RNA screen results (Syp targets) **183**

E Summary of smFISH protein and RNA screen results (200 YFP-traps) **185**

References **187**

List of Abbreviations

3'UTR . . . . . 3' untranslated region

4sU 4-thiouridine

5'UTR . . . . . 5' untranslated region

AEBSF . . . . . 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride

ALH . . . . . after larval hatching

ALNB . . . . . antennal lobe neuroblast

Br . . . . . Broad

Cas . . . . . Castor

CB . . . . . central brain

cDNA . . . . . complementary deoxyribonucleic acid

CDS . . . . . coding sequence

Chinmo . . . . . Chronologically inappropriate morphogenesis (Transcription factor)

ChIP . . . . . chromatin immunoprecipitation

CLIP . . . . . crosslinking and immunoprecipitation

COVID-19 . . . . Coronavirus disease 2019

CRISPR . . . . . clustered regularly interspaced short palindromic repeats

CV . . . . . convoluted vesicle (SARS-CoV replication organelle)

DMSO . . . . . dimethyl sulfoxide

DMV . . . . . double membrane vesicle

DNA . . . . . deoxyribonucleic acid

EcR . . . . . Ecdysone receptor

ELAV . . . . . Embryonic lethal abnormal vision

FBS . . . . . fetal bovine serum

FFPE . . . . . formalin-fixed paraffin-embedded

GFP . . . . . green fluorescent protein

GMC . . . . . ganglion mother cell

GO . . . . . gene ontology

Grh . . . . . Grainyhead

gRNA . . . . . genomc RNA (SARS-CoV-2)

GUI . . . . . graphical user interface

Hb . . . . . Hunchback

HBV . . . . . Hepatitis B virus

iCLIP . . . . . individual nucleotide resolution CLIP

Imp . . . . . IGF-II mRNA binding protein

INP . . . . . intermediate neural progenitor

Insc . . . . . Inscuteable

kb . . . . . kilobases

KH . . . . . hnRNP K homology domain

Kr . . . . . Kruppel

LN2 . . . . . liquid nitrogen

MB . . . . . mushroom body

MBNB . . . . . mushroom body neuroblast

MERS-CoV . . . Middle East Respiratory Syndrome

mRNA . . . . . messenger RNA

N . . . . . Nucleocapsid (SARS-CoV-2)

NB . . . . . neuroblast

ncRNA . . . . . non-coding RNA

NMJ . . . . . neuromuscular junction

NSC . . . . . neural stem cell

OL . . . . . optic lobe

ORF . . . . . open reading frame

OrR . . . . . Oregon-R wild-type *D. melanogaster* strain

PBS . . . . . phosphate buffered saline

PCR . . . . . polymerase chain reaction

pgRNA . . . . . pre-genomic RNA (HBV)

Pros . . . . . Prospero

PSF . . . . . point-spread function

RBP . . . . . RNA-binding protein

RDV . . . . . remdesivir

RIP-seq . . . . . RNA immunoprecipitation sequencing

RNA-seq . . . . . RNA sequencing

RNase . . . . . ribonuclease

RNP . . . . . Ribonucleoprotein

RRM . . . . . RNA recognition motif

RT-qPCR . . . . quantitative reverse transcription PCR

RVC . . . . . ribonucleoside vanadyl complexes

S . . . . . Spike (SARS-CoV-2)

SARS-CoV-2 . . . Severe Acute Respiratory Syndrome Coronavirus 2

scRNA-seq . . . . single-cell RNA sequencing

SDS-PAGE . . . . sodium dodecyl sulphate polyacrylamide gel

sgRNA . . . . . subgenomic RNA (SARS-CoV-2)

smFISH . . . . . single-molecule fluorescence *in situ* hybridisation

Sqd . . . . . Squid

SSC . . . . . sodium saline citrate

Svp . . . . . Sevenup

Syp . . . . . Syncrip, synaptotagmin binding cytoplasmic RNA interacting protein

T2NB . . . . . type 1 neuroblast

TF . . . . . transcription factor

UMI . . . . . unique molecular identifier

UV . . . . . ultraviolet (wavelength)

VNC . . . . . ventral nerve cord

VOC . . . . . variant of concern (SARS-CoV-2)

VP . . . . . vesicle packet

vRNA . . . . . viral RNA

YFP . . . . . yellow fluorescent protein

1 Introduction

One of Britain's greatest scientists, Francis Crick, was the first to suggest the term 'Central Dogma' in 1958, which is now a keystone of molecular biology: the flow of information from genetic code to a functional protein (F. Crick, 1970; F. H. Crick, 1958). For the correct biological outcome, it is critical that each step of this process is tightly regulated. Indeed, multifold trans-acting factors, such as transcription factors (TFs), RNA-binding proteins (RBPs) and kinases, regulate various stages of gene expression via their unique mechanisms of action. Although the regulation of transcription has been well-established to influence many biological processes, post-transcriptional control has only started to gain more attention recently (Buccitelli & Selbach, 2020). The relay of genetic information from DNA to messenger RNA (mRNA) fundamentally depends on transcription. However, the complexity of the life of mRNA can be much greater than that of DNA as it can be edited, transported and form complex structures. The interactions between RNA cis-elements and RBPs, therefore, offer an additional layer of regulation that can be exploited by highly polarised cells, cells requiring rapid response to stimuli, and RNA viruses that hijack host RBPs (Buccitelli & Selbach, 2020; Garcia-Moreno et al., 2019; Z. Liu et al., 2015; Mayr, 2017; Pilaz & Silver, 2015).

In this Chapter, I will introduce two examples of RNA metabolism control: (i) in *Drosophila melanogaster* neurogenesis and (ii) in SARS-CoV-2 infection and viral RNA replication. RNA

exists in various lengths and conformations with or without protein-coding potentials. In this thesis, I mainly focus on mRNAs and viral RNAs (vRNAs) encoding proteins. Investigating RNA regulation requires intricate tools that must be highly sensitive and specific to the target transcript being studied. Similar methodologies can be applied to both systems, offering new insights into the biology of RNA metabolism, which will be the main focus of this thesis.

1.1 mRNA metabolism in the *Drosophila* nervous system

The brain is the most complex organ in many organisms and is responsible for diverse functions to control behaviour, memory and learning. A variety of neurogenesis mechanisms have evolved in order to facilitate neurogenesis at varying scales. A vast number and diversity of neurons are produced from a small population of neural stem cells (NSCs) in order for a fully functioning brain to develop. Tight regulation of NSC behaviour is therefore critical for the correct development of a brain and to prevent dysregulated proliferation leading to pathogenesis.

1.1.1 *Drosophila* neural stem cells as a model of neurogenesis

Mechanisms of neurogenesis

Some organisms, such as *Caenorhabditis elegans* largely employ a 'direct neurogenesis' programme, where an NSC division gives rise to one post-mitotic cell that will mature and take a neural fate (Hobert, 2010) (Figure 1.1A, *left panel*). As the simplest mechanism, such one-to-one genesis does not scale well in organisms with larger nervous systems. *D. melanogaster*, with a more complex central nervous system (CNS), predominately employs the 'indirect neurogenesis' method, in which a *D. melanogaster* NSC, a neuroblast (NB), divides asymmetrically and generates a precursor cell, called a ganglion mother cell (GMC) (Figure 1.1A, *centre panel*). The GMC matures and terminally divides, giving rise to two daughter neurons or glia, depending primarily on the identity of the NB that gave birth to that GMC (Homem & Knoblich, 2012). The most abundant Type I NBs in the fly central brain primarily utilise 'indirect neurogenesis' (Figure 1.1B). On the contrary, Type II NBs generate progenitor cells that can divide multiple times before terminal differentiation (Figure 1.1A, *right panel*). These intermediate neural progenitors (INPs)

(Boone & Doe, 2008) are reminiscent of mammalian outer radial glial cells (D. V. Hansen et al., 2010; Pollen et al., 2015), which maintain stemness and unique temporal identity to generate many more GMCs (Bayraktar et al., 2010; Janssens et al., 2014). Effectively, INPs enable an exponential increase in the NB-to-neuron ratio.

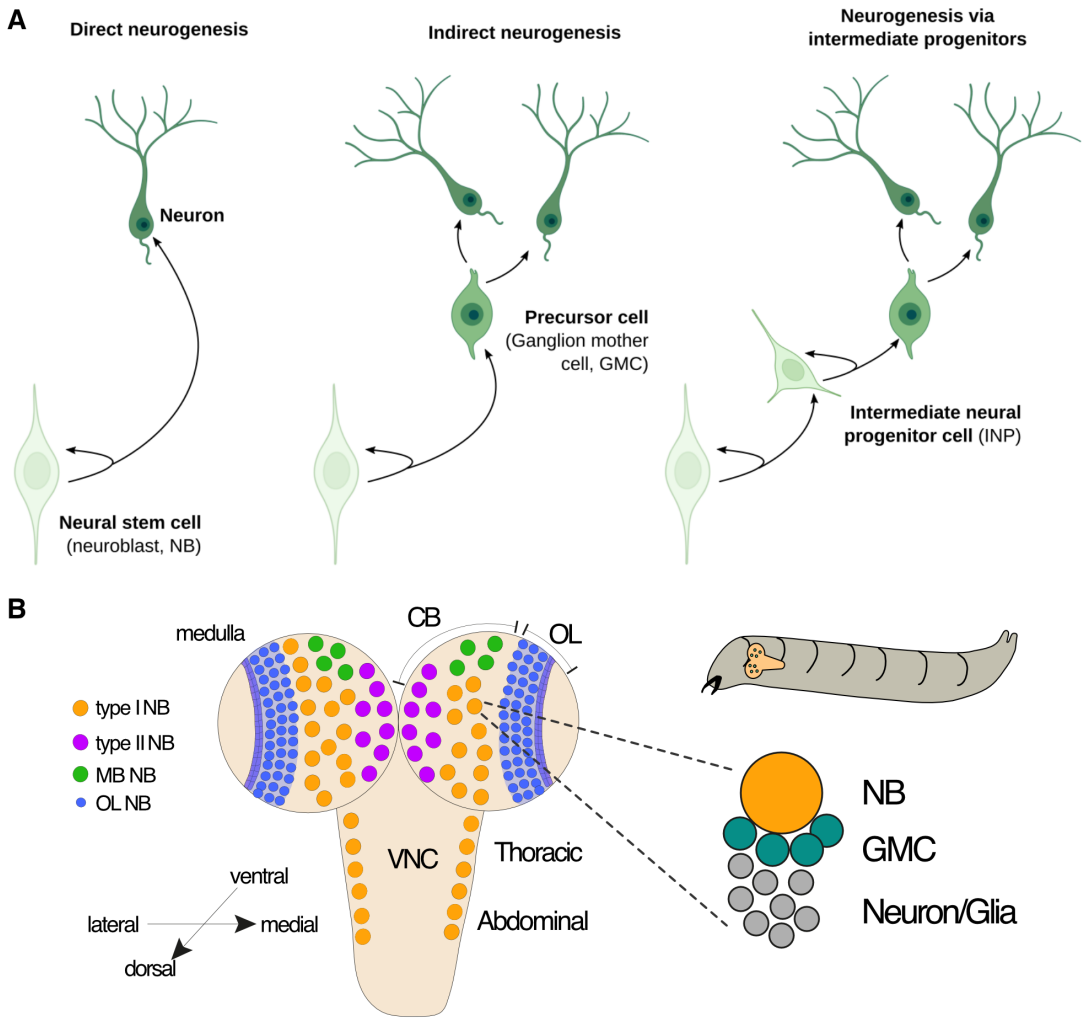


Figure 1.1: Mechanism of neurogenesis. (**A**) Three common mechanisms of neurogenesis. NBs divide asymmetrically to generate a daughter cell destined to differentiate and another to self-renew. In 'direct neurogenesis', a single stem cell division leads to the generation of one neuron. *D. melanogaster* Type I NBs utilise 'indirect neurogenesis', which produces a precursor cell (GMC) that can give rise to two daughter cells. Type II NBs produce intermediate neural progenitors (INPs), which maintain stemness to divide asymmetrically and yield a greater degree of cell amplification. (**B**) Overview of the third instar larval (L3) brain, showing the distribution of different NB lineages: Type I NBs (*orange*), Type II NBs (*magenta*), mushroom body NBs (*green*), and optic lobe NBs (OL, *blue*). A schematic layout of Type I NB lineage is also shown, which is composed of one NBs, 4-5 GMCs and their progeny cells (immature and mature neurons or glia).

Asymmetric neural stem cell division

Most Type I and Type II NB division produces two distinct cell types: a self-renewing NB, and a progeny GMC or INP. These cells are different in size and have very different proliferative potentials. The progeny GMCs and INPs undergo further divisions and exit mitosis entirely. It is crucial that the progeny cells do not inherit the self-proliferating behaviour of stem cells as it can lead to uncontrolled tumour growth (Homem & Knoblich, 2012). On the contrary, NBs must maintain their proliferative fate throughout their division lifetime. The production of different cell fates from a single stem cell division can be achieved through extrinsic factors (e.g. from the local environment), or the segregation of intrinsic determinants. Interestingly, asymmetric NB divisions during post-embryonic neurogenesis do not require extrinsic factors, as they continue to occur normally in isolated brain cultures (Ceron et al., 2006). Therefore, the intrinsic factors required for *Drosophila* NB asymmetric division have been studied in great detail, and the processes involved in establishing cell polarity, localisation of cell fate determinants and segregation of fate determinants are now well understood (Homem & Knoblich, 2012).

The polarity of NB is established when it first delaminates from the ventral neuroectoderm (Knoblich, 2008), and then subsequent NB divisions are aligned relative to the previous division axis via centrosomes (Rebollo et al., 2009). Internally, the cell polarity is established by localisation of the Par complex to the apical cortex of the NBs, which consists of Bazooka (Baz), Par-6, and atypical protein kinase C (aPKC) (Suzuki & Ohno, 2006). The Par complex regulates the spindle assembly and cell cleavage during NB divisions. Baz initiates a cascade of protein bindings on the apical side of the NB membrane through sequential Baz - Inscuteable (Insc) - Pins - Gai - Mud interactions (Knoblich, 2008). Mud, the homolog of NuMA, recruits astral microtubules to one of the spindle poles and anchors it to the apical cortex, thereby establishing an asymmetric spindle formation. Consequently, the cell cleavage site is shifted basally, producing smaller progeny cells, and mutations in any of these proteins lead to dysregulated asymmetric division and mislocalised fate determinants (Kaltschmidt & Brand, 2002).

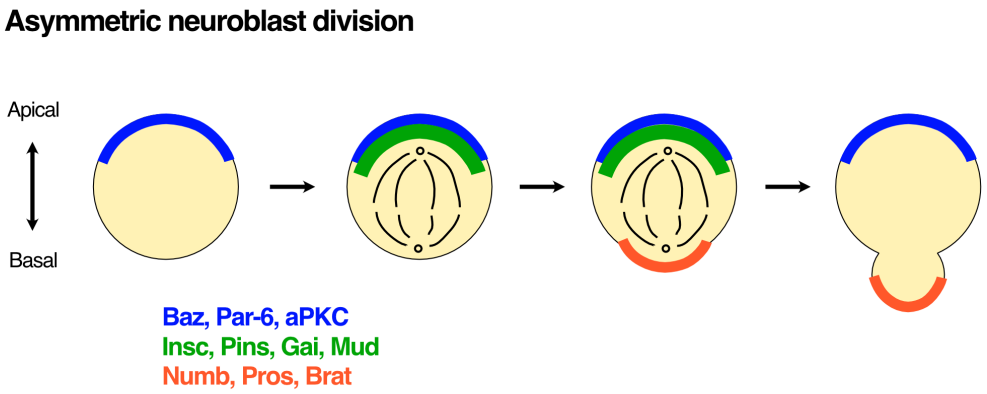


Figure 1.2: Asymmetric segregation during neuroblast division. During neuroblast (NB) division, the Par complex (Baz, Par-6 and aPKC) is assembled on the apical cortex of the cell. Baz then sequentially recruits Insc, Pins, Gai, and Mud, which establish asymmetric spindle pole formation. Numb, Pros and Brat are recruited to the basal side of the cortex so that they are segregated to the progeny cell upon asymmetric division. Figure adapted from (Homem & Knoblich, 2012).

The apical Par complex induces the accumulation of pro-differentiation determinants at the opposite, basal side of the NBs so that they are inherited by the progeny cell. Here, three key fate determinants segregate to the smaller progeny cell: Numb, Prospero (Pros) and Brain tumour (Brat) (Homem & Knoblich, 2012; Knoblich, 2008). The basal localisation of Numb is facilitated by binding to Partner-of-numb (Pon), which is activated through phosphorylation by Polo kinase (H. Wang et al., 2007). The inheritance of Numb by GMCs promotes endocytosis of the Notch receptor and inhibition of Notch signalling in progeny cells. High levels of Notch signalling promotes stem cell fate through Notch downstream targets (e.g. *worniu (wor), deadpan (dpn), klumfuss (klu), myc,* and *grainy head (grh)*) (Zacharioudaki et al., 2016), therefore the Numb-mediated inhibition of Notch promotes differentiation of progeny cells (Berdnik et al., 2002).

Pros is a homeodomain-containing transcription factor, which functions both to repress self-renewal genes and activate pro-differentiation genes (Choksi et al., 2006). In the NBs, Pros protein is asymmetrically segregated through binding to an adapter protein Miranda (Mira) (Chia et al., 2008). Mira itself accumulates to the basal side of the NB through the activity of Lethal giant larvae (Lgl), which is repressed at the apical cortex via phosphorylation by aPKC. Therefore, Mira tethers Pros to the basal NB membrane, preventing Pros activity as a transcription factor (Hirata et al.,

1995; Ikeshima-Kataoka et al., 1997; Spana & Doe, 1995). However, in the progeny cell, Mira is degraded to allow Pros to enter the nucleus (Choksi et al., 2006). Interestingly, *pros* mRNA is also asymmetrically localised by Insc and RNA-binding protein Staufen (Stau) in the embryo (Broadus et al., 1998; P. Li et al., 1997). Loss-of-function of *pros* leads to tumours in larval brains (Bello et al., 2006) or uncontrolled GMC division in the embryonic nervous system (Choksi et al., 2006).

Brat is also segregated into progeny cells through Mira, and it is thought to act together with Pros to promote a GMC fate (Betschinger et al., 2006; C.-Y. Lee et al., 2006). As an RNA-binding protein, it may regulate its targets at a post-transcriptional level, but the precise regulome of Brat is not known. The role of Brat is particularly well-studied in Type 2 NB lineages where *brat* mutants cause tumour growth because INPs revert back to NB fate and over-proliferate (Bello et al., 2006; Homem & Knoblich, 2012). The examples of Numb, Pros and Brat highlight how the mechanism of asymmetric protein localisation can promote neural differentiation. This mode of regulation may be advantageous over the transcriptional control for its rapid effect in daughter cells soon after cell division.

1.1.2 Regulation of neurogenesis

Progenitor patterning as a mechanism of neuronal fate specification

Neurons generated during brain development are not identical clones of each other but adopt diverse molecular, morphological and connectomic features (Kepecs & Fishell, 2014). The combination of these observable characteristics defines a neural 'fate' for each neuron, representing the vast diversity of cell types in the brain. So how are the specific fates determined during development? Various factors play a role in neuronal fate specification. Both NSC identity and developmental time have been observed to influence neuronal fates (Holguera & Desplan, 2018; Homem & Knoblich, 2012; Jessell, 2000; Nadadhur et al., 2018). It is thought that these two factors are inherited by their progeny through the expression of a unique combination of temporal patterning factors that vary across the spatiotemporal axis of the development. In the case of neurogenesis involving INPs, their identity and age are other factors to consider, as the INPs also express their own cascade of neural patterning factors (Bayraktar & Doe, 2013).

Additionally, extrinsic or extracellular signal cues can also influence the fate of differentiating neurons, such as hormones or TGFβ signals (Rossi & Desplan, 2020; Syed et al., 2017). The relative contribution of extrinsic versus intrinsic factors on fate patterning is not clear but this mechanism is undoubtedly part of the broader specification process.

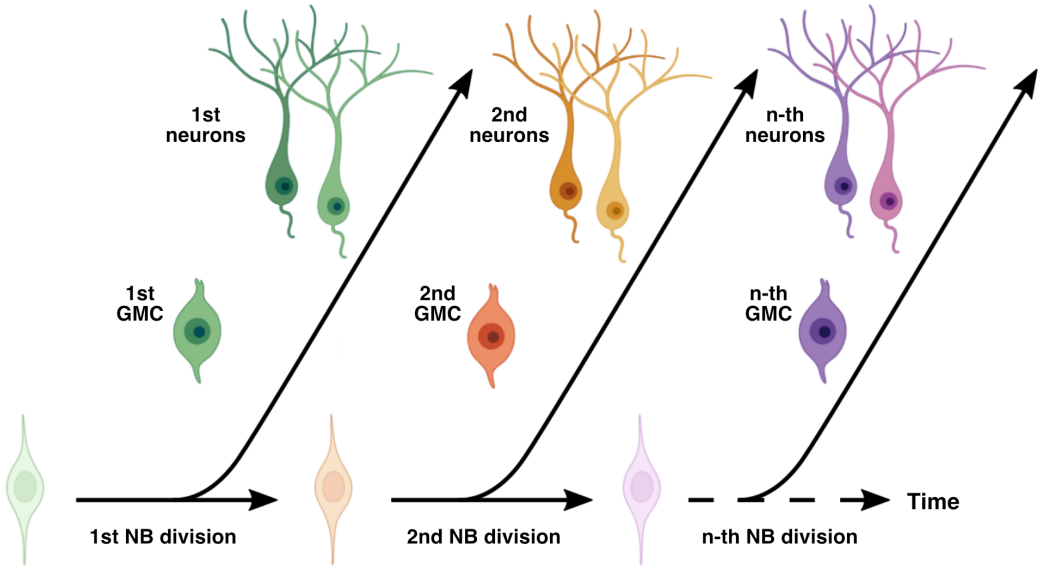


Figure 1.3: Generating neural fate patterning. A simple mechanism of generating neuron diversity in 'indirect neurogenesis'. As the NBs age and undergo asymmetric division cycles, they sequentially express temporal factors defining their age and identity (shown as different colours). This identity is inherited by the GMCs, which then pass it on to the daughter neurons or glial cells. Note that absolute development time is not to scale.

Two waves of neurogenesis

During embryogenesis, NBs are formed through delamination from the neuroectoderm, followed by permanent Notch-mediated lateral inhibition (Skeath & Thor, 2003). NB divisions are then triggered in two main waves, one in the embryo and then in the larval stages. The two waves are separated by a period of stem cell quiescence (Homem & Knoblich, 2012) (Figure 1.4). After delamination, embryonic NBs enter cycles of asymmetric division to generate a reproducible sequence of specific neuron fates (Doe, 2017a). These neurons make up the functional nervous system in the larval stage. Towards the end of embryogenesis, NBs in the abdominal regions undergo apoptosis, while cerebral and thoracic NBs enter quiescence. After hatching, larval

feeding activates insulin signaling pathways to induce re-entry of NBs into cell cycle (Chell & Brand, 2010). This second wave of neurogenesis also generates a specific sequence of neuron fates and is responsible for 90% of the neurons required for the adult fly brain. By pupal stages, most NBs exit the cell division cycle via a process known as decommissioning, which is triggered by a spike in the ecdysone hormone titre (Homem & Knoblich, 2012; C. P. Yang et al., 2017). The CNS connectomes are further remodelled during metamorphosis until eclosion, but new neuron generation becomes limited at this stage.

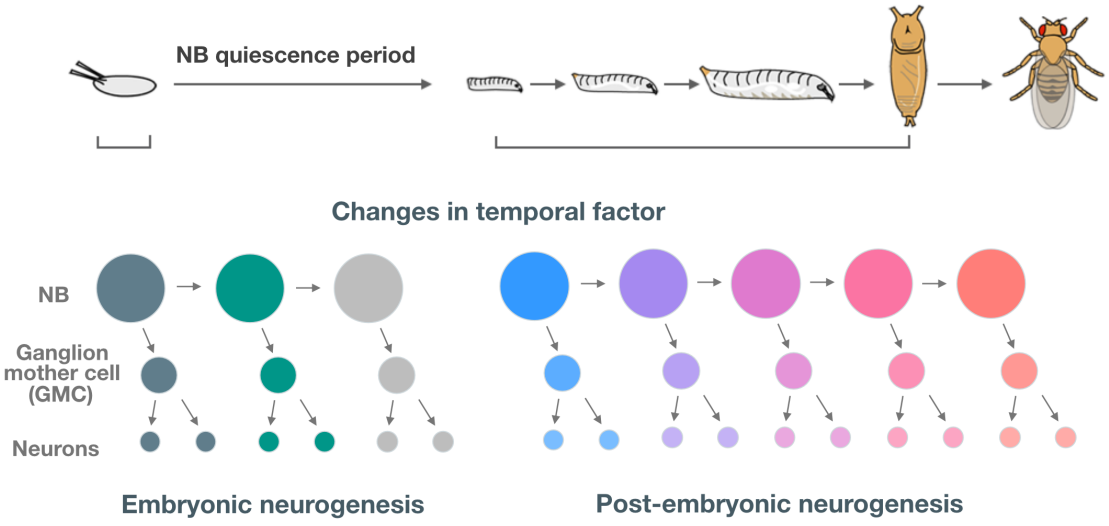


Figure 1.4: Two waves of *Drosophila melanogaster* neurogenesis. In *D. melanogaster* neurogenesis occurs during embryogenesis and in larval stages (post-embryonic), which are separated by a period of NB quiescence. Neuronal fate patterning is predominantly regulated transcriptionally by a cascade of temporal TFs. In post-embryonic neurogenesis, neuron diversity is thought to be generated by the interplay of transcriptional and post-transcriptional regulations.

Transcriptional regulation of neurogenesis

How is the NB identity regulated during the two waves of *D. melanogaster* neurogenesis? In the embryo, the NBs ensure the correct neuronal birth order by sequentially expressing a cascade of TFs. This molecular identity is subsequently inherited by the GMC when it is born. Using a series of cell ablation, transplantation and RNA *in situ* hybridisation experiments, the lineage progression regulated by each TFs has been mapped out in detail (Bayraktar & Doe, 2013; Doe, 2017a; Isshiki et al., 2001; Maurange et al., 2008) (Figure 1.5A). Here, embryonic NBs express overlapping but defined cascade of Hunchback (Hb) > Sevenup (Svp) > Kruppel (Kr) > Pdm > Castor (Cas)

TFs (Doe, 2017a). When the NB divides to produce a GMC, this TF identity is maintained in the progeny, defining the fate of the downstream neurons. It is yet to be discovered how the directionality of the TF cascades is maintained. However, previous studies have described the existence of a feed-forward loop from *hb* to *kr*, and Svp-mediated silencing of *hb* (Grosskortenhaus et al., 2006; Homem & Knoblich, 2012). Possible regulation of the sequence by the number of NB divisions has also been proposed, but this is poorly understood (Homem & Knoblich, 2012). The cascade occurs asynchronously across NBs, indicating the intrinsic nature of the progression.

Similar temporal TF cascades have also been identified in post-embryonic (larval) neurogenesis; mainly in the optic lobe (OL) NB lineages and within the INPs of the Type II lineages (Bayraktar & Doe, 2013; Holguera & Desplan, 2018) (Figure 1.5C-D). However, such a detailed cascade of TFs is yet to be identified in Type I and II NBs themselves. Several earlier studies pointed out the existence of a crude progression of Castor (Cas) + Svp > Broad (Br) > E93 TF series (Maurange et al., 2008; Syed et al., 2017; Wu et al., 2012) (Figure 1.5B). Such a cascade was evident in progeny cells that could be broadly grouped into 'early', 'mid' and 'late'-born neurons. However, the four TFs certainly do not recapitulate the full diversity of the larval neurogenesis programme. To date, the predominant approach to understand the regulation of neurogenesis involved using transcriptome-wide studies to identify key TFs expressions (Doe, 2017a; Isshiki et al., 2001; Z. Liu et al., 2015). These studies were then followed up by mutagenesis of the TFs to understand their function in neuronal fate specifications. Although it has been very powerful to our understanding of the mechanism of neurogenesis, TF activity alone cannot fully describe different pathways leading to gene expression controls (Doe, 2017a; Pilaz & Silver, 2015). Furthermore, it is not clear how these key TFs are regulated, suggesting more complex levels of control may be in place.

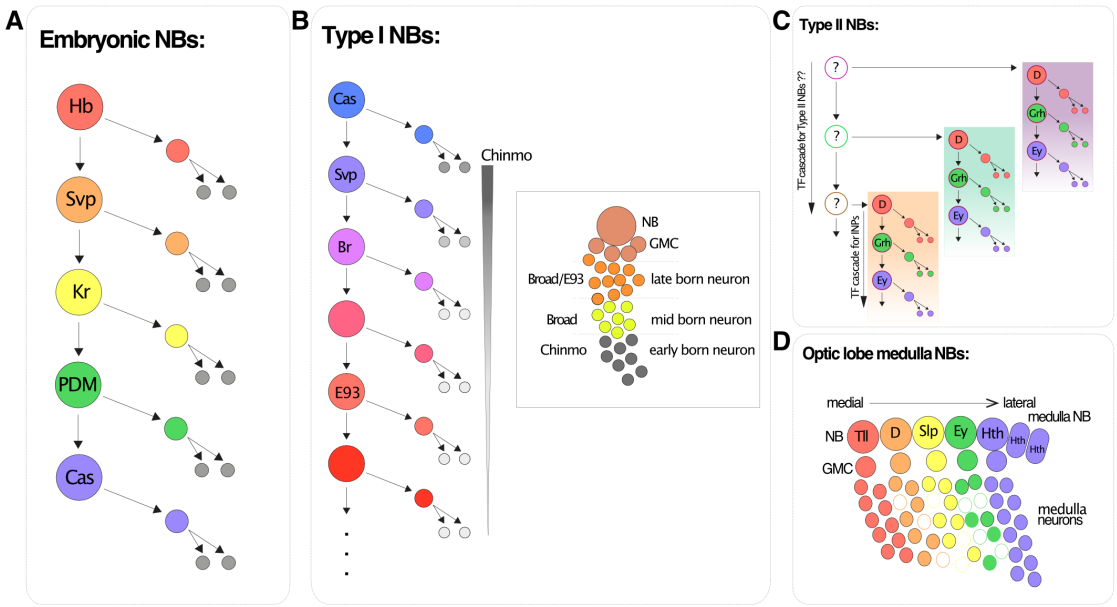


Figure 1.5: Transcriptional regulation of neuron fate determination. (**A**) Embryonic NBs sequentially express the TFs: Hunchback (Hb), Sevenup (Svp), Kruppel (Kr), Pdm, and Castor (Cas). Neurons produced during embryogenesis express high levels of Chinmo (**B**) Type I NBs express Cas after reactivation from quiescence and then switch to Svp. Multiple divisions later, NBs sequentially express Broad (Br) and E93, but the precise inbetween TFs are not known (Doe, 2017a; Syed et al., 2017). Chinmo identifies early-born neurons located deep in the central brain region. Broad and E93 expressing neurons are found on more superficial layers close to the NB. (**C**) The temporal TF cascade has not been identified in Type II NBs. However, INPs produced at different stages were shown to sequentially express Diachaete (D), Grainyhead (Grh) and Eyeless (Ey) as they age. (**D**) Optic lobe (OL) medulla NBs express Homothorax (Hth), Ey, Sloppy paired 1/2 (Slp), D, and Tailless (Tll). The cascade in the OL NBs was shown to be regulated by a feed-forward mechanism.

Post-transcriptional regulation of neurogenesis

The majority of the efforts to understand *Drosophila* brain development has targeted key TFs and their transcriptional networks. Transcription fundamentally specifies the intrinsic state of the cell as it represents the orchestration of TFs and accessible chromatin landscapes. However, transcriptional regulation alone can have limitations: different subcellular compartments cannot be regulated differently and it is slow to respond to rapidly changing external or internal stimuli. Post-transcriptional regulation encompasses many processes of RNA metabolism, including splicing, nuclear export, RNA degradation, transport, and translation control (Figure 1.6). These extra layers of regulation can offer added complexity, buffering capacity, speed, and precision of gene expression depending on the requirements of neurogenesis (Franks et al., 2017; Pilaz & Silver, 2015). RBPs are the primary mediators of post-transcriptional regulation, analogous to TFs on

transcriptional control, and they exert their function by binding to their target RNA. RBPs can have a highly specific function such as degradation (e.g. XRN1), or act broadly on different levels of post-transcriptional control. For example, ELAV can influence multiple processes including polyadenylation, RNA localisation and translational control (Carrasco et al., 2020; Pascale et al., 2008), which can induce wider functional consequences.

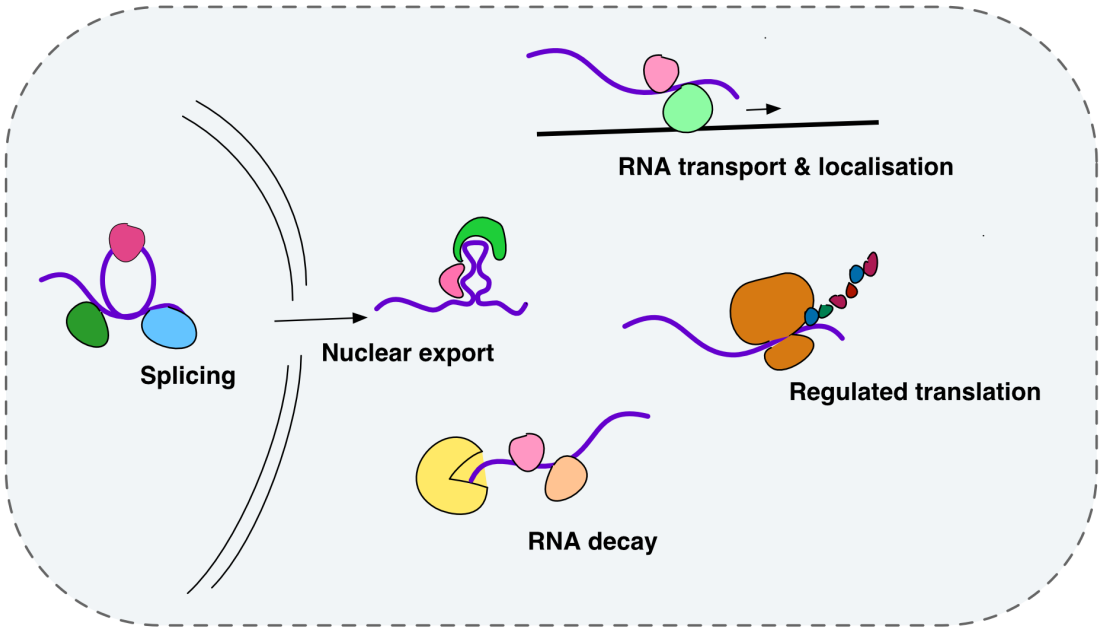


Figure 1.6: Post-transcriptional regulation of RNA metabolism. Post-transcriptional regulation influences RNA splicing, nuclear export, RNA decay, transport, localisation and translation. Each step of post-transcriptional control is mediated by RBPs. Figure adapted from Pilaz & Silver (2015) and Samuels, Järvelin, et al. (2020).

As discussed previously, neuronal fate determination is predominantly regulated by the cascade of TFs in the embryonic NBs, which imprints the information of NB age to the progeny cells (Doe, 2017a). However, in the larval central brain, the similar cascade of TFs did not continue with sufficient diversity to control more complex and longer sequences of neuronal fates. Therefore, several studies have aimed to elucidate the missing factors that regulate NB age during this stage. Initially, a mutant screening was performed to identify genes that gave rise to incorrect progression of cell fates in the mushroom body (MB) neuron lineage (Zhu et al., 2006). The MB is equivalent to the mammalian hippocampus and is developed from a subset of Type I NBs, called mushroom body neuroblasts (MBNBs). These stem cells produce a

relatively simple sequence of MB neurons, also known as Kenyon cells, during post-embryonic neurogenesis. Kenyon cells are born in the sequential order of γ, α'/β' and α/β neurons, making them a useful model to follow fate specifications. This screen identified a BTB-zinc finger TF, Chinmo, which was expressed in a decreasing gradient throughout the MBNB lifecycle (Zhu et al., 2006) (Figure 1.7). Modulating the levels of Chinmo changed the numbers of the three classes of Kenyon cells. Therefore, this observation formed a hypothesis that the regulator(s) of Chinmo would also vary over developmental time.

Indeed, Z. Liu et al. (2015) isolated MBNBs at various post-embryonic stages for transcriptomic analysis. This study found that two evolutionary conserved RBPs, IGF-II mRNA binding protein (Imp) and Synaptotagmin binding cytoplasmic RNA interacting protein (Syncrip, Syp) displayed the most dynamic changes in their abundance over time (Figure 1.8A-B). Interestingly, the two RBPs were expressed in opposing temporal gradients and manipulating their levels resulted in temporal fate transformations of Kenyon cells. It was further demonstrated that the distinct gradient levels of Imp and Syp in MBNBs post-transcriptionally regulate *chinmo*. Imp activates translation of *chinmo* mRNA, while Syp has an opposite role to specify the correct sequence of Kenyon cell fates. Importantly, the opposing gradient was observed in all Type I and Type II NBs with a moderate degree of asynchrony (Z. Liu et al., 2015; Ren et al., 2017), indicating the broad and intrinsic nature of the gradient progression.

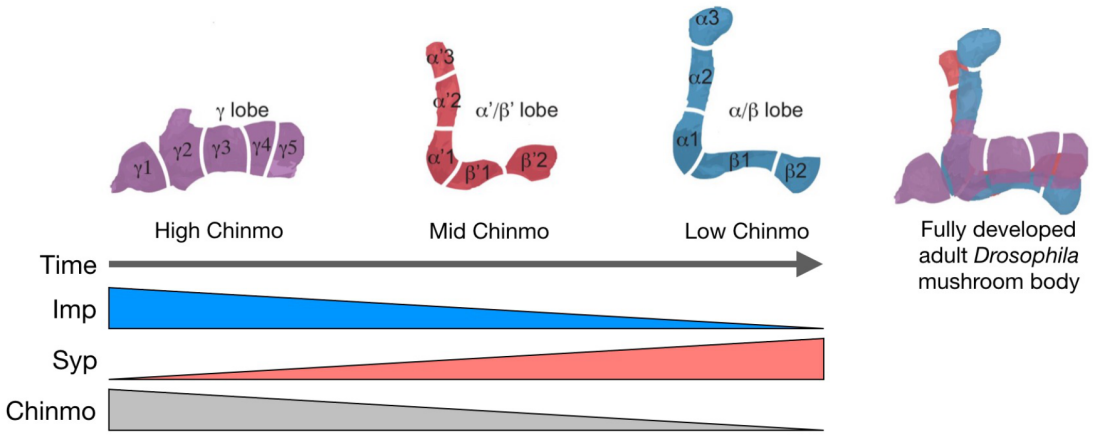


Figure 1.7: Post-transcriptional regulation of mushroom body neuron fate specification. The three mushroom body neuron fates are sequentially generated by the decreasing gradient of Chinmo in mushroom body NBs (MBNBs). High Chinmo specifies γ cell fate, while mid and low Chinmo gives rise to α'/β' and α/β, respectively. *chinmo* is post-transcriptionally regulated by Syp and Imp, where Syp suppress its translation while Imp promotes *chinmo* translation, ensuring the gradient expression of Chinmo.

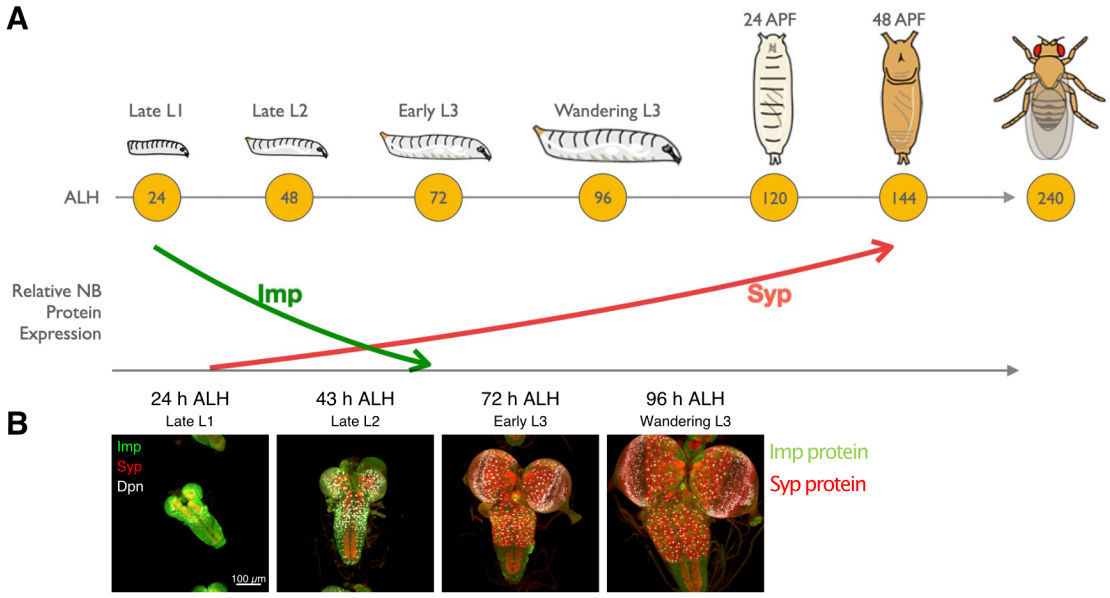


Figure 1.8: Opposing gradients of Syp and Imp during post-embryonic neurogenesis. (**A**) Schematic representation of relative Syp and Imp level in larval NBs during post-embryonic development. The time axis represents hours after larval hatching (ALH). Relative protein levels are depicted. (**B**) Immunofluorescence against Syp and Imp across four stages of the larval development, showing a brain-wide Imp-to-Syp switch. Anti-Deadpan (Dpn) was used to label NBs. Note that the neuropil staining of anti-Syp is non-specific. (n=5, representative 20 µm max-projected confocal images are shown).

Since the initial discovery, several examples of post-transcriptional regulation of neuronal fates

via Imp and Syp have been described in gene-by-gene cases (Pahl et al., 2019; Syed et al., 2017; C. P. Yang et al., 2017). However, RBPs generally regulate many different targets via different mechanisms. Therefore, identifying the targets of Imp and Syp is important to understand their roles in post-embryonic neurogenesis. Previously, the targets of Imp have been identified in the L3 brain using RNA-immunoprecipitation sequencing (RIP-seq) (Samuels, Järvelin, et al., 2020). This study was able to highlight Imp as a regulator of *myc* RNA stability, which subsequently affected NB size and division rates. However, similar genome-wide Syp targets in the brain are not known.

Syncrip

The *Drosophila* gene *syp* (previously CG17838) was first identified in the oocyte, where it was shown to regulate *gurken* RNA localisation and translation (McDermott et al., 2012). Syp contains three RNA-recognition motifs (RRMs) and an N-terminal non-canonical micro RNA (miRNA) binding domain (Hobor et al., 2018). The human SYNCRIP/HNRNPQ has another RNA-binding RGG domain in the C-terminus. Human SYNCRIP has been described to compete with polyA-binding protein (PABP) to inhibit translation (Svitkin et al., 2013) and as a component of neuronal RNA transport granules (Bannai et al., 2004). A meta-analysis revealed that loss-of-function mutations of SYNCRIP are associated with human intellectual disability (Lelieveld et al., 2016). In *D. melanogaster*, Syp was also shown to regulate synaptic growth at the neuromuscular junction (NMJ), where *syp* mutants exhibit underdevelopment (McDermott et al., 2014). Furthermore, Syp controls retrograde bone morphogenic protein (BMP) signalling, which plays a role in pre-synaptic neurotransmitter release (Halstead et al., 2014).

Both Syp and Imp have been studied individually across different model organisms and in different biological contexts. However, their roles in NB fate patterning are tightly linked. Relative levels of Syp and Imp can define NB age, where Imp is highly abundant in younger NBs and declines through the larval stage, while Syp increases to become dominant in older NBs (Z. Liu et al., 2015; Ren et al., 2017; Syed et al., 2017). Syp and Imp also mutually repress each other, so knockdown of *syp* results in prolonged Imp expression and depletion of *imp* leads to faster accumulation of Syp (Z. Liu et al., 2015; Samuels, Järvelin, et al., 2020).

Mutual repression is thought to involve post-transcriptional control of each other's transcript, but the precise mechanism of repression is unclear.

How is the Imp-to-Syp gradient maintained? The larval Cas > Svp TF switch during the high Imp stage was shown to initiate transcription of *syp* (Ren et al., 2017). Svp activates the expression of ecdysone receptor (EcR), allowing NBs to respond to external hormonal signalling and trigger Syp expression (Syed et al., 2017). Consistently, *svp* mutants completely fail to undergo normal Imp-to-Syp switch. Interestingly, however, brains with dysfunctional ecdysone metabolism eventually manage to switch to Syp albeit at a much delayed stage (Syed et al., 2017), adding further complexity to the regulation.

Syp has been shown to regulate both progeny fate and neural stem cell behaviour (Hailstone et al., 2020; Pahl et al., 2019; Ren et al., 2017; Samuels, Järvelin, et al., 2020; Syed et al., 2017), however, it is not clear precisely how it regulates these processes. RBPs, such as Syp and Imp, regulate the levels of their target through RNA stability, localisation or translation. Therefore, identifying the target transcripts is a key step to understanding its mechanism of action.

1.1.3 Methods for studying RBP targets and their regulations

How can we identify targets of Syp? Different techniques have been developed to define protein-RNA interactions on the transcriptome-wide level. Each technique has its own advantages and disadvantages, in terms of protocol difficulty, specificity, stringency, enzymatic biases, *in vitro* re-associations, or resolution (F. C. Y. Lee & Ule, 2018). Most commonly used methods are based on immunoprecipitation using antibodies against the RBP-of-interest. The interacting RNAs are pulled-down either (i) without crosslinking (RIP) in native conditions (J. Zhao et al., 2010) (Figure 1.9A), (ii) or with 'zero-distance' crosslinking via ultraviolet (UV) irradiation (CLIP) (Huppertz et al., 2014; Ule et al., 2003) (Figure 1.9B). While RIP-seq offers the simplest protocol compatible with small starting materials, it is susceptible of co-purification of contaminating RBPs and *in vitro* re-associations of physiologically irrelevant protein-RNA interactions. CLIP bypasses these problems by crosslinking RBP-RNA to retain endogenous interactions, and by using very stringent wash conditions aided by the covalent linkage. Individual nucleotide resolution

CLIP (iCLIP) goes one step further to provide single-nucleotide resolution of RBP binding maps (Huppertz et al., 2014). The high level of resolution is possible by exploiting the reverse-transcription truncation events at the RBP-RNA crosslinkage site. However, CLIP techniques rely on UV-crosslinkability of the biosample and require a relatively large amount of starting materials. CLIP protocols are also generally more difficult, leading to higher failure rates (Van Nostrand et al., 2016; Wheeler et al., 2018).

When working with tissues, RIP and CLIP do not generally yield cell-type specific information unless isolated cells are used as input. Alternatively, there have been recent developments of RBP-centric methods based on enzymatic labelling of transcripts in proximity. This is usually combined with cell-type specific drivers to express the RBP-enzyme fusion construct and avoid the need for cell dissociations. For example, the ADAR enzyme, capable of A-to-I editing, can edit transcripts close by the RBP-ADAR fusion protein. The editing events, therefore, capture the spatial proximity of the transcripts to the tagged RBP (Nguyen et al., 2020) (Figure 1.9C). Alternatively, fusing RBP to APEX (ascorbate peroxidase) enzyme can induce direct biotinylation of nucleobases in proximity (Padrón et al., 2019), which are then biochemically enriched for sequencing (Figure 1.9D). However, these techniques are often not as sensitive as CLIP and are prone to enzyme-specific biases (F. C. Y. Lee & Ule, 2018). Furthermore, it can be often challenging to express exogenous RBP-enzyme construct to the biologically relevant levels.

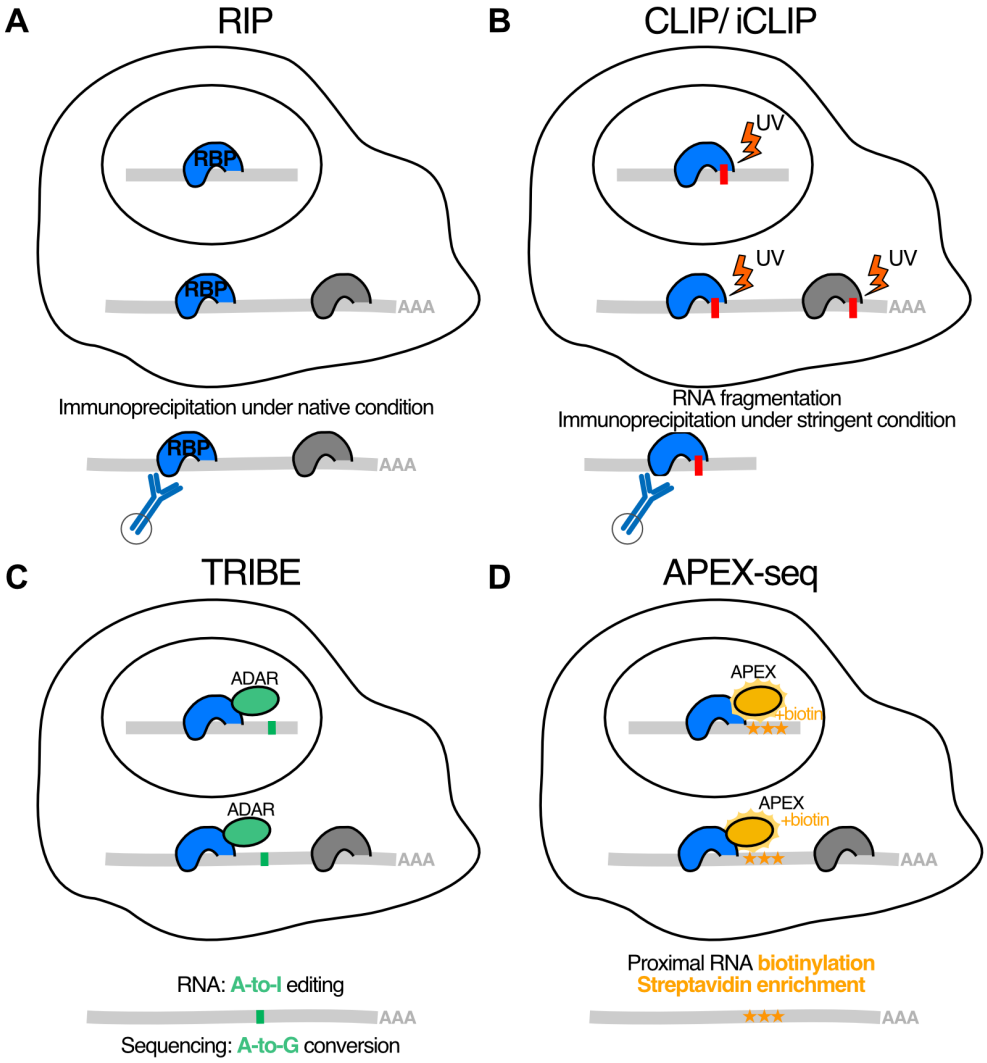


Figure 1.9: Summary of methods used to define RBP-RNA interactions at the transcriptome level. (**A**) RNA-immunoprecipitation (RIP) identifies transcripts co-immunoprecipitated with the RBP. (**B**) CLIP and iCLIP use UV to covalently crosslink *in vivo* RBP-RNA interactions. The immunoprecipitation is performed under stringent conditions and the RNAs are fragmented to avoid contaminants. (**C**) The Targets of RNA-binding proteins Identified By Editing (TRIBE) technique utilises an RBP fused to ADAR (*green*), which performs A-to-I edits on the transcript region proximal to RBP binding. (**D**) APEX-seq, utilises the APEX2 enzyme fused to an RBP (*orange*). APEX2 can biotinylate proteins and RNA within a 20 nm radius upon the addition of hydrogen peroxide and biotin-phenol.

How do we study the mechanism of post-transcriptional control mediated by RBPs? Although gene expression covers transcription, RNA metabolism and protein translation, they are rarely studied end-to-end. In order to examine the relative contribution of transcriptional and post-transcriptional control within a complex tissue, imaging-based methods are often best-suited. Such methods can retrieve spatial information on mRNA and protein distributions and quantify

changes in their abundances (Rouhanifard et al., 2018). Here, the quantitative power of single-molecule fluorescence *in situ* hybridisation (smFISH) can provide estimates of mRNA synthesis and decay rates (Bahar Halpern & Itzkovitz, 2016; Mueller et al., 2013; Raj & van Oudenaarden, 2009), while providing absolute sensitivity and high spatial precision. smFISH combined with immunofluorescence against the target protein-of-interest can reveal differences in the ratio of mRNA to protein levels and provide a gauge of translation rates (Titlow et al., 2022). In fact, such methodology is widely applied to investigate RNA localisation in the neuronal periphery as mRNAs are often present in low densities in these compartments (Holt et al., 2019; Holt & Schuman, 2013; von Kügelgen & Chekulaeva, 2020).

Alternatively, transcriptome-wide studies of post-transcriptional control have benefited from the developments of biochemical methods surveying RNA stability or ribosome occupancy. For example, 4sU-mediated metabolic labelling allows biochemical enrichment of newly synthesisied RNA and subsequent inference of RNA decay rates and stability changes in RBP mutants (de Pretis et al., 2015; Herzog et al., 2017; Jürges et al., 2018). Transcriptome-wide quantification of translation rates has classically been assessed with polysome or ribosome profiling (Calviello & Ohler, 2017; Glock et al., 2021). However, these biochemical techniques share the same disadvantage as RIP or CLIP techniques, because they often are not compatible with cell-type specific analysis. For ribosome occupancy, however, the translating ribosome affinity purification (TRAP-seq) method has been successfully applied in mouse brains by purifying translating mRNAs specifically in cells expressing exogenous GFP-tagged ribosomes (Reynoso et al., 2015; von Kügelgen & Chekulaeva, 2020).

1.1.4 Objectives

The prevailing understanding of neurogenesis, generating diversity and the large number of neurons and glia, is predominantly at the level of transcriptional regulation. Regulation of RNA stability or translation can significantly influence the protein complement of NBs and offer an additional layer of complexity that might be key to generating diversity. The opposing gradients of the RBPs, Syp and Imp, define the NB temporal progression and influence progeny fate. Thus,

perhaps there is a far greater role of post-transcriptional regulation than previously anticipated. In Chapter 3, I aim to identify the RNA targets of Syp and characterise the features of the target transcripts. Further, I highlight two key examples of Syp-mediated post-transcriptional control, other than *chinmo*. In Chapter 4, I aim to demonstrate a generalisable workflow to examine the prevalence of post-transcriptional control in a complex tissue with high sensitivity and resolution. Collectively, these Chapters aim to provide insight into the breadth of post-transcriptional regulation in the nervous system.

1.2 RNA metabolism during SARS-CoV-2 infection

The novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) emerged during my studies, leading to unprecedented public health, economic and social impacts. During the closure of the Biochemistry Department, I was given the opportunity to apply my background in working with RNA detection to investigate the poorly-understood early replication events of SARS-CoV-2 RNA genomes.

1.2.1 Lessons from Virology

The RNA biology and Virology fields are tightly linked. Investigations on viruses have led to key discoveries in RNA biology. A few examples include the identification of retroviral reverse transcriptase; the discovery of the polyadenylation signal from simian vacuolating virus 40 (SV40); discoveries of RNA interference (RNAi) using plant viruses; and the discovery of the existence of mRNA from bacteriophages (Enquist, 2009; Fitzgerald & Shenk, 1981; Taylor, 2014). Likewise, method developments in RNA biology have benefited Virology (Iselin et al., 2022; F. C. Y. Lee & Ule, 2018). Aside from the obvious translational aspects of virus research, viruses are a great basic model system as they can encode genetic information, replicate their genome, disseminate, and give rise to infection phenotypes. Therefore, studying viruses can provide new insights that improve our understanding of RNA-protein interactions and post-transcriptional controls.

1.2.2 Structure of SARS-CoV-2 genome

Coronaviruses are a family of positive-strand RNA viruses in the order Nidovirales. From the four subfamilies of *Coronaviridae* (*Alpha-, Beta-, Gamma-, and Delta*), SARS-CoV-2 is from the *Betacoronavirus* subfamily that includes highly pathogenic MERS-CoV, SARS-CoV-1 and low-pathogenic HCoV-HKU1 (Corman et al., 2018; Ke et al., 2020). Coronaviruses have high zoonotic potential, and both SARS-CoV-1 and SARS-CoV-2 are derived from zoonotic viruses that adapted to human hosts (Corman et al., 2018; Hoffmann et al., 2020; V'Kovski et al., 2021).

Coronavirus virions are composed of ~100 Spike (S) protein homotrimers used for target cell attachment, which are embedded in the outer membrane with Membrane (M) and Envelope (E) proteins. The viral genomic RNA (gRNA) is encapsulated inside the virion together with Nucleocapsids (N) as ribonucleoprotein (RNP) (Iserman, Roden, Boerneke, Sealfon, McLaughlin, Jungreis, Fritch, et al., 2020; Ke et al., 2020) (Figure 1.10A-B). SARS-CoV-2 has an unusually large genome of ~29 kilobases (kb) (Kim et al., 2020). More than two-thirds of the genome encodes two large open-reading frames (ORFs) that subsequently give rise to 16 viral non-structural proteins (nsps) (Figure 1.10C). The other 3'end portion of the genome encodes structural proteins S, M, E, and N, and a variety of accessory proteins. The structural and accessory proteins are translated from viral sub-genomic RNAs (sgRNAs) that are transcribed from the gRNA via discontinuous transcription (Kim et al., 2020).

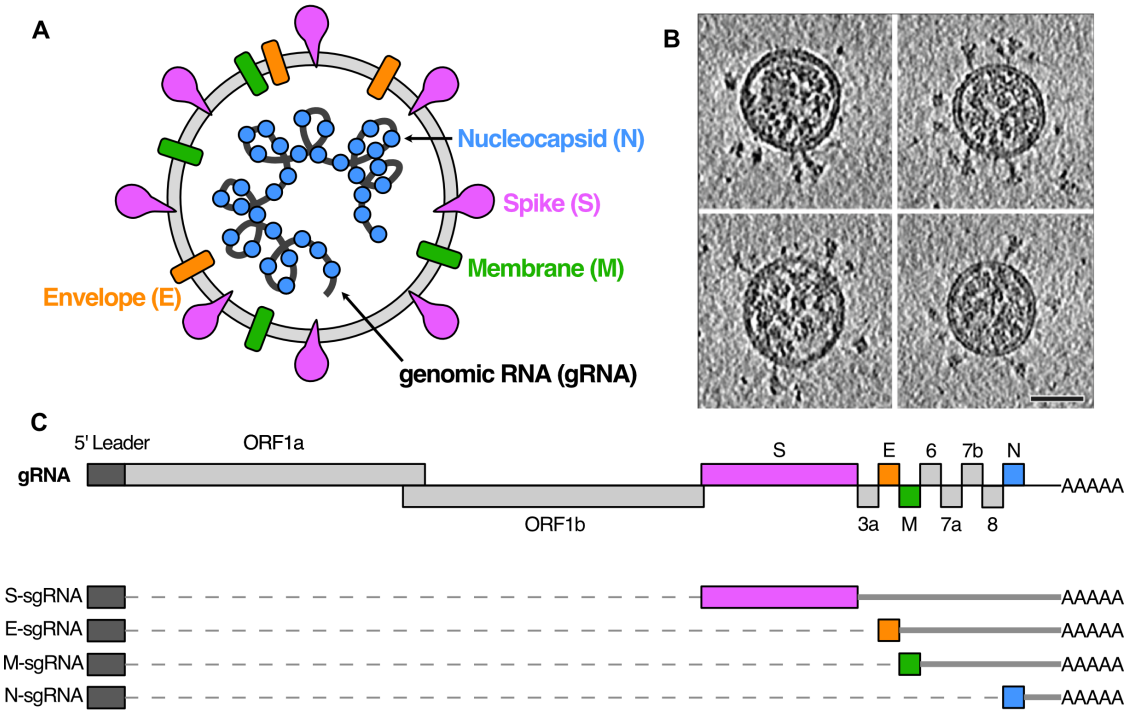


Figure 1.10: Structure of SARS-CoV-2 virus and transcriptome. (**A**) Schematic representation of the SARS-CoV-2 virion structure. (**B**) Electron micrographs of SARS-CoV-2, showing the characteristic Spike protein around the virion periphery. Figure adapted from Ke et al. (2020) (Scale bar = 10 nm). (**C**) Schematic diagram of the full-length polyadenylated SARS-CoV-2 RNA genome (gRNA). From the gRNA nine major sgRNAs are produced, which overlap with the gRNA in the 3' portion. sgRNAs are produced by discontinuous transcription, sharing the same 5'Leader sequence (*Black box*). For simplicity, only the sgRNAs encoding structural proteins are shown.

1.2.3 Infection lifecycle of SARS-CoV-2

SARS-CoV-2 mainly targets the cell lining of the respiratory tract where virus entry is orchestrated by the interaction between viral glycoprotein S and host angiotensin-converting enzyme (ACE2) (Figure 1.11). Subsequently, the transmembrane protease serine 2 (TMPRSS2) triggers the fusion of the viral and cell membranes (Hoffmann et al., 2020; M. Zhao et al., 2021). Following membrane fusion, viral gRNA RNPs are trafficked to the endoplasmic reticulum, where ORF1a and ORF1b are translated to produce polypeptides pp1a and pp1b. Polypeptides are further cleaved to produce multiple nsps that have three major functions (Finkel et al., 2021). First, most nsps are components of replication-transcription complexes (RTCs) that synthesise viral RNA (Snijder et al., 2016). The core enzyme, RNA-dependent RNA polymerase (RdRp), is derived from nsp12. Second, nsp3, 4 and 6 are involved in creating double-membrane vesicle (DMV)

compartments used as protected sites of RNA replication (Wolff et al., 2020). Finally, nsp1 and nsp3 have been described to suppress host innate immune responses such as by interfering with the host mRNA translation (Perlman & Netland, 2009; Schubert et al., 2020).

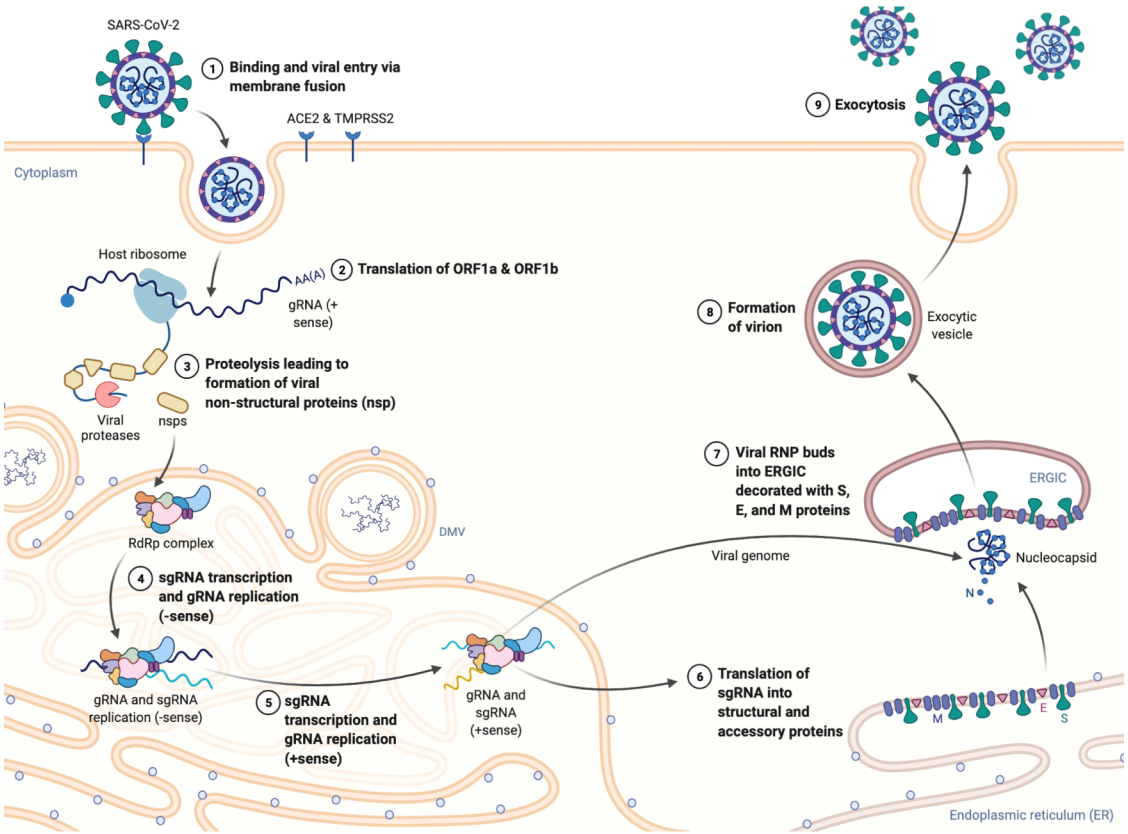


Figure 1.11: SARS-CoV-2 infection lifecycle. (1) SARS-CoV-2 binds to the host ACE2 receptor via glycoprotein Spike (S), and membrane fusion is mediated by TMPRSS2. (2) Released viral genomic RNA is translated using host ribosome machinery. (3) ORF1a and ORF1b polypeptides are subsequently autolysed to generate viral non-structural proteins (nsps). Nsps trigger remodelling of the host ER membrane and are used to create viral replication organelles such as double-membrane vesicles (DMVs). Also multiple nsps come together to assemble viral replicase complexes (RdRP complexes). (4) RdRp synthesises full-length negative-strand gRNAs and negative-strand sgRNAs. (5) Negative-strand transcripts are used as templates to produce corresponding positive-sense gRNA and sgRNAs. (6) Positive-sense sgRNAs are translated to produce viral structural and accessory proteins. (7) Virions assemble at the ERGIC, combining viral structural proteins and N-protein bound full-length gRNA. (8) Virions bud into the ERGIC and structural proteins redistribute. (9) Virions are released via exocytosis. Figure adapted from V'Kovski et al. (2021) and was created using a licensed version of Biorender.

Inside the DMVs, the assembled replicase complex uses positive-sense gRNA as a template to synthesise negative-sense (antisense) gRNAs (Cortese et al., 2020; Eymieux et al., 2021). The antisense genomes are then used to replicate new positive gRNA, thereby providing more templates that can be translated or assembled into new virions. In addition to the full-length

gRNAs, RTC synthesises a plethora of sgRNAs (Kim et al., 2020). The production of sgRNAs requires corresponding negative-sense sgRNAs as templates, which are synthesised from the discontinuous transcription of positive-sense gRNA. During transcription of negative-sense RNA, RdRP encounters transcription regulatory sequences (TRS-B), which are located upstream of each ORF of structural and accessory proteins. This interaction is capable of dissociating nascent transcribing RNA and causing it to hybridise with transcription regulatory sequence downstream of the 5'Leader sequence (TRS-L). The RdRP continues the transcription of the 5'Leader sequence to produce functional negative-strand sgRNAs, thereby producing a nested set of sgRNA products that share the same 5'Leader sequence. The negative-sense sgRNAs are transcribed to positive-strand sgRNAs with translation potential, encoding structural and accessory proteins (Kim et al., 2020; Sola et al., 2015).

The precise function of SARS-CoV-2 accessory proteins is not fully understood. Expression of these accessory proteins varies widely between *Coronaviridae* species, and also within the variants of the same species. For example, mutations introduced in the SARS-CoV-2 B.1.1.7 strain (Alpha) created novel TRS-B sites within the ORF-N sequence, leading to the production of novel ORF9b and N-star (N*) accessory proteins. Several studies described the innate immune suppressing function of ORF9b and N*, and the advantages of such mutations over the ancestral strain (Jiang et al., 2020; Parker et al., 2021; Thorne et al., 2021). Similarly, accessory proteins are generally thought to be involved in immune evasive roles, as shown for SARS-CoV-1 ORF3b and ORF6 (Kopecky-Bromberg et al., 2007).

Structural proteins are more conserved between *Coronaviridae* species, with a unified role in virion assembly. SARS-CoV-2 assembly has been described, using electron microscopy (EM), to occur inside the membranous structure derived from Endoplasmic Reticulum-Golgi intermediate compartments (ERGICs) (Cortese et al., 2020; Eymieux et al., 2021; Laue et al., 2021; Mendonca et al., 2021). These membranous structures consist of newly translated S, M and E proteins. The N protein undergoes liquid-liquid phase separation with full-length gRNA to form a compact RNP structure (Iserman, Roden, Boerneke, Sealfon, McLaughlin, Jungreis, Park, et al., 2020), which

is subsequently recruited by M protein to the ERGIC membranes. Virions bud into the ERGIC lumen and are released out of the cell by exocytosis (V'Kovski et al., 2021). These early events present a critical window of opportunity for the virus to infect or the host cell to successfully defend itself from the incoming pathogen (Boersma et al., 2020; Singer et al., 2021). However, our quantitative and kinetic knowledge of this early SARS-CoV-2 replicative lifecycle is poorly understood despite its essential role in the establishment of a productive infection.

1.2.4 Replication compartments of SARS-CoV-2

As with other positive-strand RNA viruses, the replication complexes of SARS-CoV-2 are generally associated with modified host membranes (V'Kovski et al., 2021). These membranous structures were well characterised for SARS-CoV-1. Initially, DMVs and convoluted membranes (CMs) were the first structures to be identified during the virus infection (Knoops et al., 2008; Miller & Krijnse-Locker, 2008; Snijder et al., 2016) (Figure 1.12A-C). These structures are highly dynamic and their abundance has been shown to increase throughout the course of the virus infection. Vesicle packets (VPs) were also observed, especially in later infection timepoints, which feature multiple inner vesicles encapsulated by a single outer membrane layer. Interestingly, the outer membranes of all DMVs, CMs and VPs are connected with each other and also to the Endoplasmic Reticulum (ER). This forms an elaborate membrane network around the periphery of the infected host nucleus (Miller & Krijnse-Locker, 2008). It is yet to be discovered how these structures are generated, but the host cytoskeletal vimentin network (Cortese et al., 2020) and viral nsp3, 4 and 6 have been shown to be involved (Angelini et al., 2013). Consistent with the localisation of virus replication, double-stranded RNA (dsRNA) structures were detected inside DMVs (Klein et al., 2020; Knoops et al., 2008), which represents viral replication intermediates. However, nsp8, the component of the replicase complex, was also identified in CMs (Knoops et al., 2008), so it is not known whether coronavirus replication occurs exclusively in DMVs or more generally across the membranous structures.

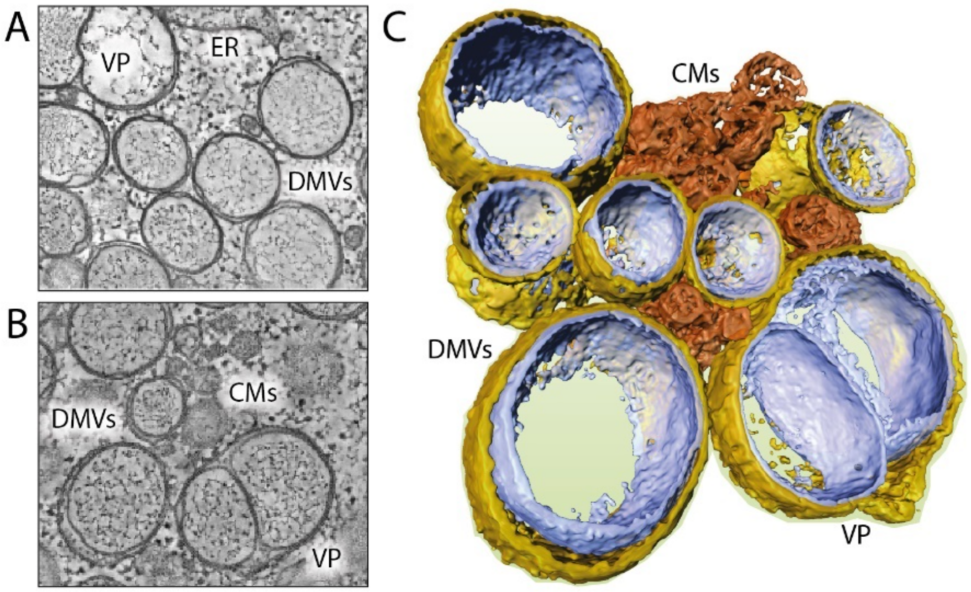


Figure 1.12: Structure of coronavirus replication compartments. (**A**)-(**B**) Electron micrographs of Convoluted Membranes (CMs), Vesicle Packets (VPs) and Double Membrane Vesicles (DMVs). (**C**) 3D reconstruction of CMs, VPs and DMVs from the electron micrographs. Figure adapted from Knoops et al. (2008).

1.2.5 Methods for studying viral replication dynamics

It is important to understand the complicated virus infection mechanisms in time and space to prevent viral disease and identify new antiviral targets. Multiple methods exist to study viral replication dynamics, which are based on the detection of virus-derived antigens, nucleic acids, or engineered reporters. Techniques such as western blots, RT-qPCR or RNA-seq have been employed extensively since the onset of the COVID-19 pandemic. While very informative, such 'bulk' ensemble methods are inadequate to examine temporally dynamic events that are asynchronous, masking cell-to-cell heterogeneity. Furthermore, these assays are often not sensitive enough to detect the virus during the first few hours of the infection, where only a single copy of gRNA is present per incoming virus particle. The advent of single-cell RNA-seq (scRNA-seq) demonstrated the ability to profile variability of viral RNA load between cells (Fiege et al., 2021; Ravindra et al., 2021), but shallow sequencing depth remains a limitation.

Alternatively, single-cell resolution viral replication dynamics have been explored using imaging approaches. Commonly used methods involve the use of immunofluorescence against

a virus-derived antigen, viral dsRNA (Weber et al., 2006), or engineered fluorescent reporters (Ramanan et al., 2016). The sensitivity of these methods is variable but is often not sufficient to profile early replication events (Boersma et al., 2020). Especially, anti-dsRNA is widely used to identify RNA virus-infected cells, but it lacks sensitivity in the early phases because the lowly abundant SARS-CoV-2 dsRNA is indistinguishable from host dsRNAs (Dhir et al., 2018; Kamel et al., 2021). Fluorescence *in situ* hybridisation methods have been applied to address the sensitivity of viral RNA detection. Combined with single-molecule sensitivity, it has offered unprecedented insights into the complex host-virus interactions in Hepatitis C, Influenza A, picornavirus and Sindbis virus infections (Boersma et al., 2020; Chou et al., 2013; Chou & Lionnet, 2018; Ramanan et al., 2016; Singer et al., 2021). However, this approach has been applied to SARS-CoV-2 in a limited capacity, and the single-molecule resolution sensitivity has largely been underutilised (Acheampong et al., 2021; Burke et al., 2021; Carossino et al., 2020; Guerini-Rocco et al., 2020; Jiao et al., 2020; Kusmartseva et al., 2020; Lean et al., 2020; J. Liu et al., 2020; Rensen et al., 2022).

The main disadvantage of the *in situ* hybridisation methods is that they require cell fixation, making it difficult to correlate live molecular events to the eventual infection outcome. With the advent of genetic tools such as 'MS2-loops' and 'TRICK-or-TREAT' (Russo & Wilusz, 2017; Tutucci et al., 2018), single RNA tracking (Singer et al., 2021), as well as visualisation of single transcript translation event (Boersma et al., 2020) became possible. Furthermore, single virus tracking (SVT) methods via fluorescently tagged proteins are increasingly becoming more popular for live and fixed cell imaging using super-resolution microscopy techniques (detailed review in (S.-L. Liu et al., 2020)). However, these techniques require genetic manipulation of the virus and have been applied to SARS-CoV-2 in limited cases (Guo et al., 2021; Senussi et al., 2021).

1.2.6 Objectives

Despite the global research effort on SARS-CoV-2, its early replication events are not completely understood. Furthermore, the recent emergence of SARS-CoV-2 Variants of Concern (VOCs), which display differential transmission, pathogenesis and infectivity, has changed the landscapes of the COVID-19 pandemic. Therefore, I used my background working with sensitive detection of

RNA molecule, to characterise SARS-CoV-2 replication kinetics using imaging and sequencing approaches. In particular, I demonstrate the utility of smFISH to perform a comparative kinetics assay comparing two different SARS-CoV-2 strains, which could be important for understanding the fitness advantage of the new VOCs (Chapter 5).

2
Materials and Methods

2.1 Fly stocks and cell culture

2.1.1 Fly stocks

Drosophila melanogaster flies were raised on standard cornmeal-agar medium and maintained in 18°C or 25°C incubators with a 12-hour day-night cycle. Fly lines used in this study are listed in Table 2.1. Fly lines used for the survey of mRNA and protein distribution of Syp targets are given in Appendix D.

2.1.2 Cell culture

Vero E6, A549-ACE2 (Klein et al., 2020) and Huh-7.5 cells were maintained in standard DMEM, Calu-3 cells in Advanced DMEM both supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 100 U/ml penicillin, and 10 μg/ml streptomycin and non-essential amino acids. All cell lines were tested free of mycoplasma and tested for their authenticity via short tandem repeat profiling at the UK Health Security Agency. Cell lines were maintained at 37°C and 5% CO_2 in a standard culture incubator.

Table 2.1: Fly stocks

Alias	Genotype	Description	Source
OrR	OregonR	wild-type	BDSC #5
sypN	PBac(PB) syp [e00286] / TM6B	syp null mutant	Harvard Exelixis Collection
Syp::GFP	w[1118]; ; syp::meGFP	eGFP-tagged Syp at the N-terminus	Titlow et al. 2018
Imp::GFP	Mi{PT-GFSTF.2} Imp [MI05901 - GFSTF.2]	FLAG-eGFP-tagged Imp	BDSC #60237
syp-RNAi	GD.33011	syp-RNAi under UAS control	VDRC #33011
e93-RNAi	TRiP.HMC04773	e93-RNAi under UAS control	BDSC #57868
prosREDr	pros{long-REDr}	pros long 3'UTR isoform mutant	Samuels et al. 2020
ok107-GAL4	P{GawB}OK107	GAL4 expression under eyeless control (mushroom body)	BDSC #854
insc-GAL4	P{GawB}inscMz1407	GAL4 expression under inscuteable control (Neuroblast and immediate progeny)	BDSC #8751
syp-f03775	PBac{w[+mC]=WH} Syp [f03775] / TM6B	syp hypomorphic mutant	BDSC #85668
syp deletion	Df(3R)BSC124 / TM6B	syp large deletion mutant	BDSC #9289
hsFLP	P{ry[+t7.2] = hsFLP}1; Adv[1]/CyO	expresses flippase under heat-shock control	BDSC #6
RFP clone	GAL4-Act5C(FRT.CD2), UAS-RFP.W / TM3	flip-out clone generator, labels clones with RFP upon flippase activity	BDSC #30558

2.1.3 Virus propagation and infection

Virus propagation was performed by Dr Peter Wing (University of Oxford), Dr Marko Noerenberg (University of Glasgow) and Louisa Iseline (University of Oxford). SARS-CoV-2 strains: VIC 01/20 (BVIC01) (Caly et al., 2020) (provided by PHE Porton Down after supply from the Doherty Centre Melbourne, Australia) and B.1.1.7 (Tegally et al., 2020) (20I/501Y.V1.HMPP1) (provided by PHE Porton Down). Vero E6 cells were infected with SARS-CoV-2 at a multiplicity of infection

(MOI) of 0.003 and incubated for 48–72 h. When cytopathies were observed, culture supernatant was harvested, clarified by centrifugation to remove residual cell debris and stored at –80°C. To determine the viral titre, Vero E6 cells were inoculated with serial dilutions of SARS-CoV-2 viral stocks for 2 h followed by addition of an overlay consisting of 1.5% carboxymethyl cellulose. Cells were incubated for 72 h and visible plaques were counted by fixing cells using amido black stain to calculate PFU/ml. HCoV-229E (Andrew Davidson lab, University of Bristol) virus was propagated in Vero E6 cells and plaque assay was performed in Huh-7.5 cells to calculate viral concentration.

For smFISH experiments with the SARS-CoV-2 strains, cells were infected at an MOI of 1 for 2 h followed by extensive washes with PBS. Non-internalised virus particles on the cell surface were removed by trypsin treatment for 2 min followed by neutralisation of trypsin using media containing FBS. Infected cells were maintained for a defined duration of time up to 48 h. For the infection with HCoV-229E, cells were infected at an MOI of 1 and were maintained for a further 24 h without washing.

2.2 Microscopy and Imaging

2.2.1 Single-molecule fluorescence *in situ* hybridisation (smFISH)

All smFISH probe sequences used in this study are listed in Appendix A.

smFISH in *D. melanogaster* tissues

smFISH was carried out as previously described (Titlow et al., 2018; L. Yang et al., 2017) with minor modifications. Tissues were dissected in PBS or Schneider's medium, rinsed once with PBS and fixed with 4% paraformaldehyde (PFA) diluted in PBS/0.1% Triton-X 100 (PBSTx) for 25 min at room temperature (RT). Samples were rinsed briefly in PBSTx and then further permeabilised with PBSTx twice for 20 min each. Samples were pre-hybridised in smFISH wash solution (10% formamide, 2xSSC) for 30 min at 37°C and hybridised in smFISH hybridisation solution (10% formamide, 2xSSC, 10% dextran sulphate) with probes diluted to 250 nM final concentrations overnight at 37°C. Samples were rinsed briefly in smFISH wash solution and then washed in smFISH wash solution twice for 30 mins each at 37°C. Finally, samples were

washed in 2xSSC for 10 min at RT and mounted on slides using either Vectashield (Vector labs) or Slowfade Diamond (ThermoFisher) anti-fade mounting media. Slides were either imaged immediately or stored at 4°C. Counterstains were included in the wash solution: DAPI (1 μg/ml), fluorophore-conjugated phalloidin (1:100, ThermoFisher).

smFISH in tissue culture cells

Cells were grown on #1.5 round-glass coverslips in 24-well plates or in 8-well glass bottom chambers (IBIDI) and fixed in 4% paraformaldehyde (Thermo Fisher) for 30-60 min at room temperature. Coverslips were cleaned in 80% ethanol with lint-free tissue and kept in 100% ethanol to maintain sterility and cleanliness. Cells were permeabilised in PBSTx for 10 min at room temperature followed by washes in PBS and 2×SSC. Then, the cells were pre-hybridised and hybridised with probes as with *D. melanogaster* tissues. For probes targeting SARS-CoV-2, concentrations of 500 nM (under 10 hours post infection, hpi) or 1 μM (24 hpi and above) were used. Counterstains were included in the wash solution: DAPI (1 μg/ml), fluorophore-conjugated phalloidin (1:2,000 - 1:5,000), CellMask Green (1:1,000,000).

For RNase-treated control experiments, 100 U/ml RNaseT1 (ThermoFisher) and/or 20 U/ml RNaseIII (NEB) were used to degrade ssRNA and dsRNA, respectively. Permeabilised cells were treated with RNases in PBS supplemented with 5 $MgCl_2$ and incubated at 37°C for 1 h and washed three times with PBS for 10 min each.

For the detection of viral negative strands, fixed and permeabilised cells were treated with DMSO, formamide, or NaOH (Singer et al., 2021; Wilcox et al., 2019). Cells were briefly rinsed in distilled water and were treated with 50 mM NaOH for 30s at room temperature, 70% formamide at 70°C for 1 h, or 90% DMSO at 70°C for 1 h. Subsequently, cells were quickly cooled on ice, washed in ice-cold PBS and then subjected to standard smFISH protocol.

smFISH in formalin-fixed paraffin-embedded (FFPE) tissue slices

Formalin-fixed paraffin-embedded (FFPE) sections of infected hamster lungs were provided by Dr Javier Salguero (Public Health England). The specimens were prepared as described in

(Dowall et al., 2021), fixed in 10% neutral buffered formalin, embedded in paraffin blocks, and sectioned to 5 μm slices. Tissue pre-treatment was performed as in (Rouhanifard et al., 2018) with minor modifications. Tissue sections were deparaffinised in 100% xylene (twice for 10 min), washed in 100% ethanol (twice for 5 min) and post-fixed in methanol-acetic acid (3:1, v/v) for 5 min. Tissues were re-hydrated in an ethanol gradient for 3 min each (100%, 85%, 70%, nuclease-free water). Crosslink reversing was performed by heating at 80°C for 1 h in antigen retrieval solution (10 mM sodium citrate, pH 6) supplemented with 4 mM Ribonucleoside vanadyl complexes (RVC). Tissues were further permeabilised in 70% ethanol overnight at 4°C. Subsequently, sections were incubated in 100% ethanol for 5 min, air-dried for 5 min. Tissue clearing was performed with 8% SDS made up in 2×SSC for 30 s followed by extensive washes in 2xSSC. Afterwards, standard smFISH procedures were followed.

2.2.2 smFISH probe design and specificity analysis

Candidate smFISH probe sequences were generated using Stellaris Probe Designer version 4.2 (`https://www.biosearchtech.com/stellaris-designer`) with the following parameters: 'masking level, 5; oligo length, 20 nt; minimum spacing length, 3 nt'. Oligonucleotides were singly labelled with ATTO633, ATTO565, or Cy3 at the 3' ends according to a published protocol (Gaspar et al., 2017) and their concentrations were normalised to 25 μM using nanodrop measurements. All probe sets used in this study had degree of labelling > 0.94. For designing virus probes, the appropriate region of the SARS-CoV-2 Wuhan-Hu-1 (NC_045512.2) reference sequence was used as the target sequence. Candidate probe sequences were BLAST screened against custom human exon and intron database to score the number of off-target basepair matches, then 35–48 sequences with the lowest match scores were chosen per probe set.

For assessing probe specificity to SARS-CoV-2, oligonucleotide sequences against ORF-1a and ORF-N were mapped against SARS-CoV-1 (NC_004718), SARS-CoV-2 (NC_045512), MERS-CoV (NC_019843), HCoV-229E (NC_002645), HCoV-NL63 (NC_005831), HCoV-OC43 (NC_006213), HCoV-HKU1 (NC_006577), Human (GCF_000001405.39), and African green monkey (GCF_015252025.1) RefSeq genome or transcriptome assembly using 'bowtie2' (2.4.4)

(Langmead & Salzberg, 2012). Following bowtie2 arguments were used to find minimum edit distance of oligonucleotide sequences to target genome/transcriptome: –end-to-end –no-unal –align-seed-mm 0, –align-seed-length 5, –align-seed-interval 1–1.15, –effort-extend 15, –effort-repeat 2. Melting temperatures were obtained using the 'rmelting' R package at 300mM sodium concentration, which reflects 2xSSC hybridisation conditions.

2.2.3 Immunofluorescence

After permeabilisation, samples were blocked in blocking solution (Licor Odyssey blocking buffer supplemented with 0.1% Tween-20) for 30-60 min at RT. Then, samples were incubated with primary antibodies diluted in blocking solution. *D. melanogaster* samples were probed overnight at 4°C while tissue culture cells were incubated for 2 h at RT. Cells were washed three times in PBS/0.1% Tween-20 (PBSTw) for 10 min each at room temperature and incubated with fluorescent secondary antibodies (1:500) diluted in blocking solution for 2 hr at RT. After further three washes in PBSTw, cells were mounted in an anti-fade mounting medium. Antibodies used in this study are listed in Table 2.2. For combined smFISH and immunofluorescence, antibody staining was carried out sequentially after the smFISH protocol. Additionally, the blocking solution was pre-treated with 1:25 RNASecure (Invitrogen) for 30 min at 50°C and supplemented with 2 mM RVC to prevent RNase activity.

2.2.4 Microscopy

Prepared samples were imaged on an Olympus SpinSR10 spinning disk confocal system equipped with Prime BSI and Prime 95B sCMOS cameras or on an Olympus FV3000 Inverted Laser Scanning Microscope. Objectives used were ×20 dry (0.8 NA, UPLXAPO20X), ×60 silicone oil (1.3 NA, UPLSAPO60XS2), ×60 oil (1.5 NA, UPLAPOHR60X), or ×100 oil (1.45 NA, UPLXAPO100XO). Image voxel sizes were $0.55 \times 0.55 \times 2$ μm (x:y:z) with the ×20 objective and $0.11 \times 0.11 \times 0.2$ μm (x:y:z) with the ×60 and ×100 objectives. Microscopes were driven using Olympus cellSens Dimension or FV313S software. Automatic image acquisition and image stitching were performed with cellSens using the Z-drift correction (ZDC) function. The ZDC

Table 2.2: Antibodies used for immunofluorescence

Antibody	Working concentration	Source
Syncrip (guinea pig)	1:500	McDermott et al. 2014
Syncrip (rabbit)	1:200	Rossi et al. 2020
Imp (rabbit)	1:500	Gift from P. Macdonald
Pros (mouse)	1:10	DSHB, MR1A
H3K9me3 (rabbit)	1:500	Abcam, ab176916
Deadpan (rat)	1:200	Abcam, ab195173
Broad-core (mouse)	1:100	DSHB, 25E9.D7
ELAV (rat)	1:10	DSHB, 7E8A10
Fas2 (mouse)	1:10	DSHB, 1D4
E93 (guinea pig)	1:500	Syed et al. 2017
SARS-CoV-2 Nucleocapsid (human)	1:2000	Huang et al. 2020
J2 dsRNA (mouse)	1:500	Scicons, 10010200

corrects the sample drift by infrared-laser detection of glass coverslip, which often fails when the refractive index of the glass and immersion oil matches perfectly. Therefore, to aid robust registration of coverslip positions during automated imaging, samples were mounted in 10-30% Vectashield (refractive index ~ 1.35) to force refractive index mismatch. Images were uploaded and stored in the University of Oxford OMERO server (Allan et al., 2012), and OMERO.figure was used to generate image visualisations.

2.3 Biochemical techniques

2.3.1 iCLIP

UV crosslinking and lysate preparation

For whole-larval and larval-head iCLIP, samples were collected and snap-frozen in liquid nitrogen. Frozen biosamples were ground into fine powder using liquid a nitrogen-cooled mortar and pestle and were subject to 254 nm UV-C irradiation four times at 150-250 mJ/cm2 using a Stratalinker 2400 (Stratagene). For larval brain iCLIP, wandering L3 brains (older than 96 h after larval hatching, ALH) were dissected in cold PBS supplemented with 0.2 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) in batches of 50 brains, and UV-C irradiated four times at 150 mJ/cm2 with shaking in-between. Crosslinked samples were lysed

in CLIP lysis buffer supplemented with 0.2-2 mM AEBSF (Sigma), 1-2x cOmplete protease inhibitor (Roche), 1 mM Ribonucleoside vanadyl complex (RVC, New England Biolabs), and 40 U/ml RNAsin plus (Promega). Lysates were sonicated using a Bioruptor (Diagenode) with 6-10 cycles of 30 s on/off periods at the low setting. Lysates were cleared by centrifugation and filtration using Co-star 0.22 μm spin filters (Sigma), and protein levels were quantified using the Pierce 660 Protein Assay (ThermoFisher). Quantified lysates were diluted to 0.5-1.5 mg/ml, snap-frozen and stored at -80°C.

iCLIP library preparation

iCLIP sequencing libraries were generated using the iCLIP2 protocol with several modifications from infrared CLIP (irCLIP) and enhanced CLIP (eCLIP) to avoid radioactivity, improve RNA recovery and enhance reliability of binding site discovery by preparing additional size-matched input libraries (Buchbender et al., 2019; Lee et al., 2018a; Van Nostrand et al., 2016; Zarnegar et al., 2016). Detailed optimisation steps are described in the relevant sections of the Chapter 3. Crosslinked RNAs in the lysates were partially trimmed with 0.15-0.5 U/μl of RNAse I (Ambion) depending on the endogenous RNAse activities. GFP-tagged Syp (L3 brains and whole-larvae) was immunoprecipitated using GFP-Trap Magnetic Agarose Beads (Chromotek) ,and Syp (L2 heads) was immunoprecipitated using guinea-pig anti-Syp (McDermott et al., 2012) coupled to Dynabeads Protein A (Invitrogen). FLAG-tagged Imp (Nagarkar-Jaiswal et al., 2015) was pulled down with Magnetic M2 FLAG Beads (Sigma). Immunoprecipitations and washes were performed as described in the iCLIP2 protocol, and the RNAs were 3'end dephosphorylated by PNK and FastAP phosphatases. Then, infrared IR800-dye conjugated 3' adapters were ligated overnight using RNA ligase 1 in high percentage PEG8000 buffer to improve ligation yield. Molecular weight distribution of the protein-RNA complexes was visualised after SDS-PAGE and transfer to nitrocellulose membranes, then appropriate size regions (~75 kDa above the expected molecular weight of the RBPs) were excised to extract RNA using proteinase K in an SDS-containing buffer. The RNAs were further purified with phenol/chloroform/isoamyl alcohol mix (pH 6.6) followed by clean-up with RNA Clean & Concentrator Kit (Zymo research). The

purified RNAs were reverse transcribed using SuperScript IV (Invitrogen), and the cDNA was purified with MyONE Silane beads (Thermofisher). Second DNA adapters containing unique molecular identifiers (UMIs) and experimental barcodes were attached to the 3' end of the cDNAs overnight using RNA ligase 1, and the reaction clean-up was performed using MyONE Silane beads. The cDNAs now flanked by Illumina adapters at both ends were amplified using a two-step PCR strategy. First, cDNAs were amplified with 6 PCR cycles using short Solexa P3/5 primers to allow beads-based size selection with ProNex Beads Size-selection Chemistry (Promega). In order to avoid over-amplification, the optimal number of cycles for the second amplification was calculated using 1 μl of the first PCR product from qPCR Ct values. Both PCR and qPCR were performed using Pfusion High-Fidelity Polymerase Master Mix (NEB) and full-length Solexa P3/5 primers, and Evagreen (Biotium) was added as a reporter dye in qPCR reactions. After the second PCR amplification, libraries were subject to a second round of size selection by 2.4X ProNex Chemistry to deplete excess primers and molecules smaller than 150 bp.

Size-matched input library preparation

Size-matched input (SMInput) libraries were prepared in parallel to iCLIP experiments. 10 μl of the RNAse I digested lysate were separated via SDS-PAGE alongside the paired immunoprecipitation sample. After transfer to a nitrocellulose membrane, the same molecular weight region was excised, hence size-matched, and RNA was extracted, which represents the background RBP crosslink-ome. Purified RNAs were 3'end dephosphorylated and adapter-ligated in solution followed by MyONE Silane clean-up after each step. Subsequently, the SMInput RNAs were processed alongside the immunoprecipitation sample at the reverse transcription stage.

2.3.2 Western blot

Tissue lysates were prepared by homogenising biosamples using a hand-held pestle grinder or by making frozen powder using a liquid nitrogen-cooled pestle and mortar. Proteins were extracted in CLIP lysis buffer on ice for a further 10 min and the lysates were cleared by centrifugation. Protein concentrations of lysates were measured using the Pierce 660 Protein Assay kit (ThermoFisher)

Table 2.3: Antibodies used for western blot

Antibody	Working concentration	Source
Syncrip (guinea pig)	1:2000	McDermott et al. 2014
GFP (rat)	1:1000	Chromotek 3H9
FLAG (mouse)	1:1000	Sigma, F1804
H3K9me3 (rabbit)	1:1000	Abcam, ab176916
Tubulin (mouse)	1:1000	Sigma, T5168

and then 10-50 μg of proteins were denatured by boiling at 95°C for 10 min in lithium dodecyl sulphate (LDS) sample buffer (Invitrogen) containing reducing agents. Proteins were separated by SDS-PAGE on NuPAGE 4-12% Bis-Tris or 3-8% Tris-Acetate gradient gels (Invitrogen) and then transferred to PVDF membranes with the Trans-Blot Turbo Transfer System (Biorad). Membranes were blocked in 50% Odyssey Blocking Buffer (Licor) diluted with PBS for 1 h at RT. Primary antibodies were probed overnight at 4°C. After washes with 0.1% PBSTw, the membrane was probed with fluorescent secondary antibodies (1:10,000, Licor) for 2 h at RT. Membranes were washed with 0.1% PBSTw and imaged using Licor Odyssey XF imaging system. Antibodies used for immunoblotting in this study are listed in Table 2.3.

2.3.3 RT-qPCR

RT-qPCR was performed using primers specific to a transcript of interest or Illumina sequencing adapters. Reaction mix was prepared using SYBR Green Master Mix (PCRBio) or Phusion HF Master Mix (NEB) and the reactions was performed with the CFX96 touch Real-Time PCR Detection System (Biorad). Cycle threshold (Ct) values were calculated from the Biorad CFX software using the second derivative maximum method where the fluorescence curve of the sample turns sharply upward. RT-qPCR of SARS-CoV-2 infected cells were performed by Dr Peter Wing in the Category 3 containment lab (Nuffield Department of Medicine). Briefly, infected cells were harvested in RLT buffer and RNA extracted using the RNAeasy kit (Qiagen). SARS-CoV-2 RNA was quantified using the One-step RT-qPCR kit (Takyon) using primer probes directed against the

SARS-CoV-2 N gene (SARS primer probe, IDT) and beta-2-microglobulin as an internal control. SARS-CoV-2 qPCR reactions were carried out using a Roche 96 Light cycler.

2.4 Image analysis

2.4.1 Cell segmentation

Cell segmentation was performed either manually in ImageJ (National Institute of Health) or automatically with the MorphoLibJ ImageJ plugin (Legland et al., 2016) or with a Cellpose python library (Stringer et al., 2021). Monolayer tissue culture cells were segmented using 2D maximum intensity projected images of CellMask stains. Cellpose parameters for ×60 and ×100 magnification images were 'model_type = cyto, diameter = 375, flow_threshold = 0.9, cellprob_threshold=-3'. For 20× stitched images, the CellMask channel was deconvolved with a constrained iterative module using cellSens (five iterations, default spinning disk PSF, Olympus), then the following Cellpose parameters were used: 'model_type = cyto, diameter = 55, flow_threshold = 0, cellprob_threshold=-6'. Brain neuroblasts were segmented in 3D using phalloidin staining using 'Morphological segmentation' tool of the MorphoLibJ library. Brain nuclei were automatically segmented with DAPI staining using Cellpose in 3D mode with 'model_type = omnipose', which performs better than 'model_type = nuclei'.

2.4.2 smFISH quantification

FISH-quant

For the analysis of smFISH datasets using FISH-quant (Mueller et al., 2013), ImageJ region of interest (ROI) files were converted to FISH-quant outline files using a custom Python script. Then, the smFISH channels were Laplacian of Gaussian filtered (sigma = 7, 3 px) and pre-detected using local maximum mode with the 'allow smaller z region for analysis' option enabled. Pre-detected diffraction limited spots were fitted with 3D Gaussians and thresholded in batch mode based on filtered intensity, amplitude, and σz. The intensity threshold was defined by identifying the plateau range of the number of filtered spots across multiple fluorescence intensity values or by the uninfected 'Mock' condition samples. The filtering also removed non-specific autofluorescence

and rare dust particles because these contaminants usually show lower fluorescence intensity and are highly variable in shape. Viral factories were defined as gRNA smFISH signals with spatially extended foci that exceed the point-spread function of the microscope and intensity of the reference single molecules. In FISH-quant, the foci were quantified using the TxSite quantification mode (xy:z = 500:1200 nm crop per factory) with normal-sampled averaged single-molecule image (xy:z = 15:12 px) from the batch mode output. Then, the 'Integrated intensity in 3D' method was used to compare the integrated intensity of the viral factory to that of averaged single-molecule RNA.

BigFISH

Large smFISH datasets were processed with a custom Python pipeline using Bigfish (Imbert et al., 2021), skimage, and numpy libraries, which is detailed in relevant sections of the Chapter 5. Raw image files were converted to a numpy array, and individual cells were segmented using the Cellpose library as described above. Images, where cells were labelled with the CellMask stain, were pre-processed with a median filter, radius = 50 px. The background signal in the smFISH channel was subtracted with the skimage.white_tophat algorithm (radius = 5 px, individual z frames were processed in 2D due to memory constraints, results were indistinguishable from 3D-processed images). Threshold setting for smFISH spot detection was set specifically for each set of images collected in each session. Cells with high viral RNA ($>10^{5\text{-}6}$ RNA counts) were quantified by integrating smFISH channel intensities within entire cellular volumes and comparing to the reference integrated intensity of single molecules derived from cells with a lower infection density. Reference single-molecule images were obtained using the 'build_reference_spot()' function iterating over the entire image dataset in Bigfish.

In Bigfish, viral factories were resolved using the 'decompose_cluster()' function to find a reference single-molecule spot in a less signal-dense region of the image, which was used to simulate fitting of reference single-molecule spots into viral factories until the local signal intensities are matched. The candidate factories were filtered based on the previously reported radii of DMVs measured by electron microscopy (150 nm pre-8 hpi and 200 nm post-8 hpi)

(Cortese et al., 2020). In addition, I applied a threshold of 3–7 RNA molecules per factory as a technical cut-off to prevent over-clustering of viral factories at later infection timepoints.

RNA half-life calculation

RNA half-lives were calculated using TransQuant (Bahar Halpern & Itzkovitz, 2016). TransQuant software was used in MATLAB to obtain the probe library weighting factor, which accounts for the positional information of each oligonucleotide on the target transcript. Transcription rate, decay rate, and half-lives were calculated using the following equations. An elongation rate of 1.5 kb per minute was used for the calculation (Samuels, Järvelin, et al., 2020). For the *e93* transcript, Chromosome Fraction was 1 and Gene Copy Number was 2.

$$Transcription\,Rate = \frac{\frac{Nascent\,RNA}{Weighting\,Factor} * Elongation\,Rate}{Transcript\,Length}$$

$$Decay\,Rate = \frac{Chromosome\,Fraction * Transcription\,Rate * Gene\,Copy\,Number}{Mature\,RNA}$$

$$Half\,Life = \frac{ln(2)}{Decay\,Rate} * 60$$

2.4.3 Fluorescence intensity quantification

Immunofluorescence images were background subtracted using rolling ball subtraction method (radius = 50-150 px) in ImageJ. Antibody stain was quantified by integrating fluorescence intensity across the z-stacks of region of interest (ROI) divided by the volume to obtain signal density. The signal density was then normalised to the average signal density of background (area with no tissue) or the average signal density of uninfected 'Mock' condition cells in the case of anti-dsRNA J2. Fluorescence intensity profiles were obtained using ImageJ 'Plot profile' tool across 3 μm, region on maximum intensity projected images.

To assess co-localisation of viral N protein with SARS-CoV-2 RNA, ellipsoid objects of size approximating the point-spread function of the microscope (xy radius = 65 nm, z radius

= 150 nm) were placed at the centroid coordinates of smFISH spots using ImageJ 3D suite. Integrated density of N protein channel (background subtracted, radius = 5 px) fluorescence within the ellipsoid mask was calculated and compared to the equivalent signal in the uninfected condition or randomly distributed ellipsoids.

2.4.4 Simulating spatial distribution of cells

SARS-CoV-2 infected cells were hybridised with +ORF1a smFISH probes and imaged using a 20x air objective. A large area covering at least 70% of the 8-well chamber (IBIDI) area was scanned with image stitching. The spatial distribution of 'super-permissive' infected cells were assessed by testing the complete spatial randomness hypothesis by comparing the average nearest-neighbour distance of 'super-permissive' cells to an equal number of randomly selected coordinates (Ripley, 1979). The coordinates of 'super-permissive' cells were obtained using the 3D object counter with manual intensity thresholding on the +ORF1a smFISH channel (ImageJ). To prevent placement of simulated coordinates to void regions where cells are not present, a mask image was created by segmenting the DAPI channel. Placement of simulated coordinates were confined to regions defined by the DAPI mask. Nearest-neighbour distances were calculated using the KDtree algorithm (Maneewongvatana & Mount, 1999) using the Python Scipy library function 'spatial.KDTree'. Non-random distributions (evenly-placed or clustered) were simulated by randomly placing the first point, then constraining the placement of subsequent coordinates to within a defined number of pixels. Nearest-neighbour index, Rn, (Pinder & Witherick, 1972) were calculated according to the following equation, where D(obs) is the average nearest-neighbour distance (μm), a is the total imaged area (μm), and n is the number of super-permissive cells. Rn value of 1 suggests a random distribution, whereas Rn < 1 indicates clusteredness and Rn > 1 indicates regular distributions.

$$Rn = \frac{D(obs)}{0.5\sqrt{\frac{a}{n}}}$$

2.4.5 Calculating RNA spatial distribution index

The spatial distribution indices of SARS-CoV-2 gRNA and sgRNA were calculated using the RNA Distribution Index (RDI) calculator (Stueland et al., 2019). smFISH channels were background subtracted using the rolling-ball method in ImageJ. Nuclei and cell boundaries were pre-segmented in imageJ using the 'Auto Threshold' function on DAPI ('method = Huang') or CellMask green ('method = MaxEntropy') channels. Thresholded images were cropped per cell, maximum intensity projected and used as input for the RDI calculator. The standard MATLAB graphical user interface was used for the analysis. The background removal tool of the RDI calculator was not used as the input image was already pre-processed in ImageJ.

2.5 Bioinformatics and statistics

2.5.1 iCLIP analysis

Raw fastq files from iCLIP were pre-processed as described in (Busch et al., 2019) with minor modifications. The data was mapped to the ENSEMBL *D. melanogaster* genome assembly version 99 using STAR (2.7.3) with default settings, with end-to-end mapping (Dobin et al., 2013). cDNAs arising from PCR duplication events filtered by removing duplicate reads containing the same UMIs using UMI_tools. Crosslink sites were extracted, which were defined as the nucleotide position preceding the cDNA insert (i.e one nucleotide before the reverse transcription truncation events). Significant crosslink sites and RBP binding peaks were called using the iCount package (<https://icount.readthedocs.io>) with a false discovery rate (FDR) threshold of 0.01. Significant binding peaks were normalised to 8 nt length using the R script provided in (Busch et al., 2019) and were filtered by counting crosslink events (normalised to library size within each sample) and comparing the counts between iCLIP, SMInput and IgG control libraries.

2.5.2 RNA-sequencing analysis

Reanalysis of published datasets

Published RNA-seq datasets were re-analysed using the Kallisto-DESeq2 method (Bray et al., 2016; Love et al., 2014). Raw fastq files were downloaded from Gene Expression Omnibus

(GEO) and were filtered for *D. melanogaster* ribosomal RNA and low-complexity reads using BBDuk. Filtered reads were pseudoaligned to the ENSEMBL *D. melanogaster* transcriptome assembly version 99 using Kallisto (Bray et al., 2016). The custom transcriptome index was created using Kallisto-Bustools to include flanking intron sequences (25 nt) at either ends of exons, which allows recovery of pre-processed mRNA reads. Estimated counts and transcripts per million (TPM) for transcript isoforms were collapsed for each gene. Differential expression analysis was performed using the R package DESeq2 (Love et al., 2014). Normalisation for library size were performed within DESeq2 analysis and multiple hypothesis testing was done via the Independent Hypothesis Weighting (IHW) method (Ignatiadis et al., 2016) implemented in R package 'IHW'. For differential expression of genes in neuroblast timecourse datasets (Z. Liu et al., 2015; Ren et al., 2017), the pseudoaligned counts were processed using maSigPro (Conesa et al., 2006). Identified timecourse groups were further clustered manually depending on the overall shape of the expression pattern.

Analysis of SARS-CoV-2 infected A549-ACE2 timecourse datasets

The computational analysis of RNA-seq from SARS-CoV-2 infected A549-ACE2 cells was done by Dr Aino Jarvelin (University of Oxford) in collaboration. Total RNA from infected cells was extracted by Dr Peter Wing (University of Oxford) as with RT-qPCR and the sequencing libraries were prepared by Natasha Palmalux (University of Glasgow) using the Illumina Total RNA prep with the Ribo-Zero Plus library kit according to the manufacturer guidelines. Raw fastq files were mapped to a combined human (ENSEMBL GRh38.99) and SARS-CoV-2 (NC_045512.2) RefSeq assembly using STAR. Differential expression analysis was performed using DESeq2 and principal component analysis (PCA) was also performed with DESeq2 after library size correction and variance stabilisation with a logarithm transformation implemented in DESeq2. To assess relative levels of viral subgenomic (sg) and genomic RNA expression, Dr Javelin tallied the alignments (using GenomicRanges and GenomicAlignments R packages) (Lawrence et al., 2013) mapping to the region unique to the genomic RNA and the shared region

and normalised for their respective lengths. Unique contributions of sgRNA regions were then estimated by subtracting the contribution of the genomic RNA from the shared region.

2.5.3 Statistics, data wrangling and visualisation

All statistical analyses were performed using an R package 'rstatix'. Data normality was assessed using the Shapiro-Wilk test. Comparisons were performed either using the parametric tests, t-test and ANOVA-test or non-parametric tests Mann-Whitney U or Dunnett's test. p-values were adjusted using the Bonferroni method for multiple comparisons. For data wrangling, the 'tidyverse suite' of packages were used in the RStudio environment, while 'Numpy' and 'Pandas' Python libraries were used in the Jupyter notebook environment. R packages 'ggplot2', 'ggbeeswarm', 'scales', and 'patchwork' were used to create the presented visualisations. Further visual annotations were made in Affinity Designer (Serif).

2.6 Other methods

2.6.1 Neuroblast clonal analysis

GAL4-positive neuroblast clones were generated using the heat-shock inducible flippase (*hsflp*) element and the GAL4 flip-out cassette under the control of a transcription stop sequence flanked by *FRT* sites. Appropriately staged larvae were heat-shocked at 37°C for 30-60 min in a water bath to express *hsflp*, which stochastically induces GAL4-positive neuroblast clones. The heat-shocked flies were maintained at 30°C until dissection.

2.6.2 Larval locomotion assay

Larval locomotive behaviours were analysed as in (Torres-Méndez et al., 2022) with minor modifications. 3rd instar (L3) larvae were acclimatised to room temperature for 2 hours and placed at the centre of a 9-cm petri dish with 1% agarose five at a time. Larvae were further acclimatised for 2 min on stage and were allowed to crawl freely. White light was illuminated from the bottom to enhance contrast and the videos were taken at 1080p, 30 frames per second. Animals were tracked using BIOImageOperation (`https://joostdefolter.info/bio-research`) using

the standard tracking mode with static background subtractions (threshold = 0.05). Tracking was stopped when the larva touched the outer perimeter of the dish or collided with another animal on the stage. BIOImageOperation output was processed using the suite of 'tidyverse' R packages to calculate locomotion parameters. Path straightness was defined as 'total crawling distance' divided by 'distance between origin and end coordinates'. The measure of movement linearity and irregularities were calculated using the R package 'trajr' (McLean & Skowron Volponi, 2018).

3

Characterisation of RNA targets of Syncrip

3.1 Introduction

How does a small population of neural stem cells (NSCs) generate the complex human brain? This question is of major medical importance because many neurological disorders result from aberrant neurogenesis, which can be quite subtle in nature (Pilaz & Silver, 2015; Yano et al., 2016). *Drosophila melanogaster* shares many key features with mammalian neurogenesis, which make it an ideal model to study stem cell behaviour (Bayraktar et al., 2010; Doe, 2017b; Homem & Knoblich, 2012; Yasugi & Nishimura, 2016). In the embryo, neuroblasts (NBs, *Drosophila* NSCs) at different stages express unique TFs that are inherited by their daughter cells, called ganglion mother cells (GMCs), and this subsequently specifies the temporal fate of the neurons. Larval NBs, which contribute to 90% of the adult nervous system, also produce a stereotyped series of distinct neurons that arise in an invariant birth order (Doe & Goodman, 1985; Weigmann & Lehner, 1995). This raises the question of whether there is a mechanism parallel to the embryonic TF cascade that reliably assigns neuronal temporal identity. Several studies have aimed to identify the equivalent regulators of larval NBs through development. Interestingly, *IGF-II mRNA-binding protein* (*imp*) and *syncrip* (*syp*), which code for two evolutionally conserved RBPs, were shown to be expressed in opposing temporal gradients in multiple NB lineages (Z. Liu et al., 2015; Ren et al., 2017; Syed et al., 2017). It was further demonstrated that the two RBPs negatively regulate each other,

maintaining the directionality of the gradient. Although the predominant focus in the field has been on transcriptional regulation, these works highlight an additional important layer of post-transcriptional control, the extent of which has been relatively overlooked (Pilaz & Silver, 2015).

Are Syp and Imp the master regulators of the NB temporal patterning? Or do they act downstream of extrinsic signalling or other transcription factors yet to be identified? The extent to which these RBPs influence neurogenesis is unclear, however, several gene-by-gene level interactors of Syp and Imp have been described. The regulation of *chinmo* translation by Syp and Imp determines mushroom body neuroblast progeny fate (Dillard et al., 2018; Z. Liu et al., 2015), and the two RBPs are involved in various aspects of stem cell behaviour, such as regulation of cell size, rate of division, and termination of stem cell differentiation (Hailstone et al., 2020; L.-Y. Liu et al., 2020; Samuels, Järvelin, et al., 2020; C. P. Yang et al., 2017). These individual investigations illustrate the broad function of Syp and Imp and their capacity to coordinate neural development at a high level. Therefore, it is essential to understand their mechanism of action.

RBPs, such as Imp and Syp, regulate all aspects of the lifecycle of mRNA metabolism, altering its isoform choice, stability, transport, and translation efficiency (Pilaz & Silver, 2015; Van Nostrand, Pratt, et al., 2020). Therefore, to elucidate the factors that guide mRNA specificity and regulatory mechanisms of RBPs, methods that interrogate endogenous protein-RNA interactions are particularly valuable (Huppertz et al., 2014; F. C. Y. Lee & Ule, 2018). Such methods present a greater capacity to explore the principles behind the RBP-RNA interactions via their binding site and their prevalence in the transcriptome.

In this Chapter, I aimed to identify mRNA binding targets of Syp in the third instar (L3) larval brain and in the whole larva using iCLIP. This data supplements previously identified targets of Imp in the L3 brain using RNA immunoprecipitation sequencing (RIP-seq). I investigated the enrichment of Syp targets within the groups of genes known to be involved in neural-specific 3' untranslated region (3'UTR) extensions and a cohort of transcripts temporally regulated in the NBs. I then narrowed my focus to exemplify two functional examples of Syp-mediated post-transcriptional regulation in the type I NB lineage using smFISH and immunofluorescence.

3.1.1 Specific aims of this chapter

1. To identify Syp targets in the developing larval brain with iCLIP
2. To characterise the general features of Syp targets
3. To describe examples of Syp-mediated post-transcriptional regulation of its targets

3.1.2 Experimental rationale

Multiple methods exist to experimentally identify the protein-RNA interactions specific to one RBP, all of which rely on antibody immunoprecipitation or endogenous tagging of target RNAs in close spatial proximity. These methods have varying specificity and sensitivity based on their mode of RBP-RNA capture (discussed in Chapter 1, and reviewed in (F. C. Y. Lee & Ule, 2018; Wheeler et al., 2018)). I wanted to understand how Syp regulates the behaviour of NBs and their progeny in the larval brain, so I aimed to identify Syp targets specifically in the dissected brain to maximise the physiological relevance of the identified targets. While RIP-seq is a commonly used technique for this purpose, it involves purification of the RBP directly from a cell lysate without crosslinking. Because of this, immunoprecipitation is limited to gentle washes that often lead to co-purification of contaminating RBPs. Indeed, a preliminary result of Syp protein interactions profiled using mass spectroscopy (MS) (unpublished data, Francesca Robertson, Davis lab), showed extensive interaction of Syp with other RBPs. Furthermore, RIP-seq is prone to high false-positive rates due to *in vitro* RBP-RNA binding in cell lysates. To avoid this problem, I aimed to use a CLIP based protocol, which utilises UV radiation to form a zero-length covalent bond between the endogenous RBP and the bound RNA. This enables CLIP to purify RNP complexes under conditions that are stringent enough to prevent co-purification of additional RBPs or free RNAs. iCLIP takes another step to allow individual nucleotide resolution of RBP binding sites (Huppertz et al., 2014). Despite the difficulty of the technique, the identification of endogenous targets and the RBP binding sites can offer unprecedented insight into the mechanism of action of Syp. *D. melanogaster* Syp protein contains four RNA-recognition motifs (RRMs), which bind single-stranded RNAs (ssRNAs) through peptide-base interactions (Geuens et al., 2016; F. C. Y. Lee & Ule, 2018; McDermott et al., 2012; Ule et al., 2018). A crystal structure study of human SYNCRIP revealed the ability of

the N-terminal acidic protein domain to interact with miRNA (Hobor et al., 2018). This domain is also present in *D. melanogaster* Syp (Hobor et al., 2018; McDermott et al., 2012), however, the extent of Syp interaction with double-stranded RNA (dsRNA) is unclear from its peptide composition. Therefore, I used the standard iCLIP protocol, instead of hybrid iCLIP (hiCLIP) (Sugimoto et al., 2015), to study the repertoire of Syp-RNA interactions.

3.2 Identification of Syp targets

RBPs exert their function through the regulation of their target transcripts. The post-transcriptional regulation of mRNA stability, translation or transport eventually leads to the regulation of protein expression in space and time. Therefore, finding out the mRNA targets of Syp is important to dissect its regulatory mechanisms.

3.2.1 Expression of Syp in the developing larvae

Syp displays distinct spatial expression patterns across the larval central nervous system (CNS), neuromuscular junction (NMJ) and intestinal stem cells (McDermott et al., 2012, 2014). modENCODE temporal RNA-seq expression data showed the absence of *syp* expression during embryogenesis but the activation of transcription in the L3 stage, whose timing coincides with the Imp-to-Syp temporal transition in the brain and NBs (Z. Liu et al., 2015) (Figure 3.1A). Interestingly, after pupation and eclosion, the expression of *syp* is significantly reduced in females but not in males, suggesting a testes-specific role of Syp. modENCODE protein data aligned well with the RNA-seq pattern (Figure 3.1B). In summary, these data highlight the development-specific expression of Syp. Inspection of GFP-tagged Syp (Syp::GFP) (J. Titlow et al., 2020) using an epifluorescent microscope confirmed modENCODE datasets, where the enrichment of GFP signal was observed in the brain and testes of the male L3 larvae (Figure 3.1C). Taking these results, I decided to prepare iCLIP libraries with dissected brains in order to identify Syp targets specifically in the brain. Additionally, whole-larval iCLIP libraries were prepared to identify Syp targets on the organism level to gain a better understanding of general Syp functions.

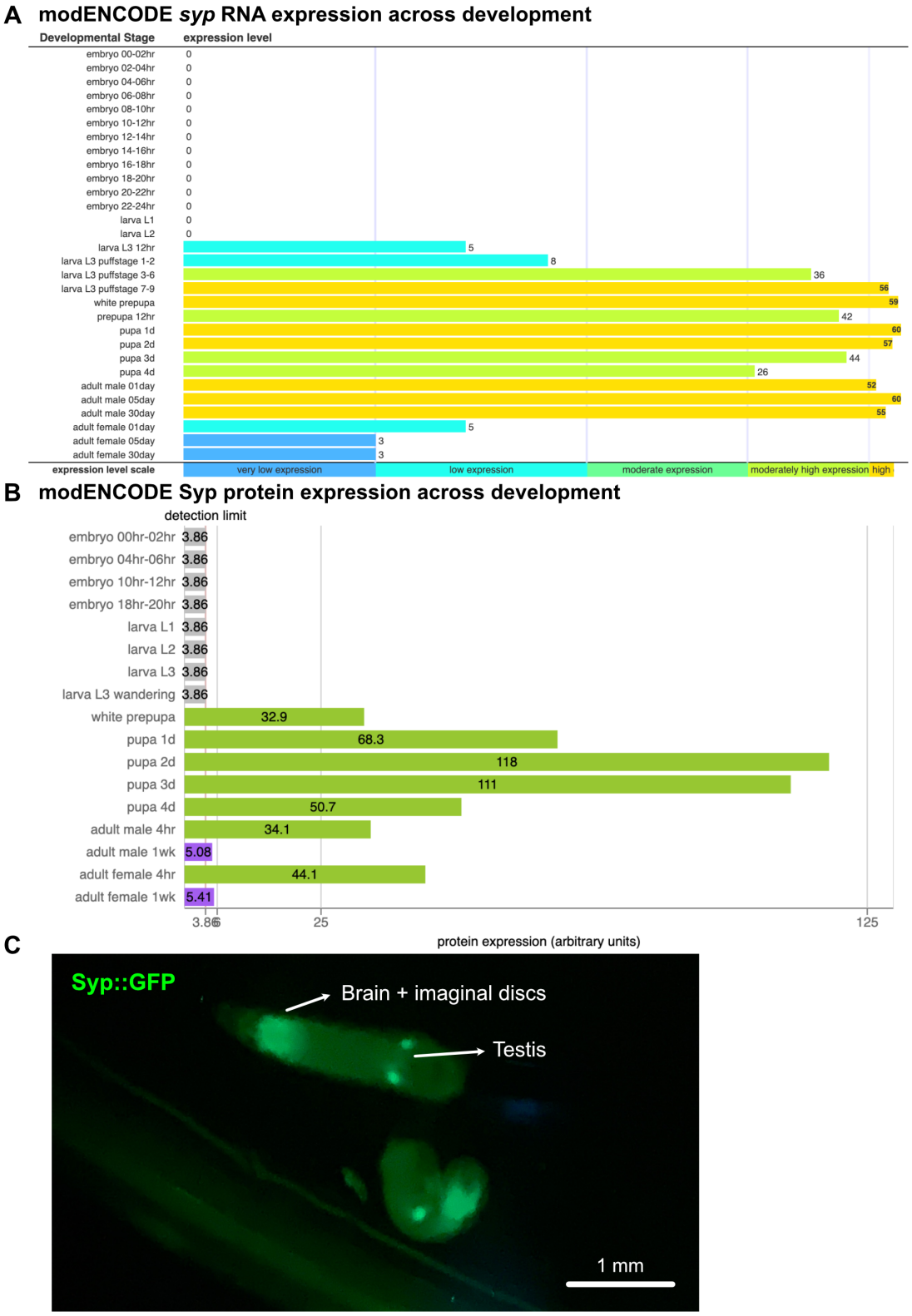


Figure 3.1: Expression of *syncrip* across *Drosophila* development. (**A**) modENCODE *syp* RNA expression as assessed by development-specific RNA-seq. (**B**) modENCODE Syp protein expression across development. modENCODE data were accessed through FlyBase. (**C**) Expression of GFP-tagged Syp in the whole larva. An epifluorescence image of 2 male L3 larvae is shown. (n=3).

In the CNS, Syp expression is induced around 40-50h after larval hatching (ALH), and Imp expression is completely turned off past 72h ALH in the central brain NB lineages (Syed et al., 2017) (Figure 3.2). To investigate dominant Syp targets in the NBs and their progeny neurons, 96h ALH brains were dissected for the iCLIP experiment. The size of the brain is also relatively large at this stage to allow better recovery of RNA per tissue.

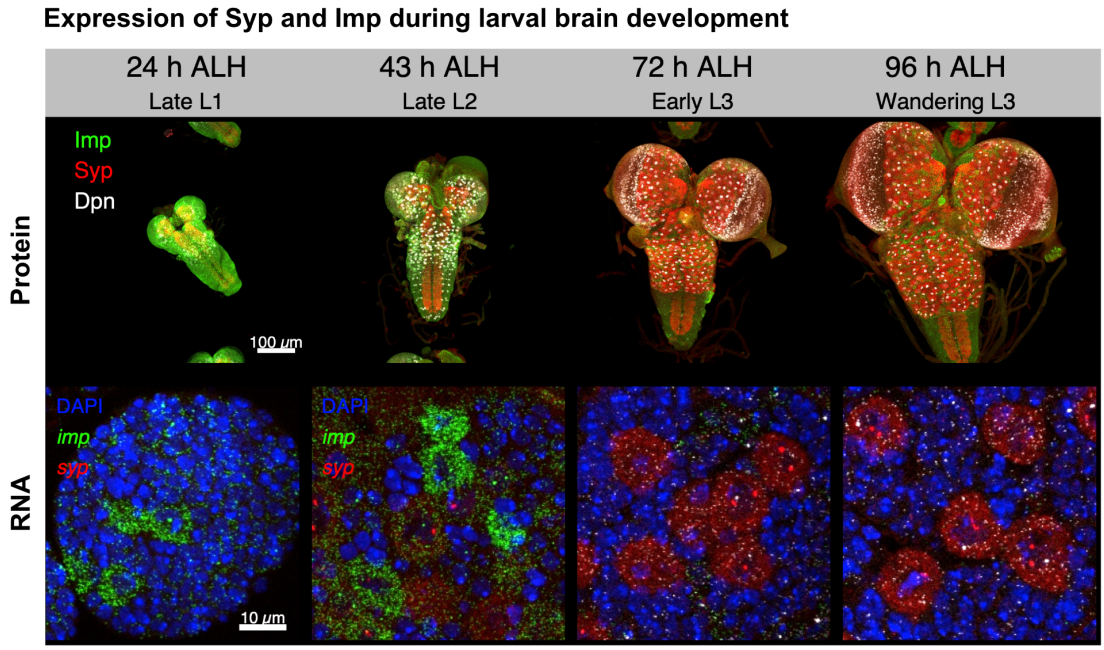


Figure 3.2: Expression of *syncrip* across the brain development. Protein and RNA expression of Imp and Syp across the brain development. Larval brains at designated hours after larval hatching (ALH) were dissected and stained with anti-Imp and anti-Syp or hybridised with *imp* and *syp* exon probes. Anti-Deadpan (Dpn) was used to label neuroblasts. Maximum intensity projected confocal images (50 µm, *upper panels*; 2 µm, *lower panels*) are shown. (n=5 for Immunofluorescence; n=3 for RNA stain).

3.2.2 Adaptation of iCLIP to samples from larval tissues

For the construction of iCLIP libraries, Syp::GFP (N-terminal tagged) fly line (J. Titlow et al., 2020) was used for immunoprecipitation using high-affinity GFP nanobodies (Chromotek). The libraries were prepared using the recently developed iCLIP2 protocol (Buchbender et al., 2019) with modifications to improve RNA recovery when working with a small amount of starting materials. A schematic of the protocols is presented in Figure 3.3. In this protocol, crosslinked Syp-RNA complexes are immunoprecipitated and the bound RNAs are partially trimmed to yield RNA fragments suitable for Illumina sequencing. Immunoprecipitated complexes are separated

via SDS-PAGE and transferred to nitrocellulose, which removes free RNA contaminants. From the nitrocellulose membrane, appropriate size regions containing the RBP-RNA complexes were excised and the RNA recovered via proteinase K digestion. Purified RNAs were then reverse transcribed, where the truncation of the reverse transcription occurs at the site of crosslinkage. cDNAs were ligated with Illumina sequencing adapters and size-selected using solid-phase reversible immobilisation (SPRI) beads. Major changes from the original iCLIP2 protocol are as follows:

1. Use of infrared adapter (Zarnegar et al., 2016) to visualise crosslinked RBP-RNA complexes to avoid radioactivity.
2. Use of adapter ligation buffer free of reducing agents to prevent unwanted elution of Syp::GFP during on-beads enzymatic reaction.
3. Preparation of size-matched input (SMInput) libraries, which represent the background crosslink-ome (Van Nostrand et al., 2016; Wheeler et al., 2018). These libraries were prepared by running the 10 µl of crosslinked lysate on SDS-PAGE and transferring it to nitrocellulose. The same size region, hence size-matched, was excised for the cDNA preparations. SMInput samples were used to filter for high-confidence binding sites during the computational analysis (Busch et al., 2019).

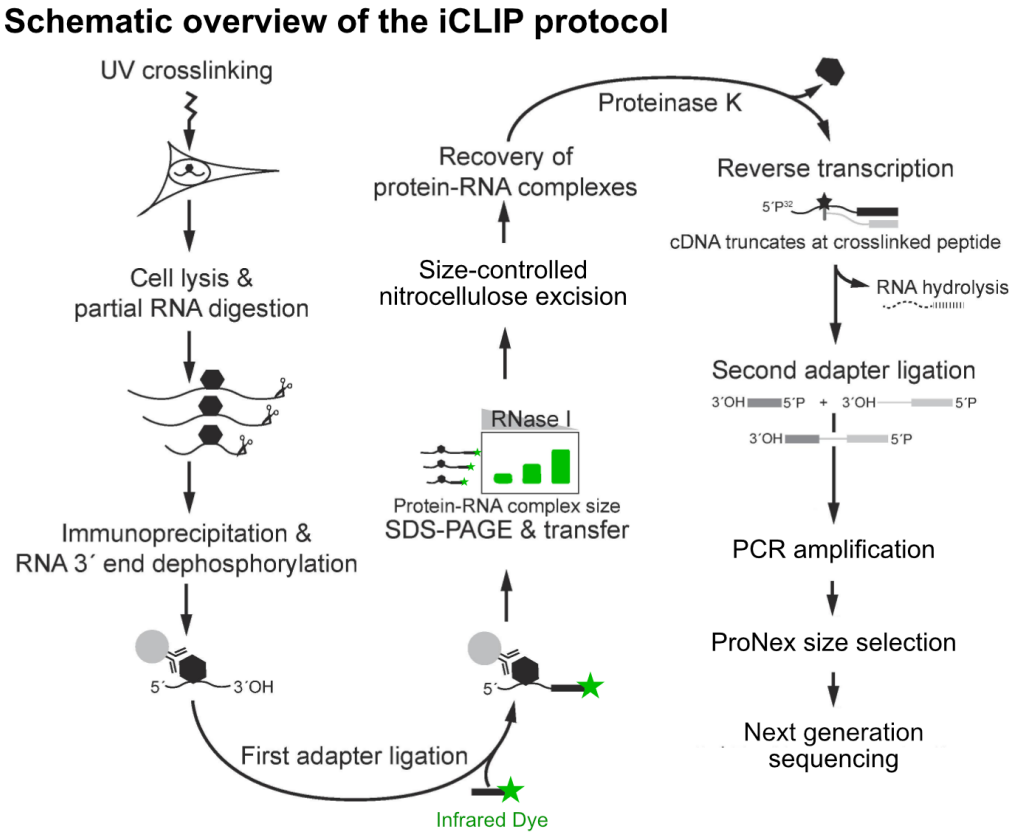


Figure 3.3: Overview of the modified iCLIP2 method. Schematic representation of the different steps of the iCLIP2 workflow. Note that infrared adapters were used instead of radioactive labelling in the original protocol (see Chapter 2 for details).

UV crosslinking strategy

UV-crosslinking at 254 nm induces the formation of a covalent bond between RNA and protein, capturing the *in vivo* binding pattern (Ule et al., 2003). Highly specific interactions are recovered as the crosslinking events only occur when both protein and RNA are in extreme proximity. L3 brains are small in size for allowing proper UV penetration, making them amenable to apply the original iCLIP method, using direct crosslinking. Dissected brain tissues were irradiated multiple times for even coverage of dorsal and ventral sides. In contrast, whole larval samples were not fit for direct crosslinking because of the large size and opaque cuticle layer. Therefore, I adapted a crosslinking strategy that has been applied to plant tissues (Meyer et al., 2017). Here, the specimen is ground up into frozen powder in liquid nitrogen and subsequently UV irradiated

in powder form. This way, the endogenous RBP-RNA binding events are recovered by 'freezing' the interactions. Crosslinked samples were then homogenised in lysis buffer.

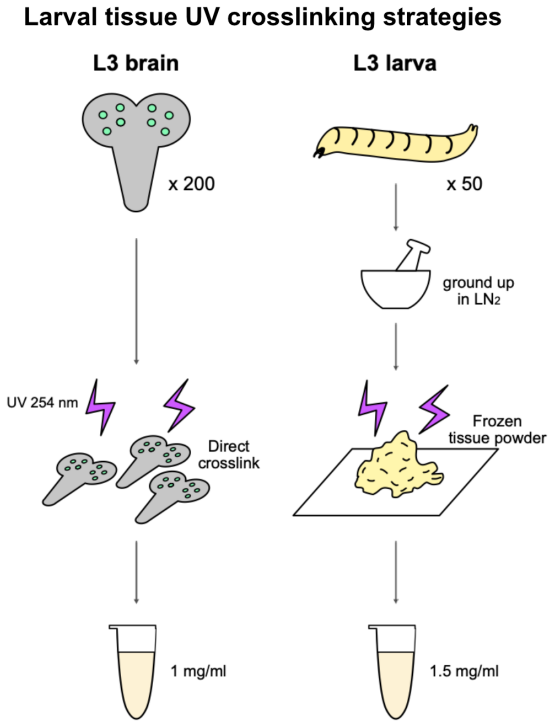


Figure 3.4: UV crosslinking strategies of the larval tissues. Schematic representation of the UV crosslinking method used to covalently link Syp::GFP and its target RNA. Brain tissues were directly crosslinked using 254 nm irradiation. Whole-larval samples were snap-frozen in liquid nitrogen and ground into fine frozen power and subsequently crosslinked.

I assessed the result of UV crosslinking by purifying Syp::GFP-RNA complexes from the lysates. After the immunoprecipitation of Syp::GFP, the infrared adapter was ligated to the Syp::GFP-bound RNAs (Figure 3.3). Following SDS-PAGE separation and nitrocellulose transfer, I found Syp::GFP-RNA complexes were visible as a strong infrared signal that migrated at a higher molecular weight (MW) range from 'free' Syp::GFP (Figure 3.5). As expected, the signal was abolished in non-irradiated samples, indicating successful crosslinking of Syp::GFP with its target RNAs. The signal often appears as a smear, representing the different lengths of the target transcripts immunoprecipitated with Syp::GFP. I noted that stronger UV irradiation (250 mJ/cm^2) did not yield improvement in the infrared RNA signal and therefore low-level UV (150 mJ/cm^2) was used throughout the experiment. Syp::GFP-RNA complexes were liberated from

the membrane by excising the membrane regions highlighted in the yellow dashed box in Figure 3.5, followed by proteinase K treatment. The purified RNA is still covalently attached to a short Syp::GFP peptide, which stalls and truncates reverse transcriptase activity at this site, hence providing the individual nucleotide resolution of Syp::GFP binding.

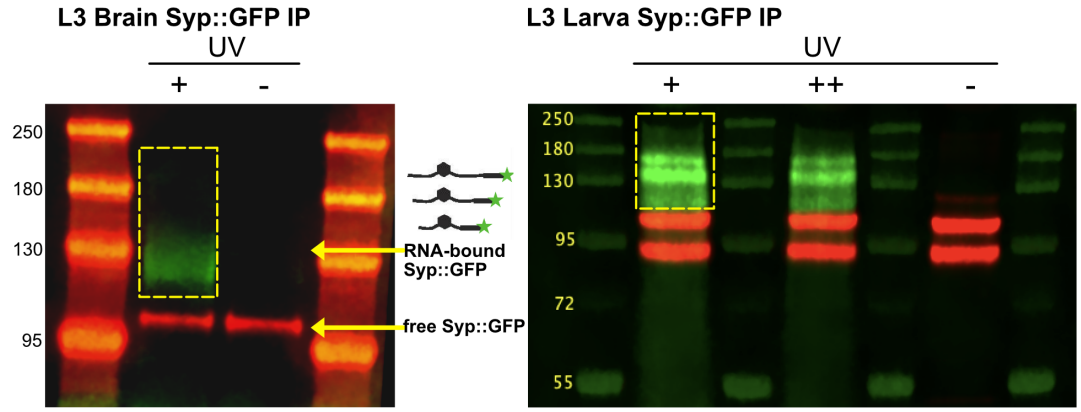


Figure 3.5: Biochemical purification of Syp::GFP-RNA complexes. Infrared images of Syp::GFP-RNA complexes transferred to a nitrocellulose membrane after SDS-PAGE. Western blot using anti-Syp was used to visualise 'free' unbound Syp::GFP (red). Syp::GFP-bound transcripts were visualised by ligating the infrared IR800 adapter using RNA ligase I (green). UV crosslinked and non-crosslinked samples are shown. '+' indicates UV strength of 150 mJ/cm^2 and '++' indicates 250 mJ/cm^2. Representative nitrocellulose image from three replicate experiment is shown.

iCLIP library preparation and assessing the reproducibility

Next, the truncated cDNAs were ligated with Illumina adapters containing unique molecular identifiers (UMIs). The UMIs label distinct cDNA molecules with a unique stretch of sequences, which are used to collapse PCR-duplicated reads after sequencing. Such methodology greatly reduces potential biases caused by over-amplification of low-complexity libraries (Landskron et al., 2018), which is often the case with CLIP-based protocols (Buchbender et al., 2019; Lee et al., 2018a; Van Nostrand et al., 2016). To examine the complexity of Syp iCLIP libraries I examined the optimal number of PCR cycles required to amplify enough material, which was assessed based on the visibility of DNA libraries on TBE PAGE gels with SYBRGOLD (Huppertz et al., 2014). I found that my iCLIP (IP) and SMInput (Input) libraries were sufficiently amplified with an average of 12-13 PCR cycles (Figure 3.6A-B). In contrast, my negative control samples (Control), where immunoprecipitation was performed without the GFP-nanobody, required more than 20

cycles. The range for IP and SMInput samples was well below the maximum number of cycles recommended by the iCLIP2 protocol (Buchbender et al., 2019), suggesting a high complexity of these libraries. The high cycle number required for the Control samples also demonstrates specific enrichment of Syp::GFP-RNA complexes in my experimentation.

After sequencing, raw reads were aligned to the *D. melanogaster* genome (ENSEMBL v99) to map the crosslinked sites. Then, I extracted the UMI sequences from the reads and collapsed duplicated reads using 'umi_tools' (Smith et al., 2017). Consistent with the low PCR cycles required for the amplification, IP and SMInput samples on average contained 12% of PCR-duplicated reads, while the no-nanobody control libraries were heavily duplicated beyond 75% (Figure 3.6C).

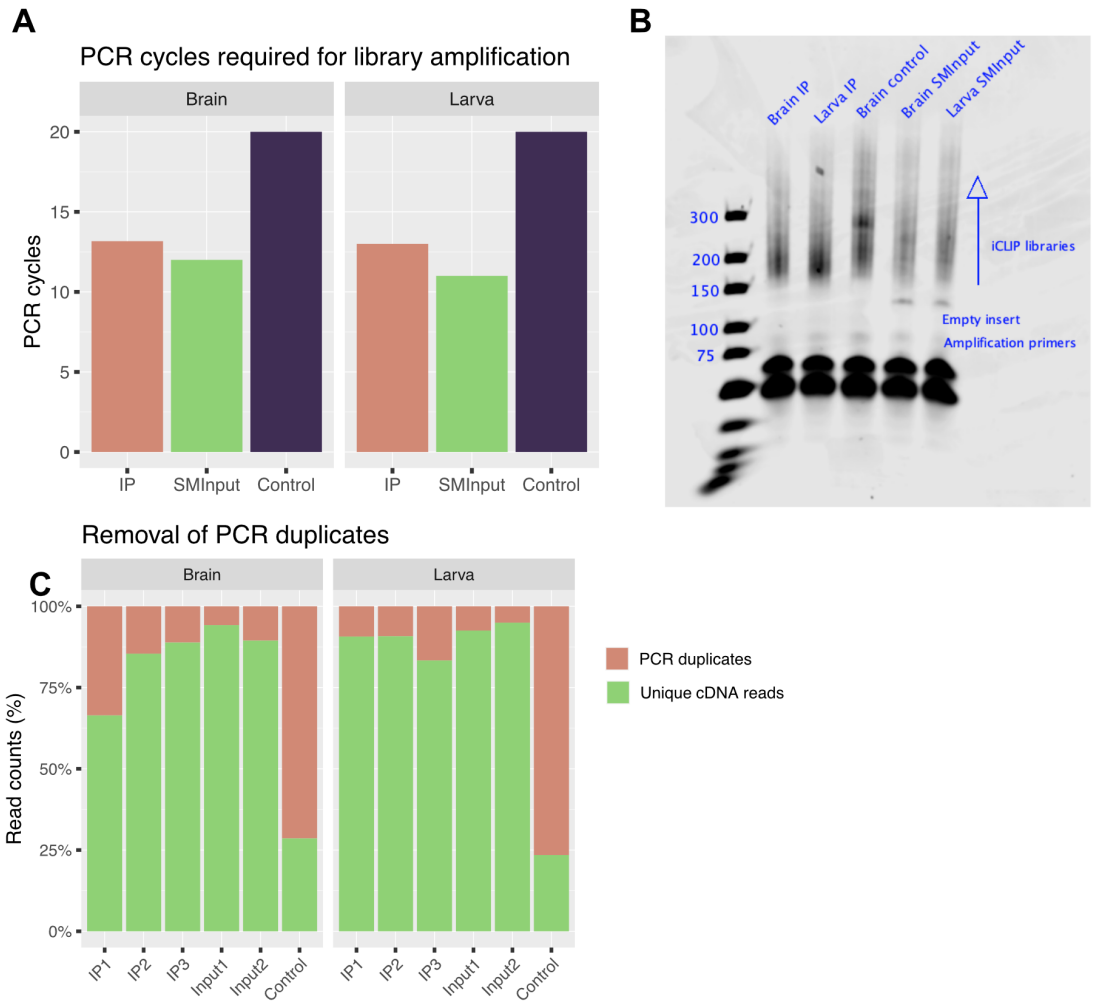


Figure 3.6: Quality control of iCLIP libraries. (**A**) The number of PCR cycles required to amplify a sufficient amount of cDNA molecules. A sufficient amount of library molecules was assessed by visualisation with SYBRGOLD after 6% TBE PAGE gel electrophoresis. IP=Syp immunoprecipitated iCLIP libraries, SMInput=Size-matched input libraries, Control=No-nanobody control. (**B**) Example TBE PAGE of amplified libraries stained with SYBRGOLD. IP, SMInput and Control libraries were amplified with 13, 12 and 20 PCR cycles, respectively. Empty insert and adapter DNAs were removed using beads or gel purification before sequencing. (IP libraries, n=3; SMInput libraries, n=2, Control libraries, n=1 for brain and whole-larval biosamples). (**C**) Percentage of PCR duplicated reads after deduplication of sequencing reads using unique molecule identifier (UMI) sequences.

I assessed the reproducibility of the iCLIP dataset by extracting the number of crosslinked reads per gene, which was normalised to the library size (reads per million, RPM). Then, I performed correlations between the replicates. The correlation plots showed the biological replicates are very well correlated in both the brain and whole-larval datasets (R=0.92-0.98) (Figure 3.7A-B). Taken together, these results highlight the successful preparation of high complexity Syp iCLIP libraries.

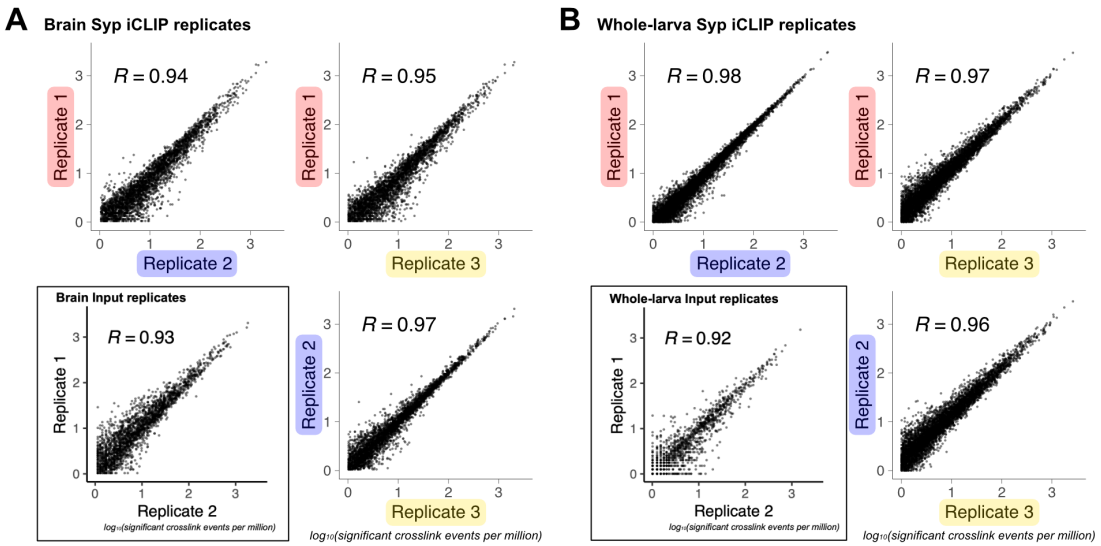


Figure 3.7: Correlation plots to show reproducibility of iCLIP. Correlation plots between the triplicates of IP samples or duplicates of SMInput libraries in the (**A**) L3 brain or (**B**) L3 whole-larva. Each symbol represents crosslink density (normalised to library size) for each gene. The plots are shown in logarithmic scales. The correlation was measured by Pearson's method in base R. (IP libraries, n=3; SMInput libraries, n=2 for brain and whole-larval biosamples).

3.2.3 Identification of transcriptome-wide Syp binding sites

Peak calling

Syp crosslinks scattered throughout the transcriptome were further processed to define Syp binding sites (Busch et al., 2019) (Figure 3.8A). Typically, RBP binding on its target RNA leaves a footprint reminiscescent of 'peaks', which represents a cluster of truncated reverse transcription events (F. C. Y. Lee & Ule, 2018) (see Figure 3.8B for schematic diagram). The 'peak-calling' procedure is equivalent to the identification of transcription factor binding sites used in ChIP-seq or similar protocols. Therefore, these binding sites describe the nature of Syp interaction with its target RNA. Next, high confidence binding sites were determined. The number of crosslinks (reads per million, RPM) values were extracted and compared between IP and SMInput samples. To increase the stringency of binding site identification, those with at least 3x RPM enrichment (Brain iCLIP) or 10x RPM enrichment (Whole-larval iCLIP) were retained. The enrichment threshold against the SMInput was determined by correlating the number of binding sites versus the RPM cutoff and identifying plateau values. Finally, I incorporated reproducibility of each binding site based on replicate information (Busch et al., 2019). I compared the top 85% quantile

binding sites between biological replicates and filtered to only keep the sites supported by all three replicates (Figure 3.9A-B). Using this procedure, I identified 16,305 high-confidence Syp binding sites in the brain and 23,112 sites in the whole larva.

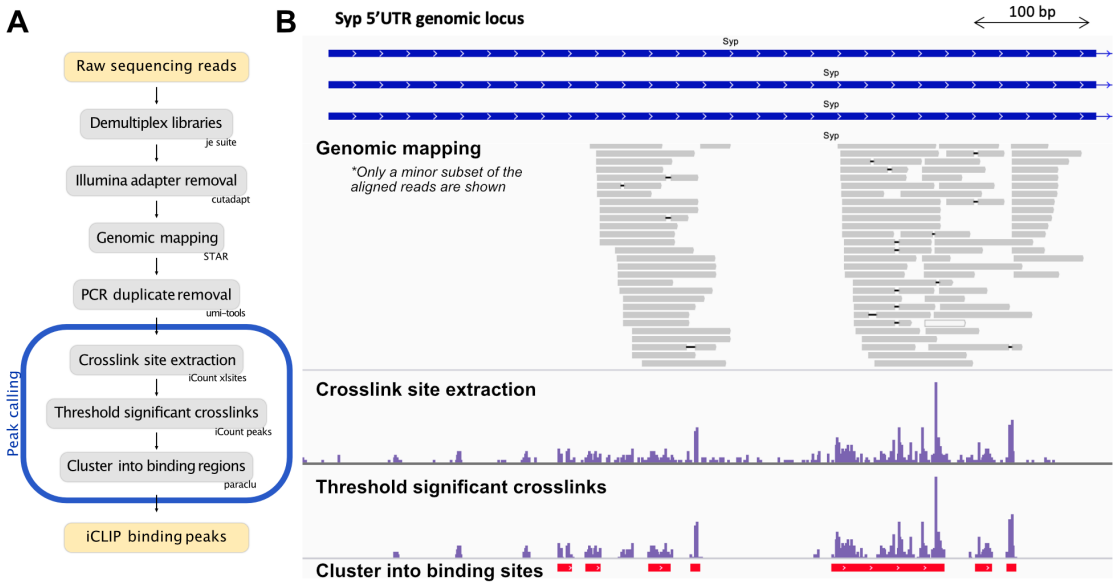


Figure 3.8: Overview of computational analysis of iCLIP. (**A**) A workflow summarising the steps of the iCLIP analysis pipeline. Tools used for each step are shown in the lower right corner. Briefly, iCLIP reads mapped to the *D. melanogaster* transcriptome were deduplicated to remove PCR amplification artefacts using unique molecule identifiers (UMIs). The crosslink sites were extracted from the aligned reads by summing the 5' most read positions where the reverse transcriptase truncations occur (Blue tracks in (B)). Then, significant crosslinks were obtained via equal density kernel comparison method (see Methods for details), which were clustered together to form binding regions (Red regions shown in (B)). (**B**) Schematic overview of peak-calling analysis (highlighted blue in (A)). *syp* 5'UTR region is shown as an example. The gray symbols represent iCLIP reads mapped to this loci. Note that only a very minor subset of reads are shown in this figure. The blue track represent the accumulation of crosslinks at each nucleotide positions. Clustered binding sites are shown in red.

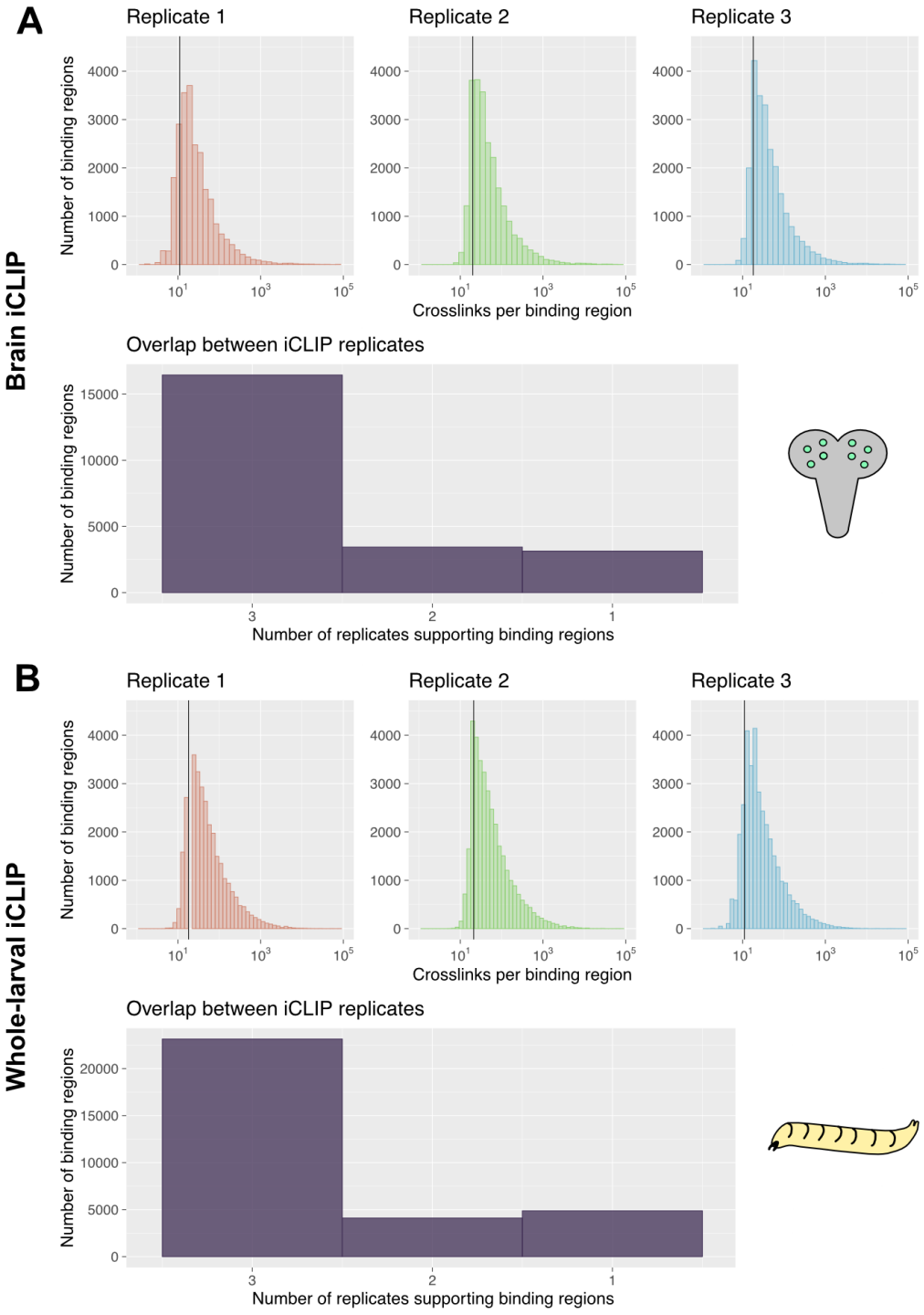


Figure 3.9: Reproducibility between replicate experiments. Histogram of crosslink events per binding site in the three Syp iCLIP replicates from (**A**) brain or (**B**) whole-larval biosamples (*upper panels*). 15% quantiles are indicated by the vertical lines. Overview of the number of binding sites that are shared between biological replicates (*lower panel*). Only the binding sites supported by all three replicates were kept for the downstream analysis.

Distribution of Syp binding sites within its targets

I found Syp binding sites were enriched in protein-coding mRNAs and non-coding RNAs (ncRNAs). Interestingly, Syp binding sites were largely present in the 3'UTR of mRNAs (Figure 3.10A), consistent with its role as a mediator of post-transcriptional control of RNA translation and stability (McDermott et al., 2014). I found that typically Syp decorates its target's 5'UTR and 3'UTR regions (exemplified in Figure 3.10B, with *jim* transcript). Indeed, a meta-transcript profile of Syp binding revealed bi-modal coverage across the transcript at the extreme ends. This highlights the preferential binding of Syp to the untranslated regulatory elements (Figure 3.10C).

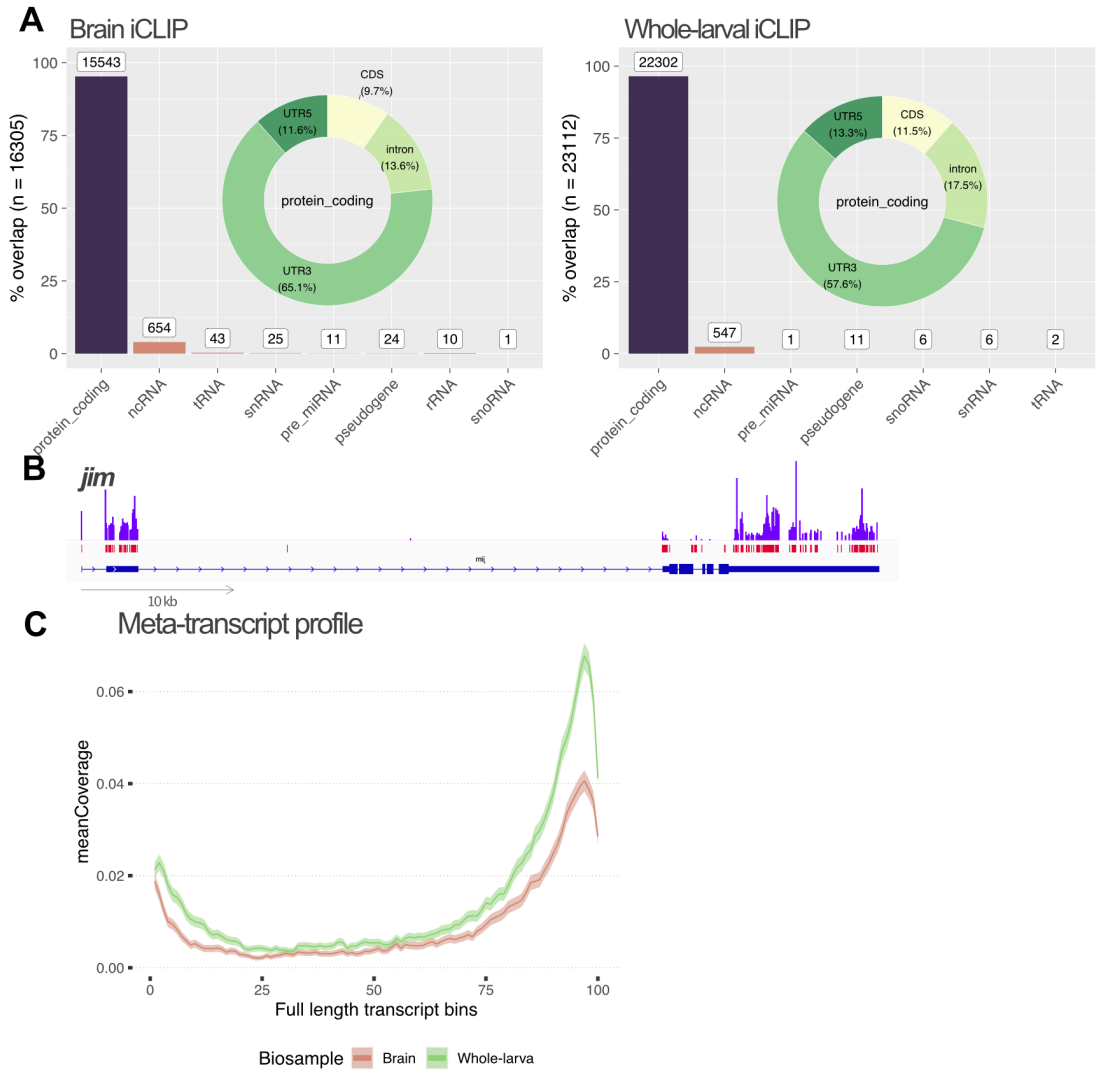


Figure 3.10: Transcriptome-wide distribution of Syp binding sites. (**A**) Assignment of Syp binding sites with respect to RNA biotypes. Binding sites within protein-coding genes were further categorised into transcript features (*pie chart*). (**B**) Typical binding site distribution of Syp on its targets. The transcript *jim* is shown as an example to illustrate enrichment of Syp binding to the 5'UTR and 3'UTR. The blue track represents crosslink density coverage (Reads per million) and the red regions represent significant Syp binding sites that are enriched over the size-matched input libraries. The coverage shows average of the three biological replicates. (**C**) Meta-transcript profile of Syp binding site distributions. The ribbon widths represent 90% confidence intervals. The meta-profile analysis was limited to protein-coding genes.

Next, I analysed Syp binding at the target gene level. I found that Syp targets 1,476 genes in the L3 brain and 1,910 genes at the whole-organism level (Figure 3.11A). The full lists of Syp target transcripts are given in Appendix B and Appendix C. Notably, when I arranged the target transcripts based on the number of binding sites, there were frequent occurrences of transcripts

encoding nucleic acid-binding proteins (Figure 3.11B). Additionally, cytoskeletal binding proteins were featured in larval samples, potentially hinting at the function of Syp as a regulator of cytoskeleton reorganisation (J. Titlow et al., 2020). A comparison of the brain and whole-larval target genes showed that there is a significant overlap of ~1,000 genes, but around ~500 targets were only recovered in the brain iCLIP (Figure 3.11C). Many genes in this group turned out to be specifically expressed in the larval CNS, therefore the use of dissected brains for iCLIP has recovered brain-specific targets of Syp and their binding sites.

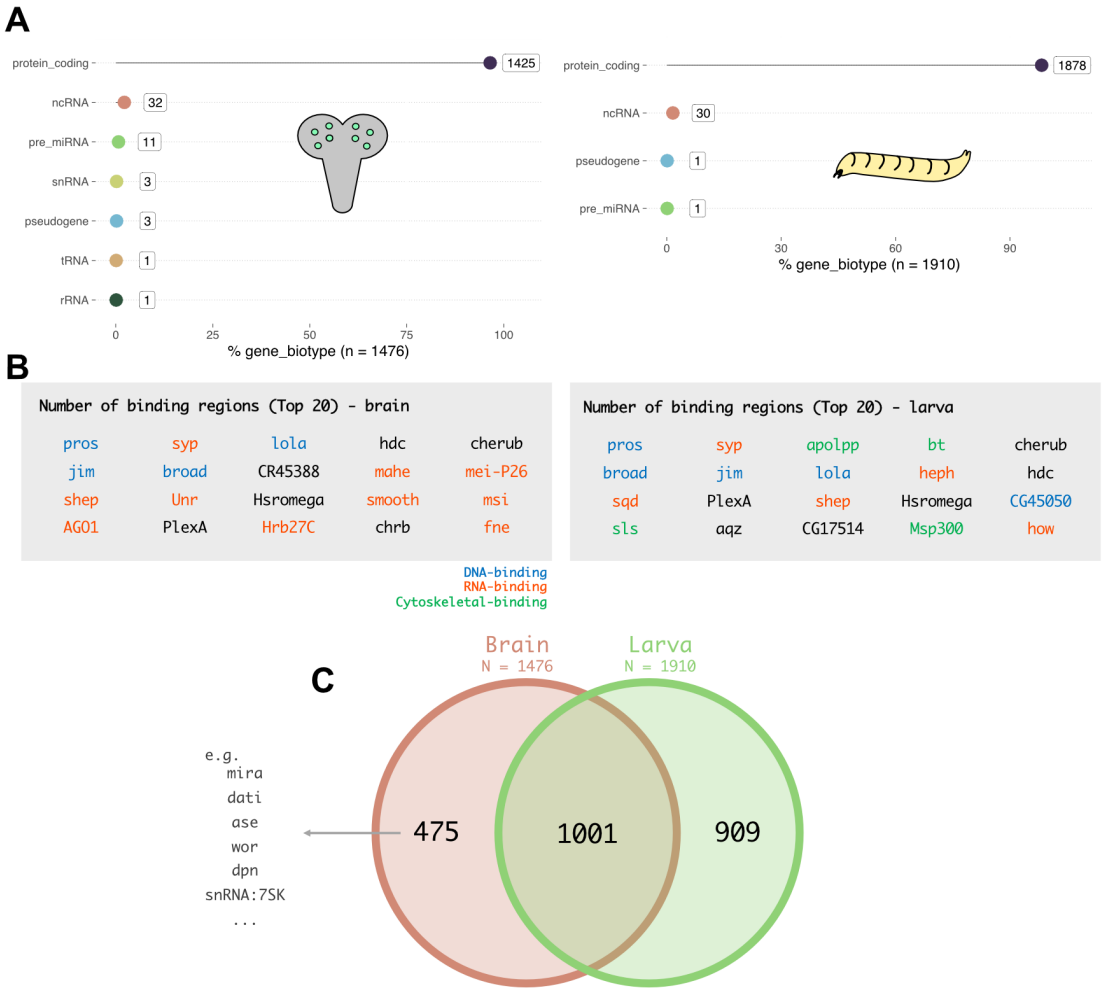


Figure 3.11: Identification of Syp targets. (**A**) Assignment of Syp binding that falls within genes of a given type. (**B**) Top 20 Syp target genes, arranged by the number of binding sites. The genes are colour-coded based on their molecular function (DNA, RNA or cytoskeletal binding). (**C**) Comparison of Syp targets between the brain and whole-larval samples. Examples of brain-specific targets are shown.

3.3 General characteristics of Syp targets

iCLIP has identified thousands of targets of Syp in the developing larvae, which collectively shapes the Syp regulatory network. To gain further understanding of the role of Syp within this complex network, I explored the functional annotation of Syp targets and cross-referenced the targets to the lists of genes with functional relevance in *D. melanogaster* nervous system development.

3.3.1 Syp targets in the brain

Functional annotations

To explore the biological processes orchestrated by Syp and its targets, I performed gene ontology (GO) enrichment analysis of Syp targets. For the background, an L3 brain transcriptome (Samuels, Järvelin, et al., 2020; Samuels, Arava, et al., 2020) was used to avoid false enrichment of nervous system-related terms. GO analysis revealed a highly significant enrichment of terms related to 'neurogenesis' and 'nervous system development' (Figure 3.12), consistent with my hypothesis of Syp as an important regulator of larval neurogenesis. Notably, GO terms related to 'RNA-binding' were also highly enriched. Corroborating this result, Interpro domain analysis of Syp targets showed enrichment of RNA-recognition motifs (RRMs) and K-Homology (KH) domains, mainly found in RBPs. KEGG pathway analysis also highlighted several signalling pathways that have been shown to regulate self-renewgal behaviour of NB (Notch, Hippo), neuronal differentiation and synaptic transmission (Hedgehog, TGF-beta) (Arefin et al., 2019; Contreras et al., 2018; Halstead et al., 2014; Poon et al., 2016; Rossi & Desplan, 2020).

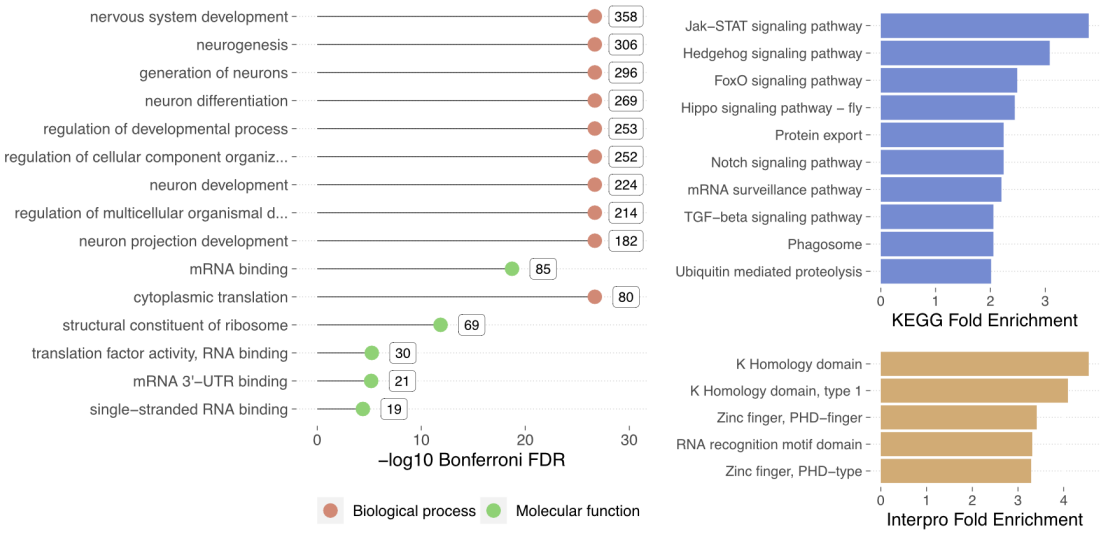


Figure 3.12: Functional annotation of Syp targets in the brain. Gene ontology (GO), KEGG pathway and Interpro domain enrichment analysis of Syp targets in the brain. Brain transcriptome was used as a background for the analysis. Each term is colour-coded based on the ontology type. GO analysed with the R package 'topGO'. KEGG and Interpro analysis was performed through the 'DAVID' server.

Enrichment of RBPs and TFs among Syp targets

I was interested in further exploring the types of targets bound by Syp. The prevalence of nucleic-acid binding proteins such as TFs and RBPs within Syp targets could raise an interesting hypothesis that Syp could be regulating the level of different TFs that are classically thought to determine cell fate. Therefore, I retrieved a curated set of gene annotations from GLAD (Hu et al., 2015) to quantify the enrichment of TFs and RBPs. Additionally, I included a list of Imp targets identified using RIP-seq (Samuels, Järvelin, et al., 2020) in the L3 brain to assess the same feature. Compared to the frequency in the protein-coding genome, Syp and Imp targets were enriched for TFs (DNA-binding) (Table 3.1). Interestingly, targets of both Syp and Imp showed even greater enrichment. This result suggests that both Syp and Imp could be functioning upstream of multiple TFs governing cell fate and function. In line with this idea, the regulation of TFs such as *chinmo* and *myc* by Syp/Imp has been described (Z. Liu et al., 2015; Samuels, Järvelin, et al., 2020). While the precise function of many more TFs is less well understood, the dysregulation of target TFs may explain the phenotypes of *syp* and *imp* mutants. Syp and Imp targets also showed an enrichment of transcripts encoding RBPs (Table 3.1). Many RBPs known to be involved in the nervous

Table 3.1: Enrichment of TFs and RBPs in Syp and Imp targets

Category	Number of protein-coding genes	DNA-binding	Fisher's p-value (TF)	RNA-binding	Fisher's p-value (RBP)
Whole genome	13936	1146 (8.2%)	NA	576 (4.1%)	NA
Syp targets	1351	218 (16.1%)	3.9e-24	190 (14.1%)	1.5e-56
Imp targets	605	101 (16.7%)	2.8e-12	60 (9.9%)	2.3e-10
Syp & Imp targets	255	53 (20.8%)	2.1e-10	49 (19.2%)	8e-20

system function (eg. *elav, squid, pum, heph, hrb27c, and brat*) were targeted by both Syp and Imp. This result supports the idea of Syp/Imp as components of a larger post-transcriptional regulatory network. To date, *lin28* has been described in detail in for its role in Kenyon cell fate determination (Z. Liu et al., 2015), but the roles of other RBPs in neuron specification remain to be elucidated.

Syp binds to transcripts with long 3'UTRs

iCLIP data uncovered Syp binding is mostly found in the UTRs of the target transcripts (Figure 3.10A). UTRs often act as a cis-regulatory elements of post-transcriptional regulation, and they are capable of undergoing evolution to diversify the regulatory regimes without changing the coding potential of the transcript. The transcriptomic landscapes of 3'UTRs are especially diverse in the nervous system. Earlier reports have described general lengthening of the 3'UTR and the preference towards longer 3'UTR isoforms in neural tissues (Hilgers et al., 2011, 2012; Malka et al., 2017; Smibert et al., 2012) that may impact neuronal functions via regulation of mRNA transport or stability (Tushev et al., 2018). In *Drosophila*, the *mei-P26* gene, encoding a ubiquitin ligase, harbours the longest 3'UTR in the genome reaching 19 kb. Interestingly, mei-P26 is a target of both Syp and Imp. In the Davis lab, we have also examined another co-target, *prospero (pros)*, encoding a TF involved in NB cell cycle termination, that features a 15.5 kb 3'UTR.

I analysed the max 3'UTR lengths of all transcripts genome-wide, in the brain transcriptome, and those that interact with Syp and/or Imp. The analysis showed there was a general enrichment of transcripts with longer 3'UTRs in the brain transcriptome, consistent with previous reports (Bae & Miura, 2020; Hilgers et al., 2011) (Figure 3.13A). Examining the max 3'UTR lengths of the Syp/Imp co-targets resulted in a further enrichment for longer 3'UTRs. Notably, the targets included many of the transcripts with the longest 3'UTRs in the brain transcriptome. Next, I cross-referenced the list of transcripts that have been shown to undergo neural-specific extension

of 3'UTRs (eUTR) (Hilgers et al., 2011; Smibert et al., 2012). Interestingly, I found a highly significant overlap of neural eUTR transcripts with the targets of Syp (219/400 genes) (Figure 3.13B). The enrichment of long 3'UTRs may suggest the functional role of longer regulatory sequences that are bound by RBPs such as Syp and Imp (Mayr, 2017). Indeed, brain iCLIP revealed that Syp binds to these extended 3'UTR sequences (Figure 3.13C). Previous Imp iCLIP in *Drosophila* S2 culture cells (H. T. Hansen et al., 2015) also reported predominant localisation of Imp binding sites to 3'UTRs. However, because S2 cells do not typically express the extended 3'UTR isoforms (Wei et al., 2020), it remains to be identified whether Syp and Imp both exert their function through the extra regulatory sequences in the brain.

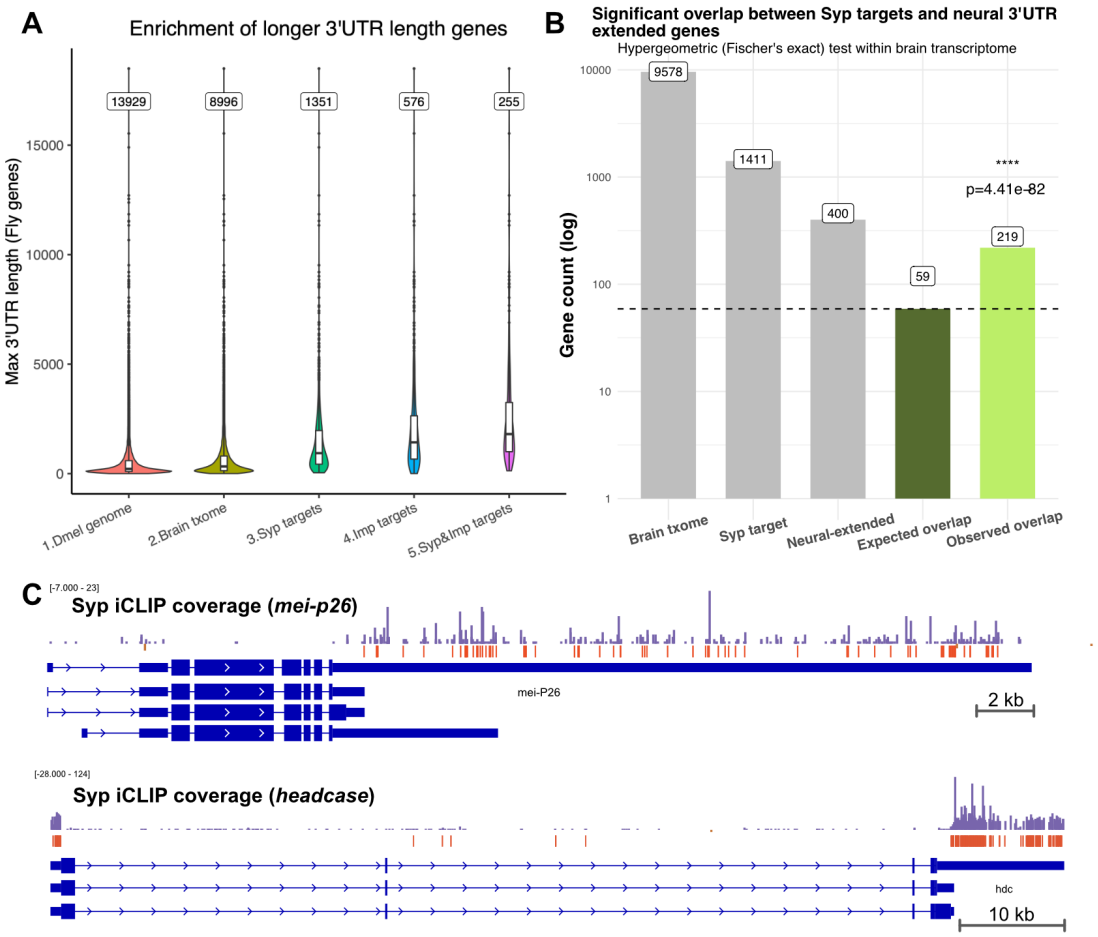


Figure 3.13: Syp binds genes with long 3'UTRs. (**A**) Analysis of maximum 3'UTR lengths across the *D. melanogaster* genome, brain transcriptome, Syp targets and Imp targets. The text labels represent the number of genes in each category. (**B**) Significant overlap between Syp targets and transcripts that have extended 3'UTRs specifically in the nervous system. The statistic of the overlap was calculated using a hypergeometric test implemented in base R. ****p<0.0001. (**C**) Syp iCLIP coverage of two neural-specific 3'UTR extended transcripts, *mei-p26* and *headcase*, showing extensive Syp binding in the extended 3'UTR. iCLIP crosslink densities are shown in blue and binding sites are shown in red. Read per million (RPM) ranges are shown in the upper left corner. The average RPM of the three replicate IP libraries are shown.

Syp binds to temporally regulated transcripts within NBs

Syp has received attention as a potential regulator of NB temporal fate patterning (Doe, 2017b; Z. Liu et al., 2015; Syed et al., 2017). Therefore, I asked whether Syp targets are enriched for temporally regulated transcripts during the NB development. I re-analysed RNA-seq data of purified antennal lobe (ALNB, Type I), Type II (T2NB), and mushroom body (MBNB) NBs sampled at 4 different stages of the Imp-to-Syp gradient progression (Z. Liu et al., 2015; Ren et al., 2017). To identify transcripts that change abundance over time, I used MaSigPro (Conesa et

al., 2006) to find clusters of transcripts with particular temporal expression patterns. I considered 6 scenarios based on the shape-profile of the clusters (Figure 3.14A-C):

1. 'Down-regulated' genes show a decrease in abundance with increasing Syp.
2. 'Up-regulated' genes show an increase in abundance with increasing Syp.
3. 'V-regulated' genes show higher expression only with high Imp or high Syp, demonstrating a 'V-shape'.
4. 'Peak-1' genes show peak expression in high Imp conditions.
5. 'Peak-2' genes show peak expression in high Imp + low Syp conditions.
6. 'Peak-3' genes show peak expression in low Syp + high Syp conditions.

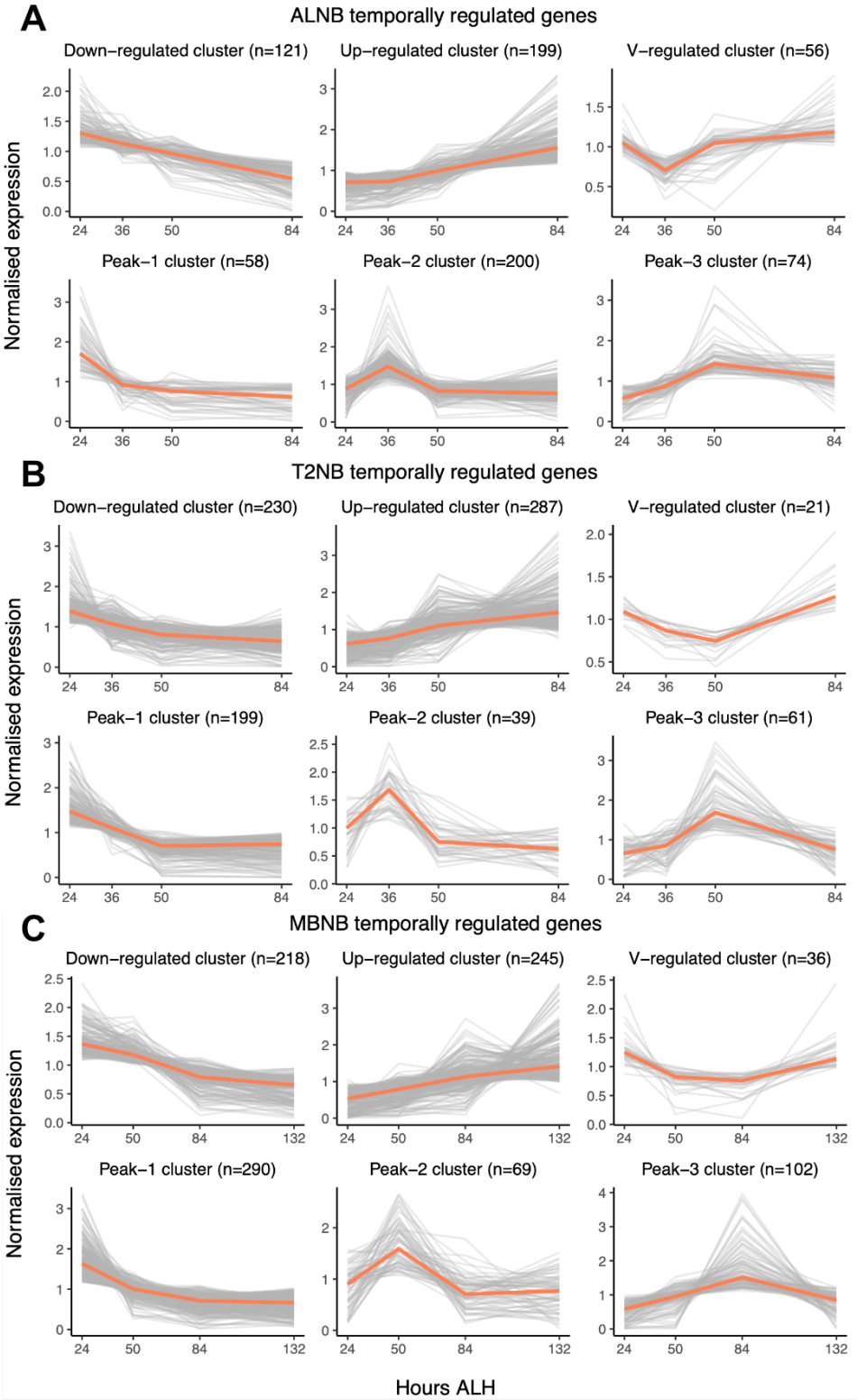


Figure 3.14: Cluster analysis of temporally regulated NB transcripts. Cluster analysis of temporal RNA-seq obtained from purified (**A**) antennal lobe (ALNB), (**B**) type II (T2NB) and (**C**) mushroom body (MBNB) neuroblasts at four stages during larval neurogenesis. 6 clusters were identified using MaSigPro. The median expression pattern for each cluster is shown with the red line. Gray lines represent all temporal profiles of genes in that cluster.

I identified 500-600 genes showing distinct temporal expression patterns from the three NB datasets, which were classifiable into the six clusters (Figure 3.14A-C). Then, I examined the proportion of Syp targets within each cluster. The analysis showed that Syp targets are enriched in all 6 categories above the random chance of intersection (Figure 3.15), however a particularly high level of overlap was observed with the 'Down-regulated' and 'V-regulated' clusters. The 'Down-regulated' transcripts included genes like *chinmo*, which have been described to be negatively regulated by Syp to suppress early neuronal fate in Kenyon cells (Dillard et al., 2018; Z. Liu et al., 2015). The defining feature of 'V-regulated' genes was unclear, but perhaps this pattern may be a indication of a complex interplay of both post-transcriptional and transcriptional regulations. In summary, these data illustrate that Syp binds to many temporally regulated transcripts in NBs with a particular preference for transcripts displaying specific temporal profiles.

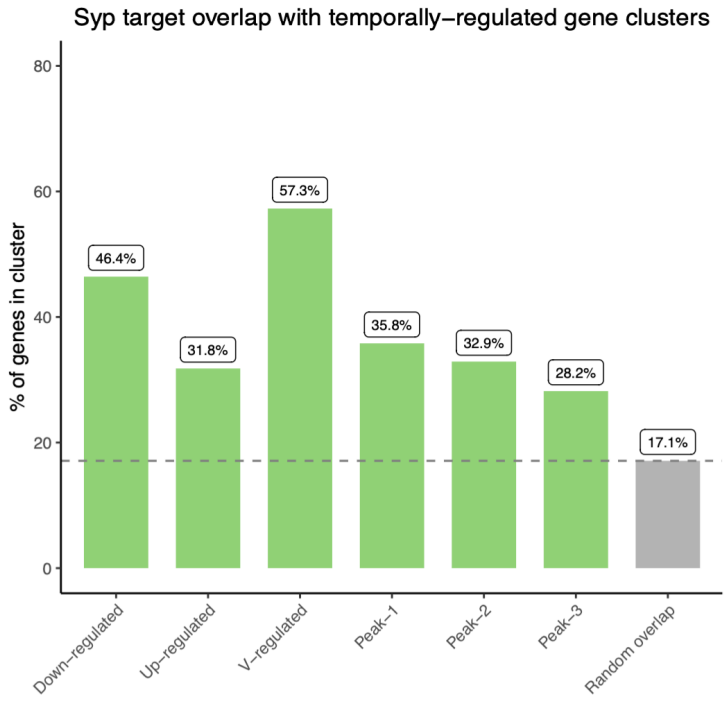


Figure 3.15: Syp binds to temporally regulated transcripts in the NB. Overlap between Syp targets and the clusters of temporally regulated transcripts in NBs (see Figure 3.14). The dotted horizontal line shows the level of overlap that could be expected by chance. The text label represents the percentage of Syp targets within each cluster.

3.3.2 Syp targets in the whole-larva

Functional annotations

At the organism-wide level, a different feature of the Syp targets was highlighted (Figure 3.16). GO analysis revealed enrichment of terms regarding 'actin filament-based processes' in addition to 'axon guidance'. 'RNA-binding' terms were still enriched, but the preeminent molecular function terms involved 'cytoskeleton binding'. Similarly, Interpro protein domain analysis also revealed primary enrichment of the Calponin homology domain, which is present in many cytoskeleton remodellers (Korenbaum & Rivero, 2002). Interestingly, in the Davis lab, the role of Syp in modulating NMJ synaptic transmission has been described in detail (Halstead et al., 2014; J. Titlow et al., 2020). Specifically, Syp has been shown to regulate the translation of *msp300*, which organises actin-filaments around new synapses in response to an activity-dependent stimulus (J. Titlow et al., 2020). Therefore, this result points to a potentially broader role of Syp in proper functioning of motoneurons and muscles via regulation of cytoskeletal binding proteins.

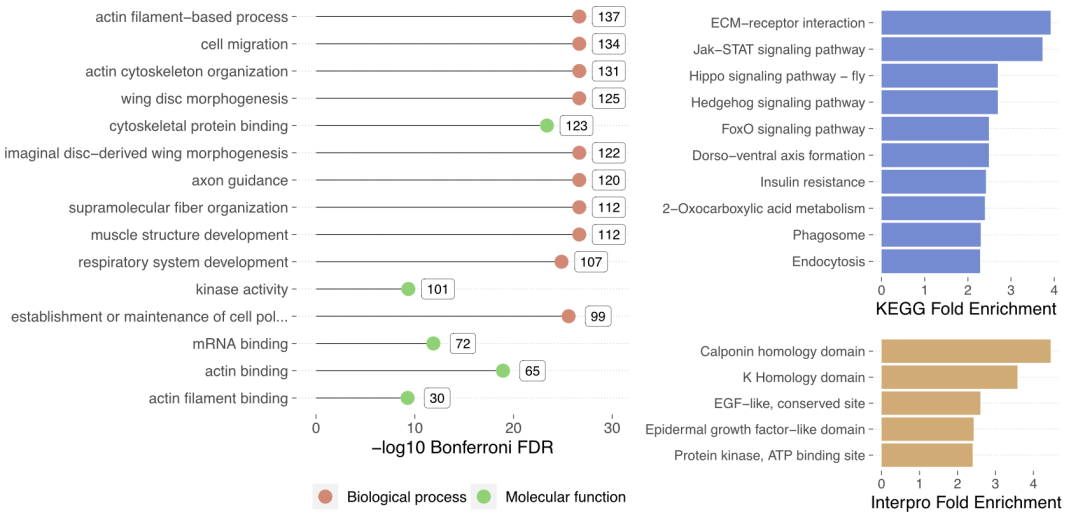


Figure 3.16: Functional annotation of Syp targets in the whole-larva. Gene ontology (GO), KEGG pathway and Interpro domain enrichment analysis of Syp targets in the whole-larva. All *D. melanogaster* genes were used as a background for the analysis. Each term is colour-coded based on the ontology type. GO analysed with the R package 'topGO'. KEGG and Interpro analysis was performed via the 'DAVID' server.

Syp mutant larvae show a locomotion defect

Given the overrepresentation of 'cytoskeleton binding'-related targets, I was interested to investigate whether loss of *syp* leads to defective larval locomotion, which can indicate underlying dysregulation of NMJ synaptic connections. To test this, I assessed the free crawling behaviour of wild-type (*OrR*) and *syp* mutant (*sypN*, also known as syp^{e00286}) L3 larvae using a locomotion assay. The syp^{e00286} is a *P* element insertion allele that leads to degradation of *syp* transcripts via nonsense-mediated decay (McDermott et al., 2014). Larvae were allowed to acclimatise to the stage and ambient temperature (Figure 3.17A) and their free movement was tracked for 3 minutes (Figure 3.17B). I found that, compared to *OrR*, *sypN* larvae crawl much more slowly and in less straight paths (Figure 3.17C). The abnormal activity was further analysed using the R package 'trajr', which revealed the *sypN* larvae display increased irregularities in their movement, as well as more frequent head-turns (Kitamura & Imafuku, 2015). This result indicates that Syp can regulate NMJ functions via regulating its target transcripts. However, it remains to be addressed whether this phenotype is due to dysfunctional neuronal excitability or a developmental defect.

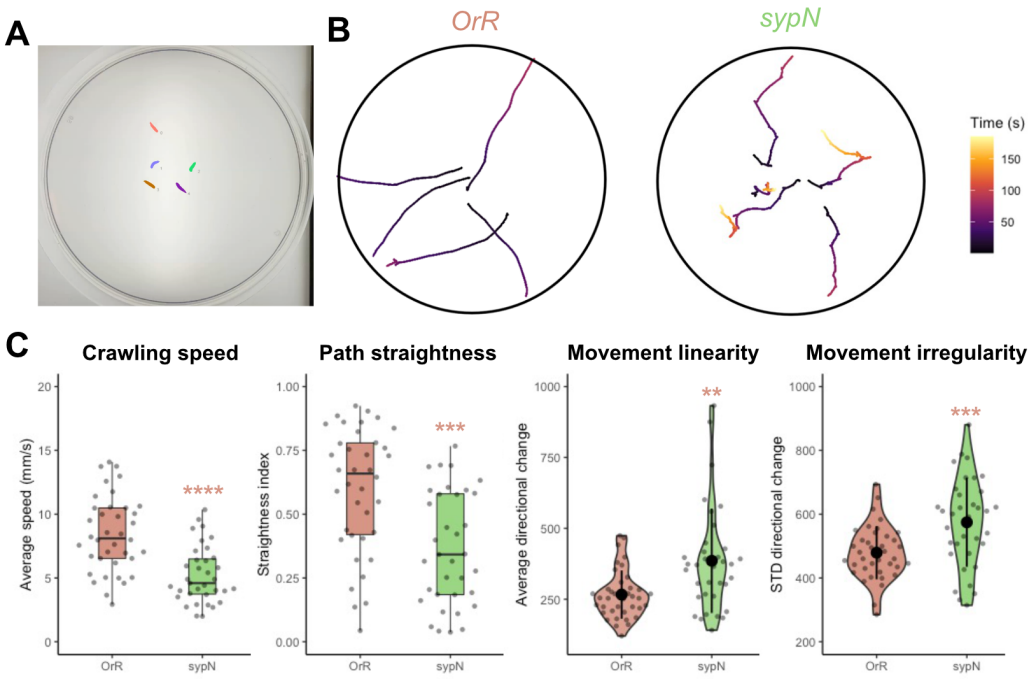


Figure 3.17: Alterations in locomotor behaviour in *syp* mutant larvae. (**A**) The stage set up used to track free-crawling larvae. (**B**) Representative locomotion tracks of free-crawling L3 larvae. Tracks are colour-coded by time. (**C**) Quantification of different parameters describing larval free-crawling behaviours. Movement linearity and irregularities were calculated using the R package ''trajr'. (n=3). **p<0.01; ***p<0.001; ****p<0.0001 (see the Materials and methods Chapter for the details on statistics).

3.4 Post-transcriptional regulation of Syp-bound transcripts

How does gene expression of the targets change when Syp is depleted? Targets of Syp, identified by iCLIP, might be regulated through their mRNA stability, localisation or translation. However, RBP binding events do not necessarily lead to a regulatory outcome, and because Syp binds to more than 1,000 targets in the brain it can be challenging to identify functionally relevant transcripts. Therefore, I aimed to narrow down my exploration to high Syp targets with known nervous system functions. In the next section, I present two functional examples of Syp-mediated post-transcriptional regulation.

3.4.1 RNA stability regulates cell-type specific expression pattern of Prospero

Prospero (Pros) is a TF that controls the balance between NB division and progeny differentiation (Knoblich, 2008). Pros protein is produced in the NB and is asymmetrically segregated into

the GMCs, where it acts to activate pro-neuronal genes for differentiation and also to repress self-renewal genes (Colonques et al., 2011). In the larval brain, *pros* knockdown clones develop into tumours of unregulated cell division (Betschinger et al., 2006). While the expression of *pros* is switched off in the embryonic neuronal differentiation, *pros* is sharply upregulated in differentiating larval neurons, both at the level of mRNA and protein. The role of Pros in NB termination in the pupa has been well-characterised (Maurange et al., 2008), where the nuclear entry of Pros initiates the decommissioning pathways. However, the regulation of *pros* itself and its precise role in larval neurons are unclear.

pros mRNA stability regulation by Syp has been previously shown by Dr Lu Yang, Dr Tamsin Samuels and myself in the Davis lab (Samuels, Arava, et al., 2020). Using single-molecule fluorescence *in situ* hybridisation (smFISH) and immunofluorescence, we characterised the levels of *pros* transcription, cytoplasmic mRNA accumulation and protein abundance in different cells of the NB lineage. Interestingly, we found that *pros* shows a distinct pattern of discordance between mRNA and protein expression patterns. While *pros* is transcribed in all cells of the lineage, a large accumulation of mRNA was limited to mature progeny cells (Figure 3.18A). Then, it has been shown that Syp specifically stabilises an isoform of *pros* with a highly extended 15.5 kb 3'UTR ($pros^{long}$), which is only transcribed in these progeny cells. The loss of Syp leads to destabilisation of $pros^{long}$ and the lack of a Pros protein burst in these cells, which has been confirmed using smFISH, northern blot and RNA-sequencing (Samuels, Arava, et al., 2020). The extra regulatory sequences in the 3'UTR may lead to the inclusion of novel binding sites for the RBPs, so that each UTR isoform is regulated differently. Indeed, Syp iCLIP showed heavy binding to the extended 3'UTR region of $pros^{long}$(Figure 3.18B), suggesting the possibility of a direct stability regulation. Therefore, I asked what is the functional role of Syp-mediated stabilisation of $pros^{long}$ in mature neurons.

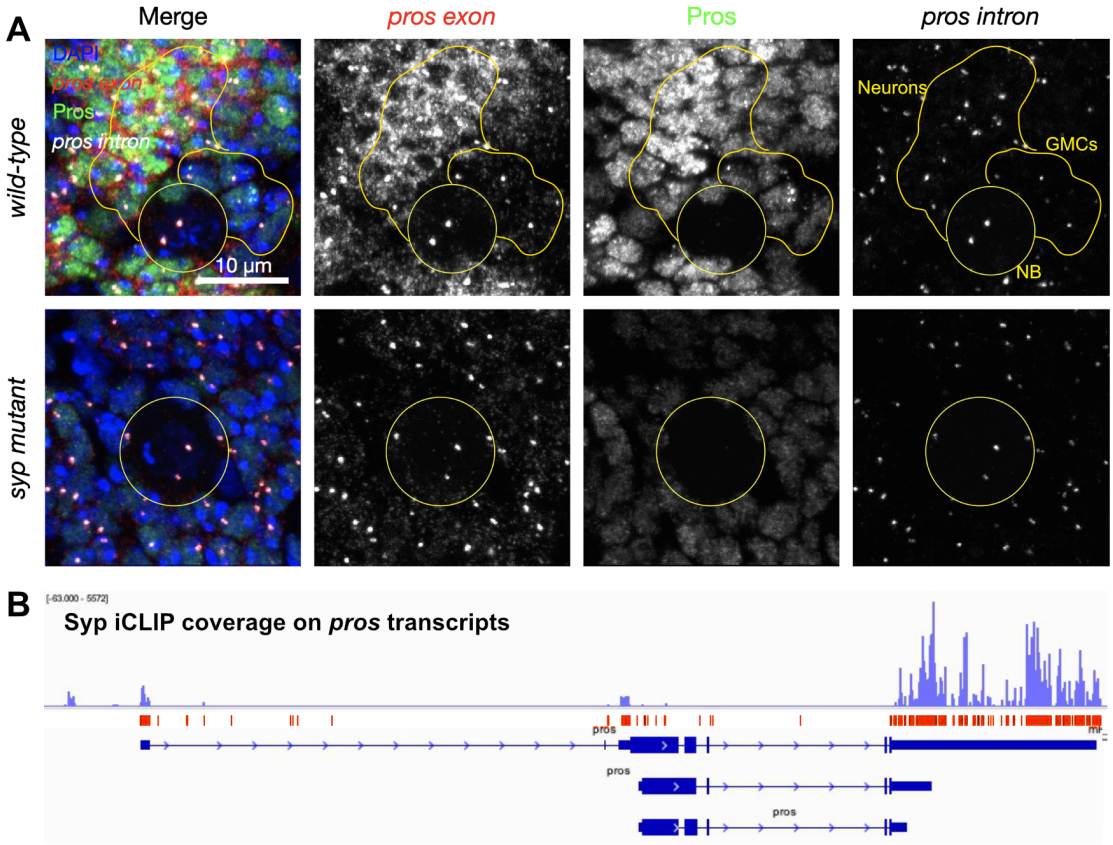


Figure 3.18: Syp regulates *pros* RNA stability. (**A**) smFISH and immunofluorescence staining in *syp* mutant (syp^{e00286}) L3 brain to show down-regulation of *pros* mRNA and Pros protein in mature progeny cells. *pros* transcription is unaltered in *syp* mutant brain. Representative confocal (2 µm projection) images are shown of central brain type I NB lineage. Note the discordant profile of *pros* transcription and Pros protein accumulation. (n=5) (**B**) Syp iCLIP coverage of *pros* transcripts showing predominant binding to the extended 3'UTR regions. iCLIP crosslink densities are shown in blue and binding sites are shown in red. Read per million (RPM) ranges are shown in the upper left corner. The average RPM of the three replicate IP libraries are shown.

CRISPR mutation of $pros^{long}$ leads to dysregulation of Pros protein expression (with Dr Tamsin Samuels)

To provide insight into $pros^{long}$ functions, Dr Tamsin Samuels generated a CRISPR mutant to remove $pros^{long}$ by inserting a viral SV40 terminator sequence just downstream of the proximal polyadenylation signal (PAS) (Figure 3.19A). The homozygous mutant showed a successful reduction of $pros^{long}$ and subsequent loss of Pros protein up-regulation in the mature progeny (Figure 3.19B). The dsRED marker that was used to screen CRISPR-transformed flies is still inserted in the reverse orientation, so we call this line *prosREDr*. The homozygous *prosREDr* flies developed normally with no obvious defects in larval development. Therefore, I used this

mutant line to explore the functional significance of neuronal upregulation of Pros.

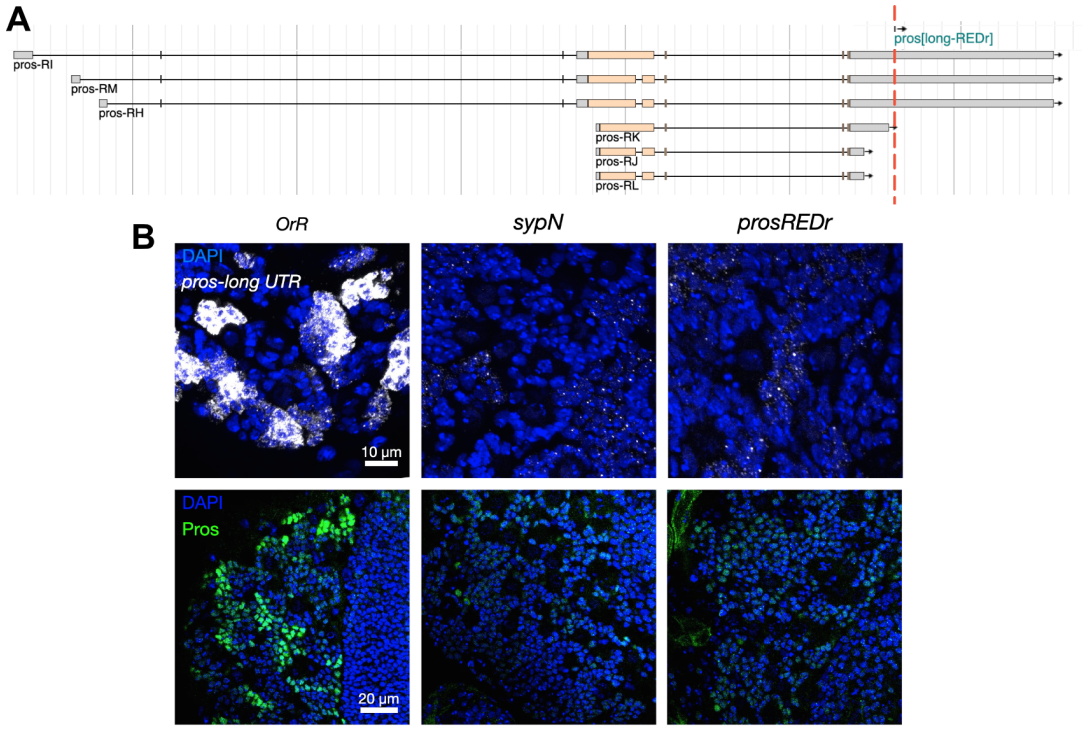


Figure 3.19: *prosREDr* is a *pros^{long}* mutant. (**A**) Genomic location of SV40-dsRED insertion used to generate the *prosREDr* mutant. CRISPR insertion was performed by Dr Tamsin Samuels. (**B**) smFISH against *pros^{long}* and immunofluorescence of Pros in *prosREDr* and *syp* mutants showing down-regulation of *pros^{long}* and Pros protein in both conditions. Representative confocal (5 µm projection) images are shown of the central brain region. (n=3).

prosREDr flies show normal mushroom body differentiation

To test the function of *pros^{long}*, I first examined if it is required for neuronal fate specification. Syp controls neuronal fate during MB development, and loss of Syp results in a complete loss of α/β neurons and precocious production of γ neurons (Z. Liu et al., 2015). Interestingly, a recent single-cell transcriptomic analysis of the adult fly brain revealed that α/β neurons express a much higher level of *pros* compared to other subtypes (Figure 3.20A) (Davie et al., 2018). To test whether loss of *pros^{long}* affects the specification of α/β neurons, I examined the MB morphology in 3-day old *prosREDr* adult brains as well as the brains where *syp-RNAi* was driven by the *OK107* driver (drives GAL4 in mushroom bodies under *eyeless* enhancer (Z. Liu et al., 2015). Immunostaining against FasII, which labels γ and α/β neurons, showed that the morphology of MB neurons in *prosREDr* adults is comparable to the *wild-type* control (Figure 3.20B). This

result indicates that α/β fate patterning occurs normally in absence of *pros*^{long} despite the high mRNA expression of *pros* in these neurons.

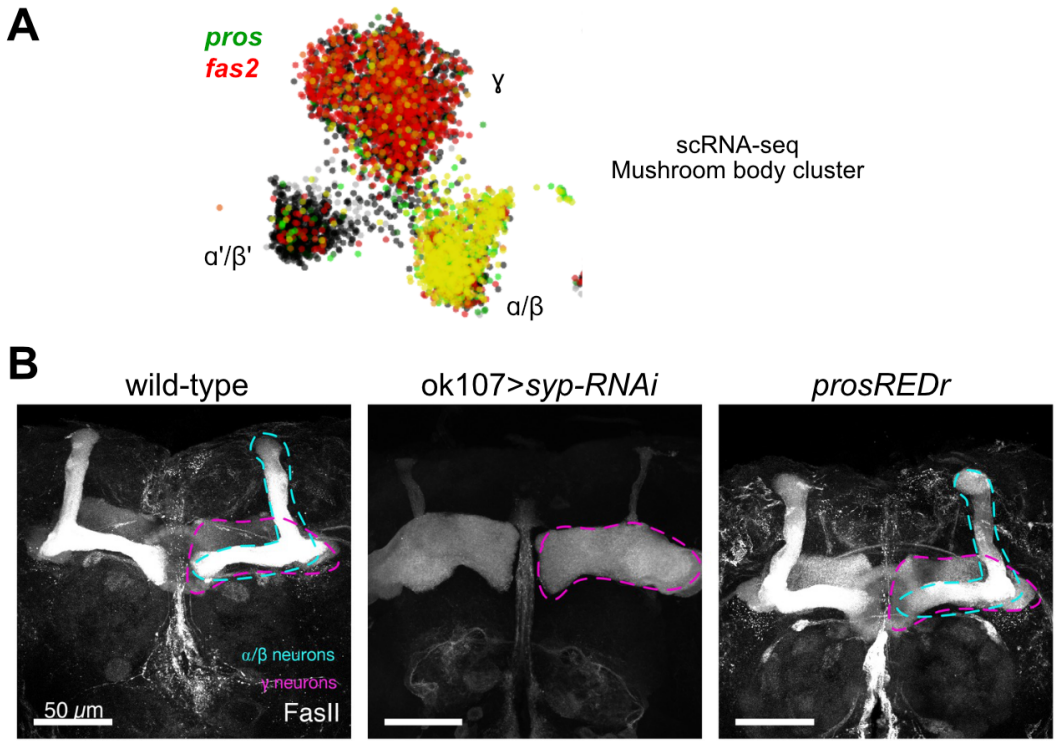


Figure 3.20: *pros*^{long} is not required for mushroom body neuronal fate determination. (**A**) single-cell RNAseq (scRNA-seq) cluster of mushroom body Kenyon cells. Each symbol represents a cell and is colour-coded based on the expression level of *pros* and *fasciculin2* (*fas2*). scRNA-seq data accessed via Scope web interface (Stein Aerts lab). Kenyon cells were identified by *eyeless* mRNA expression. (**B**) Immunofluorescence stain of γ and α/β mushroom body neurons using anti-Fas2 in adult brain of wild-type (*OrR*), *prosREDr*, or *syp-RNAi* driven with *OK107* driver (mushroom body expression). Morphologies of the neuronal subtypes are highlighted. Representative confocal (40 µm projection) images are shown. (n=3).

prosREDr flies show defective adult phenotypes

Next, I was interested to examine whether *prosREDr* flies show abnormal phenotypes in adults. I found that the majority of *prosREDr* adult flies eclose normally and are morphologically comparable to *wild-type*. However, *prosREDr* adults showed highly penetrant abnormal behaviour consistent with impaired locomotor or neurological activity (Figure 3.21A). The observed phenotype may be due to (i) loss of *pros*^{long} activity in adults or (ii) differentiation or development defect following the loss of *pros*^{long}. To examine whether a neuronal loss is associated with lack of *pros*^{long} during development, I stained 1-day old *prosREDr* adult brains with anti-Pros.

Immunofluorescence revealed a reduced frequency of Pros^{+} nuclei in the adult brain, consistent with loss of neurons (Figure 3.21B). These results indicate that loss of *pros^{long}* can lead to abnormal neuronal differentiation and subsequent adult locomotion defect. However, I cannot rule out the possible role of *pros^{long}* in adult neuron homeostasis, which will require temporal knockdown of *pros^{long}* using specific siRNA against the extended 3'UTR region.

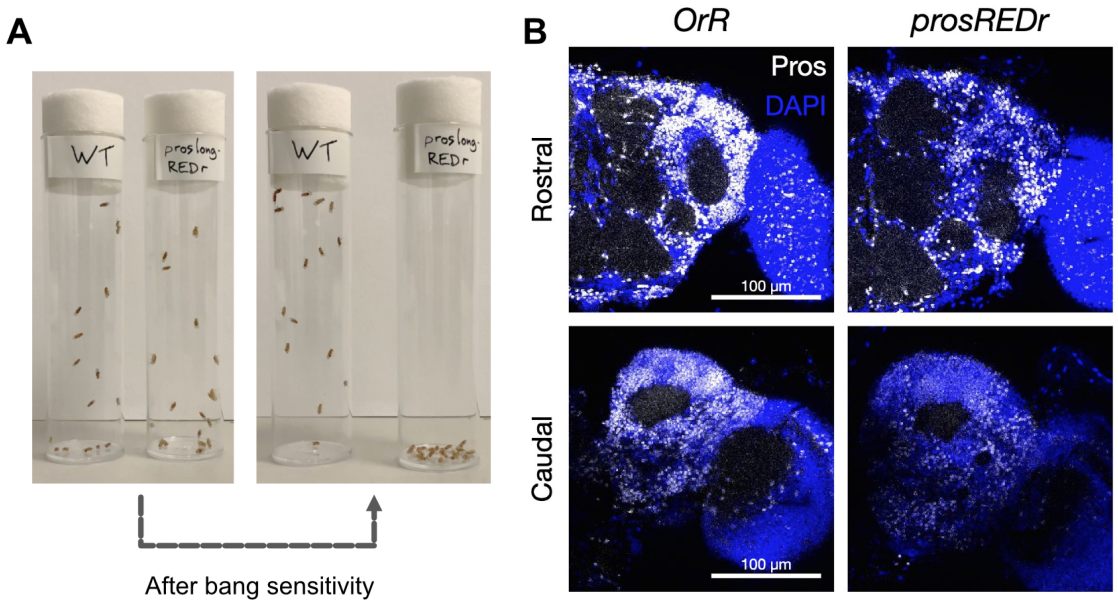


Figure 3.21: *prosREDr* adult flies exhibit defects in their activity. (**A**) 3-day old *prosREDr* homozygous adult flies were allowed to acclimatise for two hours without disturbance. The flies were bang sensitised by forcefully tapping the vials on the surface several times and the flies were allowed to freely move. The video footage can be viewed here: `https://bit.ly/3vDpub8`. (n=3). (**B**) Pros Immunofluorescence in *prosREDr* adults (1-day old) showing a reduced frequency of Pros-positive neurons. Representative confocal (10 µm projection) images are shown. (n=3).

Mature progeny of *prosREDr* shows defective heterochromatin organisation

How can loss of *pros^{long}* lead to dysfunctional neuron differentiation? Interestingly, a recent study described a novel role of Pros in phase separation in partner with heterochromatin protein, HP1a (X. Liu et al., 2020). The authors showed Pros recruits HP1a on Pros target loci to epigenetically reorganise target genes into heterochromatin. The permanent silencing of self-renewal genes (e.g. *dpn*, *stg*, *E(spl)* loci) by this mechanism has been demonstrated in the mature progeny of the NB lineage using DNA-FISH. Furthermore, loss of *pros* has been demonstrated to cause a reduced level of the H3K9me3 modification in neurons (X. Liu et al., 2020). Similarly,

chromatin accessibility is known to decrease as the GMC differentiates (Aughey et al., 2018), and the HP1a-mediated silencing of self-renewal genes occurs in mature progeny cells (Marshall & Brand, 2017). Interestingly, this silencing activity of Pros, described in Liu et al, 2020 X. Liu et al. (2020), spatially coincided very well with the upregulation of Pros and $pros^{long}$. Therefore, I decided to characterise H3K9me3 epigenetic modification levels in larval *prosREDr* brains.

Initially, I examined the level of H3K9me3 in *prosREDr* from the whole-brain lysates. Using western blots, I found that the levels of H3K9me3 protein were comparable between *OrR*, *sypN* and *prosREDr*, indicating that the repressive histone modification levels were unchanged at whole tissue level (Figure 3.22). I reasoned that the effect of dysfunctional regulation might be diluted in the whole brain sample because Pros and Syp are not expressed in all cells. Therefore, I next quantified the H3K9me3 levels on the surface of the central brain using immunofluorescence as shown in Liu et al, 2020 X. Liu et al. (2020). There are no known markers to label the progeny cells that upregulate Pros, especially in *prosREDr*; therefore I employed a strategy where I quantified every cell within the 20 µm depth from the surface of the brain. Nuclei were 3D segmented using Cellpose (Stringer et al., 2021) and were filtered based on their ELAV signal to only quantify H3K9me3 fluorescence in neurons (Figure 3.23). The analysis revealed that neurons of *prosREDr* flies display reduced levels of H3K9me3, suggesting that the phenotype observed by Liu et al, 2020 X. Liu et al. (2020) may be attributable to $pros^{long}$ (Figure 3.24). Interestingly, the reduced H3K9me3 modification was phenocopied in *sypN*, suggesting Syp could be upstream of the epigenetic modification through Pros. In summary, these results suggest an interesting regulatory network where Syp stabilises pro^{long} in order to allow a burst of Pros in mature progeny and then permanent differentiation of these cells.

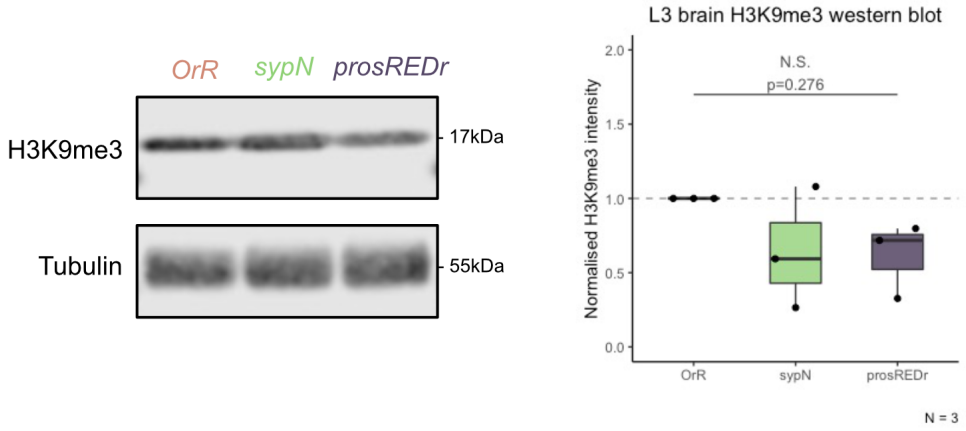


Figure 3.22: H3K9me3 modification is unaltered in the whole-brain level of prosREDr. (**A**) Western blot of H3K9me3 from the dissected brains of *OrR, sypN* and *prosREDr*. Tubulin was used as a loading control. (**B**) Quantification of western blot optical density. H3K9me3 signals were normalised to Tubulin loading control. (n=3).

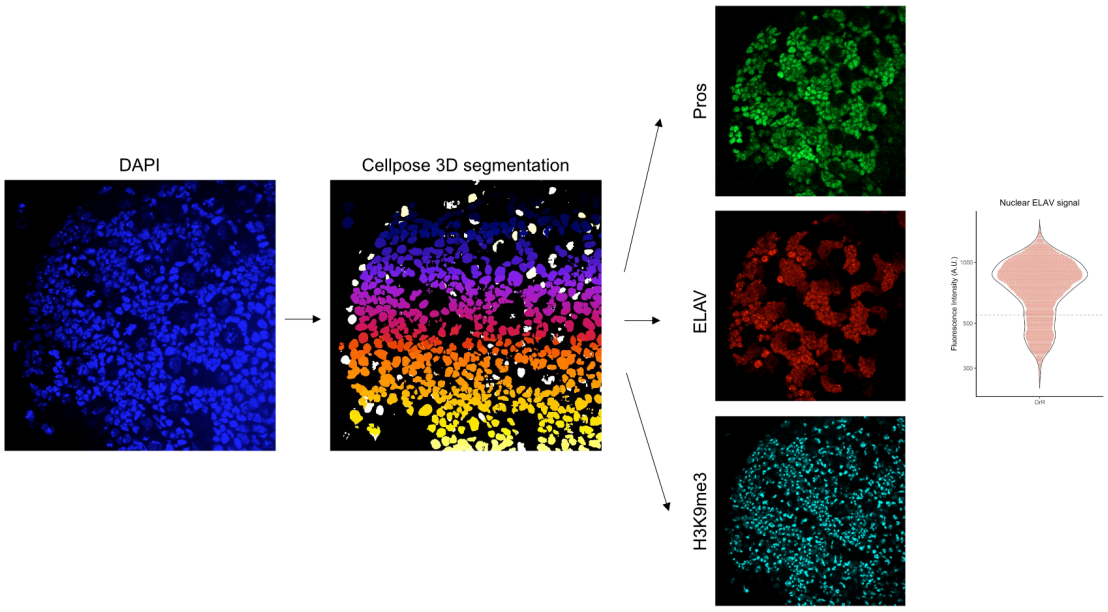


Figure 3.23: Overview of H3K9me3 immunofluorescence quantification strategy. Brain nuclei were segmented using Cellpose 3D mode on DAPI channels. 3D masks were used to quantify fluorescence density from anti-Pros, anti-ELAV and anti-H3K9me3 channels. ELAV fluorescence intensity was used to define the threshold of ELAV-positive neurons. Representative immunofluorescence channels are shown of the wild-type L3 brain. (n=3).

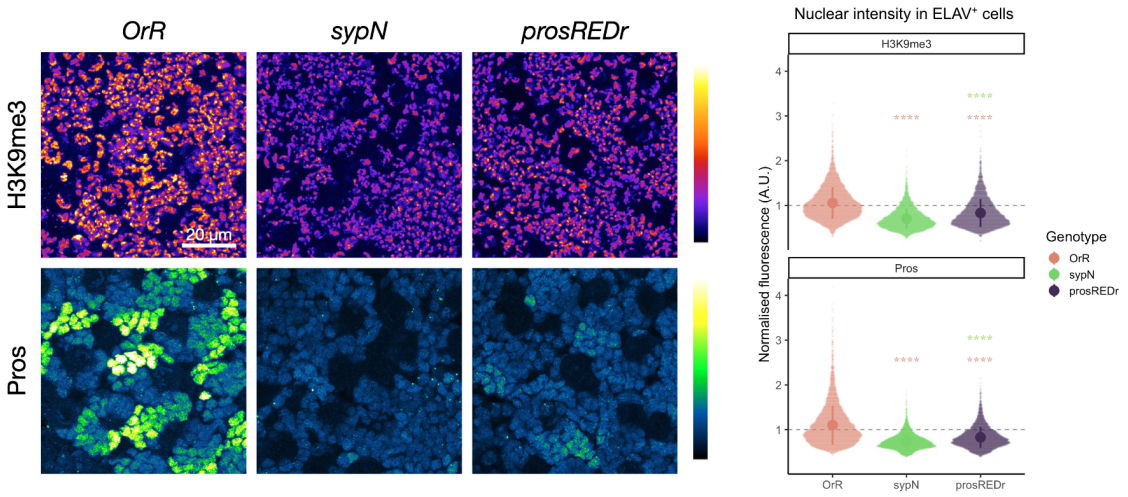


Figure 3.24: H3K9me3 modifications are suppressed in prosREDr mutant neurons. False-colour confocal (4 µm projection) images of anti-H3K9me3 and anti-Pros immunofluorescence channels. Calibration bars are shown on the right of the image panels. Quantification of H3K9me3 and Pros fluorescence density per ELAV-positive nuclei. Fluorescence intensity was normalised to the median of *OrR*. Each symbol represents a nucleus. Data presented as median±SEM. (n=3). ****p<0.0001.

3.4.2 Syp promotes translation of *e93* for the correct progression of temporal fate patterning

Next, I studied the regulation of another TF by Syp. E93, also known as Eip93F, is a member of the pipsqueak TF family first identified as an ecdysone response gene required for autophagy of larval tissues during metamorphosis. More recently, E93 has been described to be expressed in NBs in response to ecdysone as a late-stage temporal identity factor (Syed et al., 2017) (Figure 3.25)). Imp and Syp have been shown to influence the timing of E93 expression, which integrates the intrinsic temporal cues with extrinsic hormonal cues to terminate mushroom body neurogenesis in the pupa (Pahl et al., 2019). However, the precise mechanism of E93 regulation by Syp and the role of E93 in Type I lineages during larval stages are unclear.

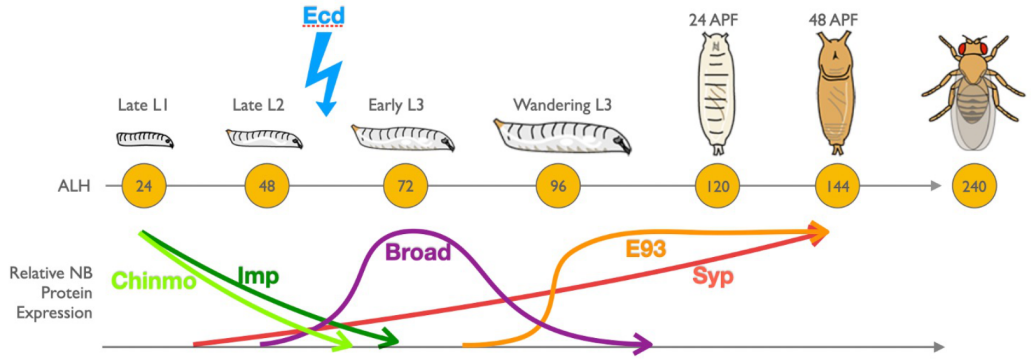


Figure 3.25: Temporal factors during larval NB development. Relative protein expression levels of temporal factors are depicted across larval neurogenesis. The expression level represents levels within NBs. Figure adapted from Syed et al. (2017).

E93 shows discordant protein and RNA expression patterns

Initially, I found *e93* from an experiment to examine the mRNA and protein expression patterns of Syp targets that are also known temporal TFs. Here, I observed an unusual gene expression pattern of *e93*. Similar to the earlier example of *pros*, *e93* showed a highly discordant pattern of protein and mRNA distribution in the NB lineage (Figure 3.26A). *e93* transcripts were expressed evenly in the NB and progeny, however, the protein was enriched in NBs, suggesting post-transcriptional control of *e93* translation in these cells. Interestingly, *e93* was also recovered from the search of Syp targets with long 3'UTR extensions as it featured a 7.7 kb extension between the short and the long 3'UTR isoforms. Indeed, Syp iCLIP coverage showed extensive binding to the eUTR region of *e93* (Figure 3.26B). To examine which isoform of *e93* is expressed in the NBs, I designed smFISH probes against the constitutive (*cmm*) and in the extended UTR (*eUTR*) sequences and labelled them with different fluorophores (Figure 3.26B). smFISH revealed that the majority of constitutive probe spots were also dual-labelled with *eUTR* probe (Figure 3.26C), indicating the $e93^{long}$ is the dominant isoform expressed in the NBs.

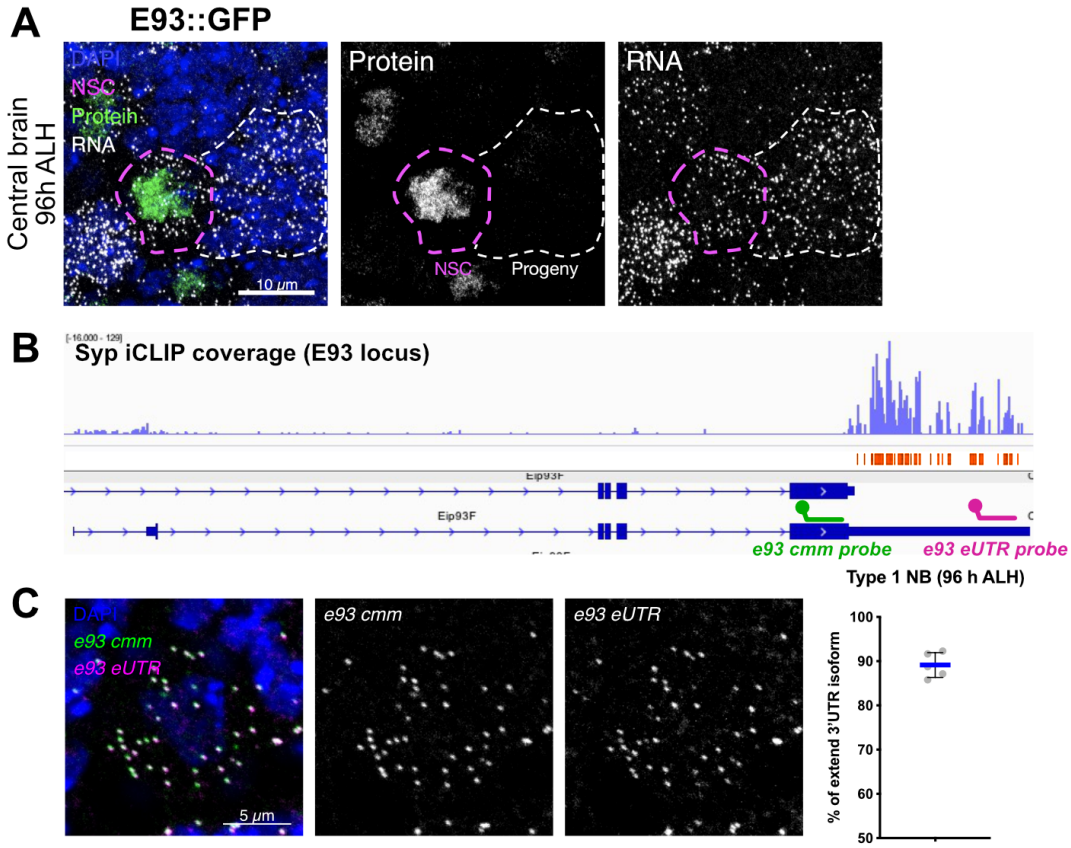


Figure 3.26: Discordant expression pattern of *e93* protein and mRNA. (**A**) Expression of *e93* protein and mRNA as assessed by E93::GFP fluorescence and smFISH against *e93*. A neural stem cell (NSC, NB) and its progeny cells are depicted. 5 µm projection confocal image is shown. (n=2). (**B**) Syp iCLIP coverage of *e93* transcript showing enrichment of binding sites within the extended 3'UTR region. iCLIP crosslink densities are shown in blue and binding sites are shown in red. Read per million (RPM) ranges are shown in the upper left corner. The average RPM of the three replicate IP libraries are shown. (**C**) smFISH analysis of *e93* 3'UTR isoform expression. smFISH probes against the constitutive region of *e93* coding sequence (*cmm*) and the extended 3'UTR region (*eUTR*) were used. Dual-labelling of the spots were assessed by co-localisation within a 250 nm distance between the two probe channels. (n=5).

E93 expression is sensitive to Syp levels

Given the NB function of E93, I was interested to examine the consequences of E93 expression when the Syp levels are modulated. Previously, Syed et al, 2017 (Syed et al., 2017) have described that E93 protein levels are unchanged in *syp* mutant NBs and concluded a lack of regulation. However, the authors used the hypomorphic allele of *syp* (syp^{f03775}), which allows low-level accumulation of Syp in the larval brain (Figure 3.27A). I re-examined the E93 level in the hemizygous null over deficiency mutant (syp^{e00286}/*Df(3R)BSC124*) NBs and found the E93 protein expression is lost in these NBs (Figure 3.27A). Furthermore, smFISH revealed that transcription

of *e93* was completely abolished (lack of nuclear transcription foci) in *syp* loss-of-function conditions (Figure 3.27B). Therefore, I conclude that Syp may regulate *e93* transcription via indirect interactions and that the expression of *e93* is sensitive to Syp levels.

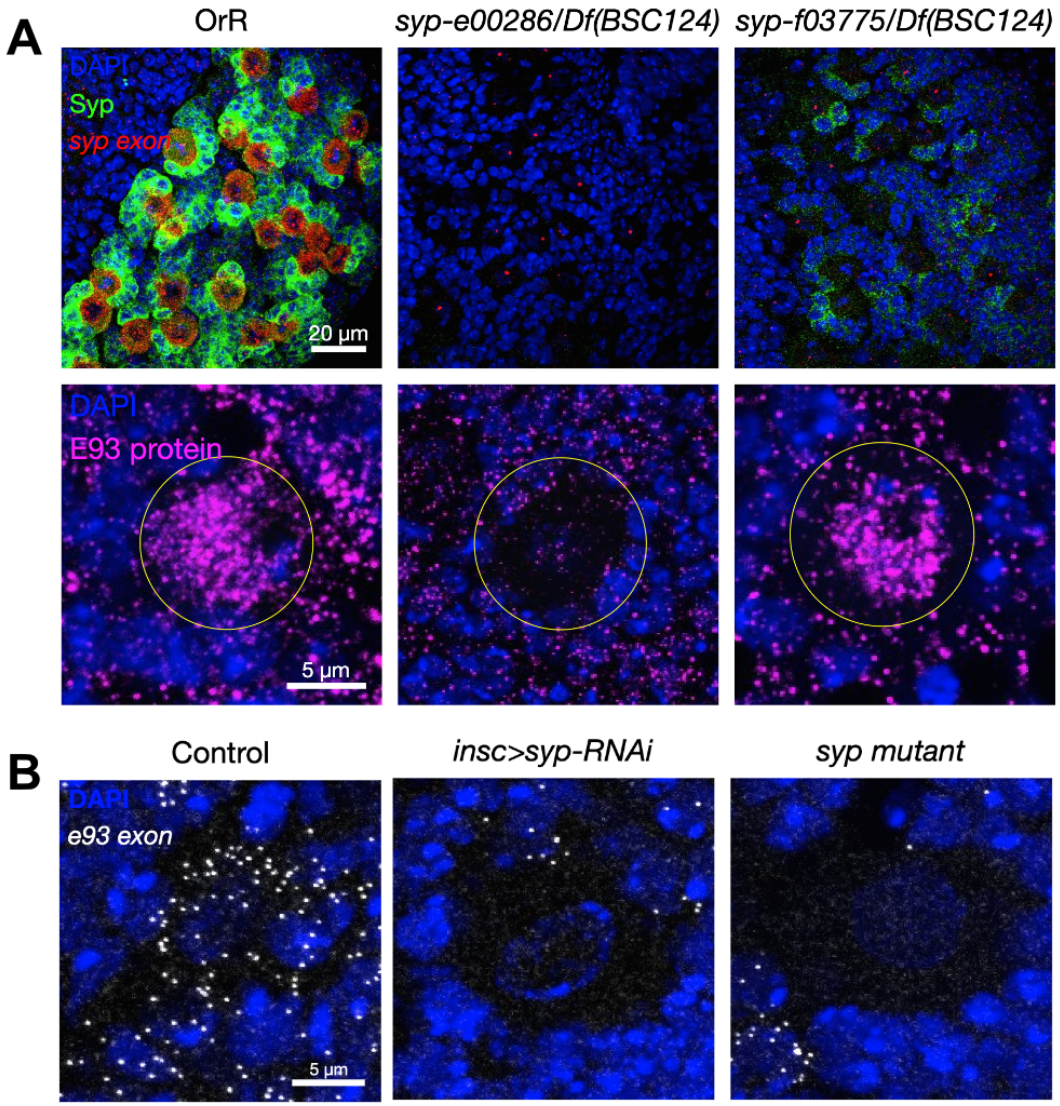


Figure 3.27: *e93* expression is sensitive to Syp levels. (**A**) smFISH and IF to show *syp* expression pattern of the two *syp* alleles. syp^{e00286} is a null allele while syp^{f03775} is a hypomorphic allele. Both alleles were tested hemizygously over the *Df(3R)BSC124 syp* deficiency chromosome. E93 protein expression is absent in the null mutant but is still visible in the hypomorphic mutant. (n=2). (**B**) smFISH analysis of *e93* transcription in *inscuteable(insc)*-driven *syp-RNAi* or the *syp* null mutant NBs. Representative confocal (5 µm projection) images are shown. (n=2).

Syp regulates translation of *e93*

In order to study the role of Syp in the post-transcriptional regulation of *e93*, I decided to deplete Syp temporally by generating *syp* knockdown clones. To set up the initial permissive conditions for *e93* transcription, larvae were allowed to develop normally until 48h ALH. At 48h ALH, I induced *syp-RNAi* clones via heat-shock, which induces expression of flippase that removes a repressor sequence upstream of the GAL4 activator. This induction occurs stochastically, so it generates a small number of NB clones that express *syp-RNAi* from this stage forward (Figure 3.28A). The induced larvae were further incubated at 30°C for 48 hours. smFISH revealed the presence of *e93* transcription in *syp-RNAi* clones (Figure 3.28B). To test if the mRNA stability of *e93* transcripts is altered by the loss of Syp, I quantified the number of *e93* mRNA molecules in the cytoplasm and those within the transcription foci. By making a steady-state assumption, the half-life of the *e93* mRNA could be calculated using these smFISH images (Bahar Halpern & Itzkovitz, 2016). The analysis showed that the half-lives of *e93* transcripts are invariant between *wild-type* and *syp-RNAi* clones, indicating that Syp does not regulate the stability of *e93* mRNA. In contrast, immunofluorescence using anti-E93 revealed that the protein level of E93 was significantly reduced in *syp-RNAi* clones (Figure 3.28C). Taken together, these results demonstrate that Syp regulates the translational efficiency of *e93* but not its mRNA stability.

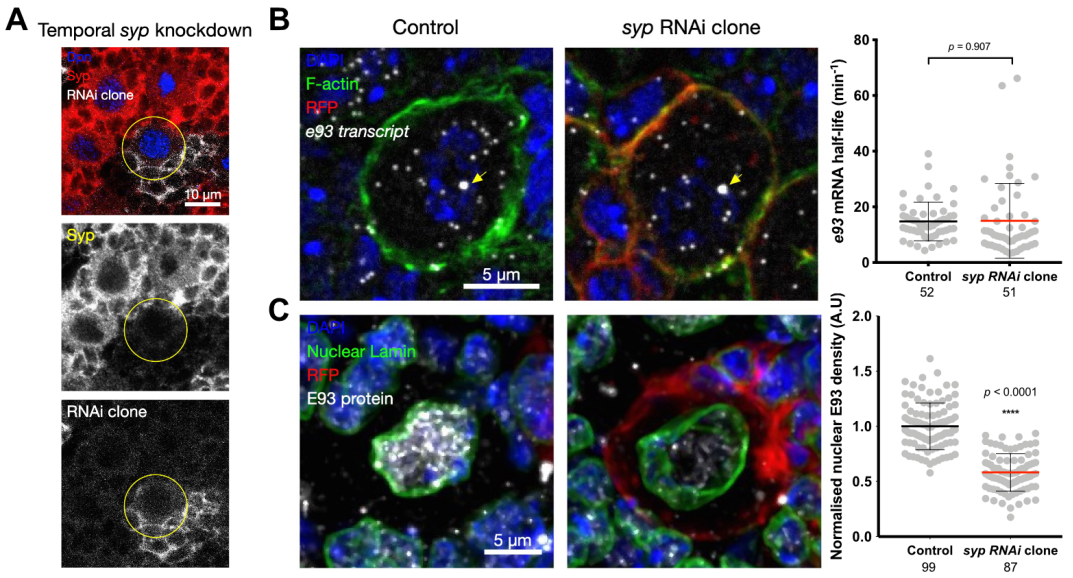


Figure 3.28: Syp regulates translational efficiency of E93. (**A**) Immunofluorescence against Syp showing knockdown of Syp in the positively labelled RNAi clone amongst the surrounding wild-type clones. (n=3). (**B-C**) smFISH and immunofluorescence analysis of *e93* mRNA stability and protein translation in temporal *syp-RNAi* clone NBs. *e93* transcription sites are depicted with the yellow arrows. mRNA half-lives were calculated by comparing nascent and mature mRNA levels under the steady-state assumption. E93 protein levels were quantified by integrating fluorescence intensity of nuclear anti-E93 signals. Representative confocal (5 µm projection) images are shown. (n=3). ****p<0.0001.

E93 is required to close Broad temporal window

What is the molecular function of E93 during larval stages? Previously, E93 has been shown to modify the chromatin accessibility of temporally regulated enhancers in the developing wing disc (Nystrom et al., 2020). Interestingly, one of the repressed enhancers was responsible for the expression of *broad* (*br*), a temporal TF that has also been identified in NBs, which just precedes the E93 window (Doe, 2017a; Syed et al., 2017). Therefore, I tested Br protein expression in *e93-RNAi* NBs at 96h ALH, a stage where Br is not highly expressed (Figure 3.25). Immunofluorescence revealed that E93-depleted NBs show abnormally extended expression of Br at 96h ALH, indicating that the correct closure of the Br temporal window requires E93 (Figure 3.29). Notably, a high nuclear Br level was also phenocopied in temporally-depleted *syp-RNAi* clones. Therefore, I speculate that Syp-mediated upregulation of E93 translation may be required to close the Br expression window. However, *br* is also a Syp iCLIP target, so the extent to which *br* repression is due to direct Syp interaction or indirectly via E93 is yet to be

established. In summary, I have characterised the post-transcriptional regulation of *e93* by Syp and the role of E93 to shut Br temporal window in the NBs.

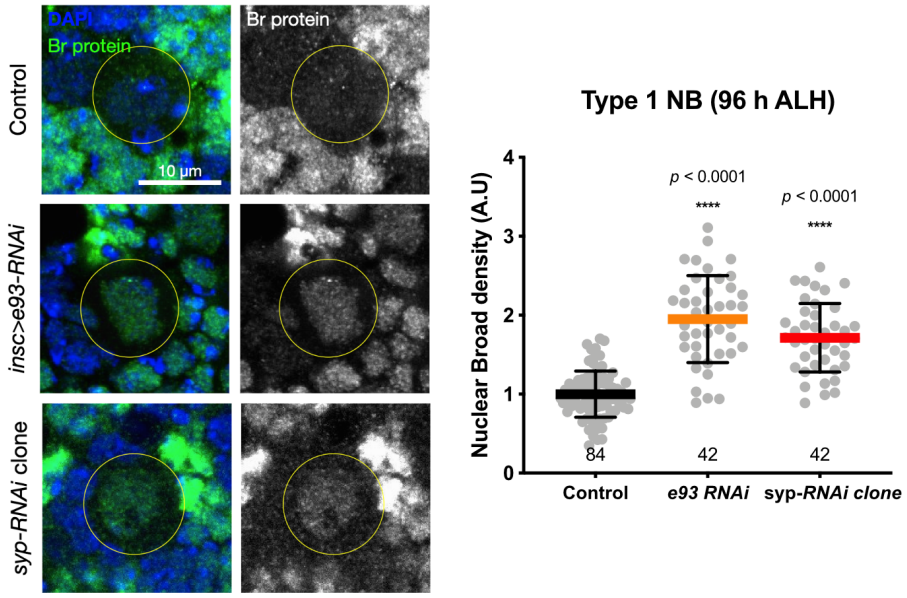


Figure 3.29: Both *syp* and *e93* are required for the correct closure of the Br temporal window. (**A**) Immunofluorescence of Br showing precocious expression of Br beyond its normal expression window in *e93* or *syp* deficient NBs (*left panels*). Representative confocal (5 µm projection) images are shown. Quantification of nuclear Br fluorescence density in the NBs. Fluorescence intensity was normalised to the control average. *Insc-GAL4* flies were used as the control. (n=3). ****p<0.0001.

3.5 Next steps

Do Syp and Imp compete for the same transcript and exert opposite regulatory effects? Does this competition lead to NB age determination? There is a significant degree of overlap between Syp targets and Imp RIP-seq targets in the L3 brain (Samuels, Järvelin, et al., 2020). However, Imp RIP at such a late stage does not recapitulate the capacity of Imp to regulate NB transcriptome (Syed et al., 2017), so the data is biased towards neuronal transcripts over NB and immediate progeny cells. Therefore, a time-resolved target and binding site identification of Syp and Imp will be required to answer these questions. I found that the protein expression of Syp and Imp in the first (L1) and second (L2) instar larval stages were highly enriched in the brain compared to the whole body (Figure 3.30A-B). This makes it possible to use whole L1 larval and L2 head samples for iCLIP experimentation without the need for individual brain dissections. The library

preparation for these samples is currently underway, and I anticipate these experiments will be able to provide further understanding of the molecular regulation of *Drosophila* neural development.

Additionally, I plan to perform computational analyses to identify binding motifs or RNA structures at or near Syp binding sites, harnessing the full nucleotide-resolution of the iCLIP data. Since Syp binds to multiple transcript features (5'UTR, 3'UTR, intron, exon), it would be worthwhile to categorise Syp binding sites and identify enriched sequence motifs separately from each feature (Van Nostrand, Pratt, et al., 2020). Together with datasets showing transcriptome changes in *syp* mutant brains (Samuels, Järvelin, et al., 2020), it would be interesting to explore whether specific binding motifs are associated with Syp-mediated RNA stabilisation or destabilisation.

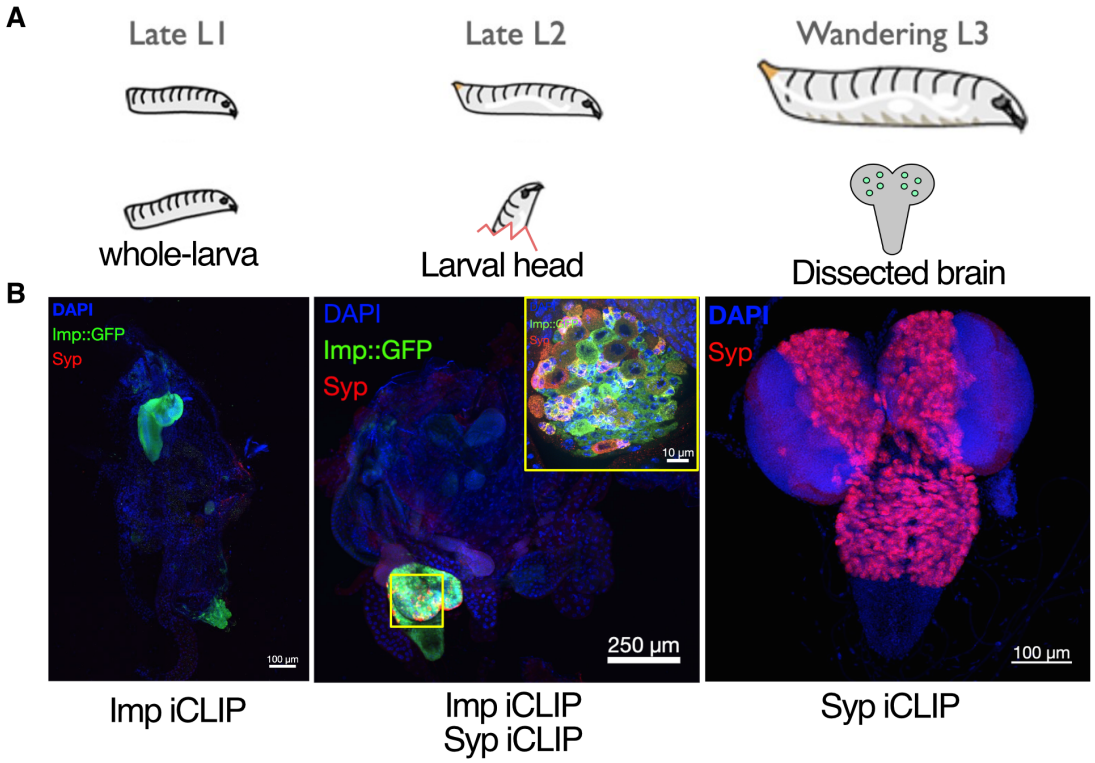


Figure 3.30: Strategy for comparative Syp and Imp iCLIP across development. (**A**) Schematics of biosamples taken at each developmental stage for iCLIP. L1 (24h ALH), L2 (48h ALH) and L3 (96h ALH). (**B**) Immunofluorescence against Syp and Imp in corresponding biosamples, showing the enrichment of both RBPs in the brain. A zoomed inset in the L2 brain shows a heterogeneous population of NBs expressing different levels of Syp and Imp. Images shown are confocal maximum intensity projections. (n=2).

3.6 Conclusion and discussion

The RBPs Imp and Syp have been shown to regulate NB age through larval neurogenesis (Doe, 2017a; Z. Liu et al., 2015). It is thought that the relative levels of Syp and Imp determine the fate of serially derived progeny cells and also regulate the decommissioning of NBs towards the end of neurogenesis (C. P. Yang et al., 2017). RBPs exert their function through post-transcriptional regulation of their target RNAs, which controls various aspects of their metabolism, including mRNA stability and translation (Pilaz & Silver, 2015). Therefore, the identification of *in vivo* RNA targets is key to understanding the regulatory functions of RBPs.

In this Chapter, I used iCLIP with dissected larval brains and whole-larval samples. I performed a computational analysis of the data to identify endogenous targets of Syp in the two biosamples. I found 1,400 and 1,900 Syp targets in the brain and whole-larva, respectively. These targets are potentially post-transcriptionally regulated by Syp and most likely contribute to the myriad of phenotypes observed when Syp is depleted. I found the targets of Syp are enriched for transcripts encoding TFs or RBPs, suggesting that Syp coordinates complex regulatory networks involving both transcriptional and post-transcriptional controls. Interestingly, I found an enrichment of targets with extended 3'UTR isoforms and evidence of Syp binding to these extended sequences. This might indicate the preference of Syp to bind to these extended 3'UTRs or just the increased likelihood of Syp binding site occurrences associated with longer sequences. Comparing iCLIP or RIP-seq of other RBPs with preferential binding to UTRs may be able to highlight if this is specific to Syp or is a general feature of RBPs. Nevertheless, the targets of Syp overlapped significantly with neural-specific 3'UTR extended transcripts, potentially highlighting the functional role of extra regulatory sequences, and this would be a useful direction for future research.

Neuronal extension of 3'UTRs is tightly regulated by multiple redundant mechanisms (Carrasco et al., 2020; Wei et al., 2020). The brain tissue-specific expression of ELAV ensures co-transcriptional masking of proximal polyA site (PAS), leading to distal PAS usage (Bae & Miura, 2020; Hilgers et al., 2012). The distal PAS usage is also conserved in primary cultures of

mammalian neurons, which requires the mammalian homolog of ELAV, Hu protein (Lee, Wei, et al., 2021). Therefore, the general extension of 3'UTRs is a conserved mechanism important for maintaining neural-specific transcriptomic landscapes (Mayr, 2016). The alternative 3'UTR isoforms can give rise to alterations in mRNA stability, mRNA localisation to subcellular regions, and differential translational control (Bae & Miura, 2020). A study using RNA-seq of NBs and progeny found that the proportion of reads mapping to the extended 3'UTRs is higher in progeny cells compared to NBs (Berger et al., 2012), which coincides with increasing ELAV protein expression along the neuronal differentiation axis (Homem & Knoblich, 2012). Therefore, it is possible that some transcripts employ 3'UTR length control to evade or accept regulation via Syp, which is ubiquitously expressed across NB lineage. It would be interesting to cross-reference single-cell 3' end RNA-sequencing data with the current Syp iCLIP to identify which transcripts are bound by Syp and also change their 3'UTR lengths along the NB, GMC and neuron differentiation route. Alternatively, long 3'UTR sequences could act as a scaffold to promote protein-protein interaction of Syp to its partner effector molecules or simply as a sponge reservoir (Bae & Miura, 2020; Mayr, 2017). An example of riboregulation, where RNA regulates RBP localisation or function, has been shown with long non-coding RNA (lncRNA) *cherub* on Syp. *cherub* controls Syp localisation to GMCs during asymmetric NB division, and it has been shown to be required for *brat*-mediated tumourigenesis (Landskron et al., 2018). The developmental function of *cherub* and other ncRNA targets of Syp found in my iCLIP is unclear. However, mammalian SYNCRIP also binds multiple lncRNAs, and one of the well-characterised targets is *BC200/BC1*, associated with neurodegeneration and several types of cancer (Duning et al., 2008; M. M. K. Hansen et al., 2018; Samson et al., 2018). It would be interesting to investigate how lncRNA and Syp regulate each other in different biological contexts.

I note that the current analysis does not examine cell-type specific targets of Syp. Given the low amount of starting materials recoverable from individual cells, performing iCLIP on FACS-sorted cells would be extremely time consuming and may yield sub-optimal data. On the other hand, iCLIP of the whole brain presents the overall snapshot of *in vivo* binding across the entire tissue, which should be useful to enhance annotation of future cell-type specific RIP-seq

experiments. It is interesting to consider for follow up work beyond this thesis, the HyperTRIBE technique (discussed in Chapter 1). The HyperTRIBE is an interesting technique that utilises UAS-GAL4 driven RBPs that are tagged with Adenosine Deaminase Acting on RNA enzyme (ADAR). This fusion construct therefore, leaves A-to-I editing events nearby RBP binding sites (Nguyen et al., 2020). Together with tissue-specific GAL4 drivers, HyperTRIBE could potentially provide cell-type specific Syp targets without the tedious cell sorting. Although the resolution of binding sites is compromised with HyperTRIBE it could easily be cross-referenced to the current iCLIP data for 'conversion' into high-resolution binding maps.

Syp also targeted an additional class of transcripts in the whole-larval samples, namely transcripts encoding cytoskeleton binders. Actin and microtubule-binding proteins play an essential role in neuronal RNA transport, synaptic plasticity and axon guidance. I found that *syp* mutant displays abnormal locomotor behaviour suggesting dysfunctional NMJ activities. It would be interesting to follow up whether a set of actin-binding proteins are regulated by Syp, or perhaps whether post-transcriptional regulation and local translation of transcripts encoding the cytoskeletal binding protein is a general phenomenon.

I presented two examples of iCLIP targets that are regulated by Syp via post-transcriptional control. The isoform-specific stabilisation of $pros^{long}$ is an interesting example, illustrating the regulatory potential of the extended 3'UTR sequences whose deletion leads to incorrect H3K9me3 modification patterns and abnormal adult behaviour. A translational control was exemplified with *e93*, which showed the conserved function in the brain and wing disc to silence the preceding temporal factor, *br*. These gene-by-gene approaches using knock-downs and imaging methods provide a detailed understanding of gene expression control with high spatiotemporal specificity. However, they are unfortunately low throughput. Transcriptome-wide identification of perturbations on RNA stability and translational control using 4sU metabolic labeling and ribosome profiling could be an attractive next step (Holt & Schuman, 2013; Jürges et al., 2018; Reynoso et al., 2015).

Interestingly, I note that both examples of *pros* and *e93* were initially identified by comparing the expression patterns of protein and RNA. A discordant pattern of protein and RNA distribution is a major hallmark of post-transcriptional control (Buccitelli & Selbach, 2020; Mayr, 2017; Titlow et al., 2022). Therefore, systematic profiling of such uncorrelated patterns can be a feasible alternative to assess the prevalence and importance of post-transcriptional regulation in the nervous system development. In the next Chapter, I describe this in more detail.

4

Investigating the prevalence of post-transcriptional regulation in the developing *Drosophila* nervous system

4.1 Introduction

Measuring mRNA and protein levels is an integral part of our effort to understand the flow of information from genomes to phenotypes. The hierarchical structure of the Central Dogma, proposed by Francis Crick, leads to the implicit assumption that protein abundance correlates with the abundance of each mRNA, which encodes the information to make each specific protein. Although largely correct, many confounders limit our ability to extrapolate from the mRNA distribution and abundance to the level and location of its corresponding protein (Buccitelli & Selbach, 2020). A key example of a confounder is the post-transcriptional regulation of mRNA by the RBPs. A hallmark of post-transcriptional regulation is a discordant distribution (lack of correlation) of individual species of protein and mRNA (Buccitelli & Selbach, 2020; Pilaz & Silver, 2015). Discordance between mRNA and protein levels and distribution often manifests itself in different cell types because of variations in mRNA stability or translation efficiency. However, post-transcriptional regulation can also occur within a cell, leading to distinct localisation of mRNA from its protein product (Holt et al., 2019; Holt & Schuman, 2013). Such intracellular

discordance can be due to various mechanisms, including mRNA degradation, local translation or directed transport of protein and mRNA.

Neurons are the most polarised cell type in multicellular organisms and it has been generally accepted that post-transcriptional mechanisms are critical to regulate protein levels at the synapses located far away from the cell body (Holt & Schuman, 2013). Localised translation is also an attractive mechanism to explain how multiple different distal synapses can be regulated differently and independently within the same neuron. The advent of single-cell transcriptomics (scRNA-seq) (Aldridge & Teichmann, 2020) and spatial transcriptomics (Marx, 2021) have transformed the profiling of RNAs in developing and mature nervous systems. However, matching single-cell proteomics techniques are largely in their infancy and often do not offer sufficient dynamic range for expression profiling (Buccitelli & Selbach, 2020; Schoof et al., 2021). Moreover, single-cell approaches often discard the peripheral compartment of cells, which is not compatible with localised RNA and proteins in neuron termini (Davie et al., 2018). Therefore, a systematic characterisation of discordance between protein and mRNA has not been carried out across a whole intact nervous system. Other methods had been more revealing for genome-wide characterisation of localised mRNA, including growing neurons on membranes with pores, allowing the isolation of the extremities for RNA sequencing (Mili et al., 2008).

In the previous Chapter, I have characterised two examples of post-transcriptional gene regulation. Expression of *pros* and *e93* was controlled through mRNA stability and translational control, respectively. Interestingly, both *pros* and *e93* were initially identified from careful observation of their protein and mRNA distributions, which were distinctively uncorrelated. Therefore, in this Chapter, I expand on this methodology to compare the level of discordance between hundreds of mRNAs and their corresponding proteins across a complex tissue in 3D. Together with Dr Joshua Titlow, Dr Ana Palanca and Maria Kiourlappou, I have devised a strategy using an existing library of protein-tagged fly lines (Lowe et al., 2014). We performed systematic visualisation of mRNA using smFISH and co-detection of tagged protein at high resolution in

the same samples. I first present the prototyping of our workflow using the targets of Syp, and then an expansion of this method to 200 genes from a collection of YFP fusions.

4.1.1 Specific aims of this chapter

1. To characterise gene expression patterns of Syp targets in the developing larval brain
2. To systematically analyse protein and mRNA expression across the nervous system

4.1.2 Publication

Contents of this Chapter have been published as a preprint, where I am a co-first author (Titlow et al., 2022):

Titlow, Joshua S., Maria Kiourlappou, Ana Palanca, Jeffrey Y. Lee, Dalia S. Gala, Darragh Ennis, Joyce J. S. Yu, et al. 2022. **"Systematic Analysis of YFP Gene Traps Reveals Common Discordance Between mRNA and Protein Across the Nervous System."** bioRxiv. `https://doi.org/10.1101/2022.03.21.485142`.

Multiple co-authors contributed to this pre-print. While I mainly present my contribution to this project, work done by others is acknowledged in the text. This preprint has an accompanying technological publication, illustrating the usage of the Zegami data exploration tool:

Maria Kiourlappou, Stephen Taylor, and Ilan Davis. 2022. **Zegami user manual for data exploration: "Systematic analysis of YFP gene traps reveals common discordance between mRNA and protein across the nervous system."** Zenodo. DOI:`https://doi.org/10.5281/zenodo.6374012`

4.2 Investigating gene expression pattern of Syp targets in the larval brain

Syp binds to more than 1,000 mRNA targets in the brain. It is not feasible to individually test, using smFISH, whether all these target transcripts are regulated post-transcriptionally. Fortunately, the *D. melanogaster* community offers a wealth of genetic tools, including protein-traps lines (Kanca et al., 2017; Morin et al., 2001). Protein-trap fly lines are produced by inserting a fluorescent protein mRNA sequence, flanked by splice acceptor and donor sites, into the coding intron of target genes. These lines are sometimes produced systematically using transposable elements (Lowe et al., 2014; Morin et al., 2001; Nagarkar-Jaiswal et al., 2015) or CRISPR-based genome editing (Lee et al., 2018a). Depending on the insertion site, many of the available protein trap lines produce functional proteins that can rescue null mutant phenotypes, correctly localise to known subcellular compartments, do not get targeted for ubiquitin-mediated degradation, and are homozygous viable (Lowe et al., 2014; Nagarkar-Jaiswal et al., 2015). Consequently, gene expression can be easily monitored with higher throughput via endogenously-tagged fluorescent protein and by using a few sets of smFISH probes targeting the fluorescent tag mRNA sequence.

4.2.1 Identification of Syp targets with protein-trap insertions

I was interested in identifying Syp targets that are post-transcriptionally regulated in the brain. My previous examples with *pros* and *e93* highlighted that genes with long 3'UTRs might contain additional regulatory sequences for RBP-binding and subsequent regulation. Therefore, I decided to focus on targets featuring long 3'UTR extensions. Using 4 kb as a cut-off, I obtained a cohort of 41 Syp targets (Figure 4.1A). Interestingly, all 41 transcripts were also Imp targets as identified using RIP-seq in the larval brain (Samuels, Järvelin, et al., 2020). Moreover, this cohort was highly enriched for transcripts encoding TFs (32%) and RBPs (46%), consistent with the hypothesis that Syp and Imp are part of a more extensive transcriptional and post-transcriptional regulatory network (discussed in Chapter 3). Of these 41 targets, 26 had existing protein-trap lines of either GFP (Morin et al., 2001; Nagarkar-Jaiswal et al., 2015; Sarov et al., 2016) or YFP (Lowe et al., 2014) insertions. I examined the genomic insertion sites and only kept the lines where the insertion

'traps' all mRNA isoforms (i.e. constitutive insertions). I reasoned that it is important to rule out the possibility of artefactual RNA/protein discordance due to mRNA isoform differences (Figure 4.1B). After this selection and elimination process, I was left with 23 protein-trap lines to characterise.

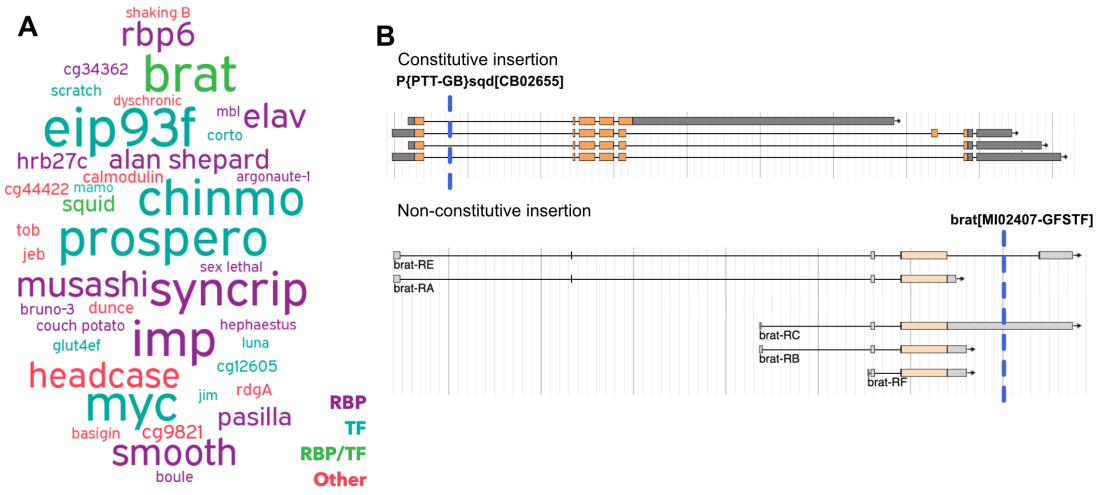


Figure 4.1: Identification of Syp targets existing as protein-traps. (**A**) Wordcloud showing Syp targets with highly extended 3'UTR longer than 4 kb. Texts are colour-coded based on RNA-binding or DNA-binding molecular functions. Font sizes are scaled with the number of Gene Ontology terms associated with 'Stem cell' or 'Nervous system function' for each gene. (**B**) Examples of protein-trap lines that were studied (Sqd::GFP, *top panel*) or not studied (Brat::GFP, *bottom panel*). The *sqd* genetic locus on the reverse strand has been horizontally flipped for visualisation purpose. Only the constitutive insertion into all mRNA isoforms were chosen for the downstream analysis. *MI02407-GFSTF* insertion in *brat* locus only captures the *brat-RE* isoform (located within the coding intron). Orange regions represent coding sequences and gray regions represent untranslated sequences.

4.2.2 Co-visualisation of protein and RNA using smFISH

To explore how gene expression is controlled in specific cell types, I used an imaging strategy to simultaneously assess transcription, cytoplasmic mRNA and protein levels throughout the whole tissue. From the 23 protein-trap lines, mRNA was detected using common *GFP* or *YFP* smFISH probes along with the fluorescence of the encoded GFP/YFP-tagged protein (Figure 4.2A). As demonstrated in the previous Chapter, smFISH images contain information about the level of transcription by visualising nascent transcripts at the endogenous gene locus in the nucleus (Figure 4.2C, *yellow arrow*). This strategy allowed me to systematically map the spatial distribution of mRNA and protein expression in many different cells and structures with high sensitivity. Neurons are known to utilise RNA localisation and local translation heavily, and the localised RNAs are

often present at low densities. I reasoned that the exquisite sensitivity of smFISH can be used in this setting to detect low-abundance transcripts. Therefore, in addition to central brain (CB) and NB lineages, I imaged the cell bodies and neuropils of the optic lobe (OL), the mushroom body (MB, equivalent to mammalian hippocampus), and the ventral nerve cord (VNC) (Figure 4.2B).

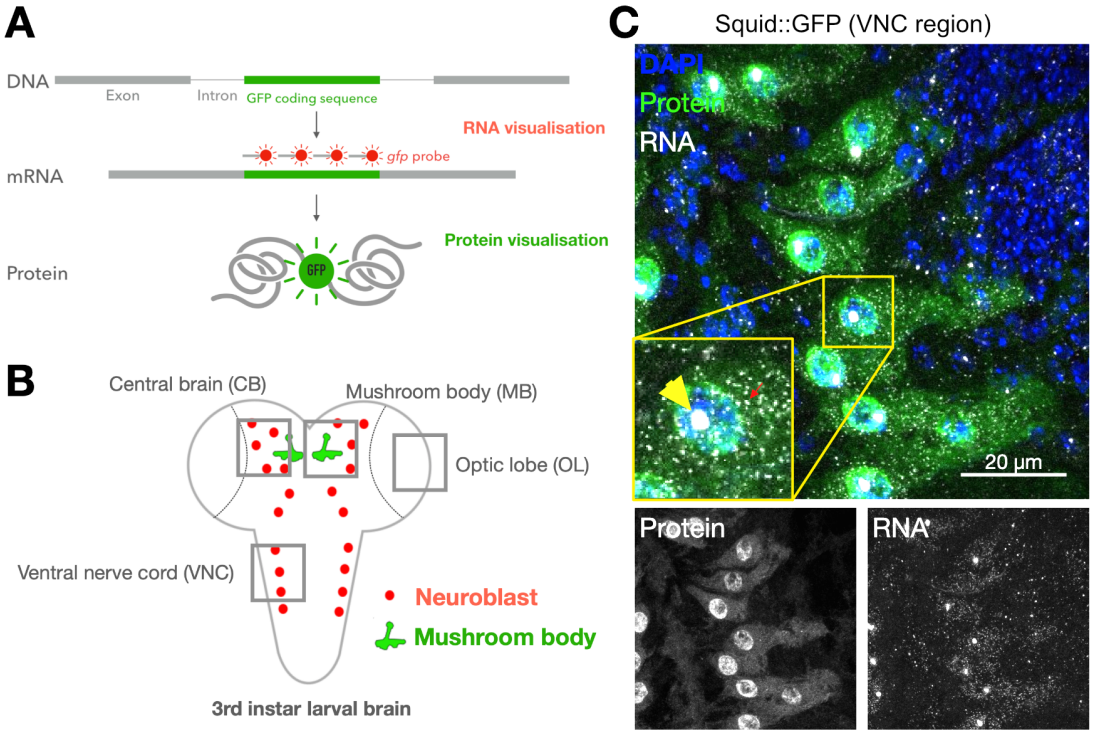


Figure 4.2: Schematic overview of the mRNA and protein expression survey. (**A**) Schematic diagram showing the strategy used to co-visualise protein and mRNA using smFISH in GFP protein-trap lines. (**B**) Compartments within the L3 larval brain that were imaged to score mRNA and protein distributions. (**C**) Representative smFISH confocal image (2 µm max projection), showing nascent transcription (*yellow arrow*), mRNA accumulation (*red arrow*), and protein expression. (n=2).

In order to determine whether the *GFP* smFISH probe can correctly detect the 'trapped' transcript, I performed dual-colour smFISH experiments on the *Squid::GFP* trap line using the *GFP* probe and *squid* (*sqd*) probe sets specific to the coding exon sequences (Figure 4.3A). I found that both *GFP* and *sqd* probes produced distinct diffraction-limited spots in the brain consistent with single-molecules, and showed a high degree of overlap between the two channels (Figure 4.3B). Importantly, the *GFP* probe did not produce a diffraction-limited signal in the *wild-type* (*OrR*) brains, demonstrating specificity.

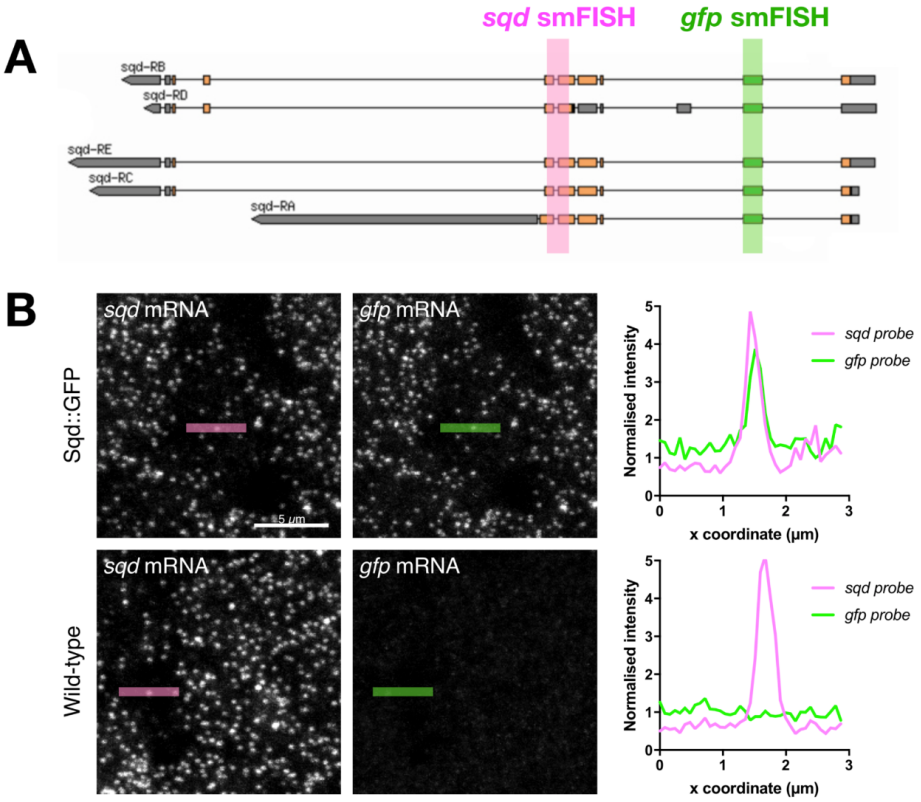


Figure 4.3: Validation of *GFP* probe specificity to mRNA containing *GFP* sequences. (**A**) Schematic diagram of the *GFP* exon insertion site (*green box*) in the gene, *squid* (*sqd*). Sequence regions targeted by the *GFP* probe and the endogenous *sqd* probe are highlighted. (**B**) smFISH image of L3 brain hybridised with *sqd* and *GFP* probes in *Sqd::GFP* (*top panels*) or *OrR* (*bottom panels*) brains. Confocal images (2 µm max projection) in the central brain region are shown. Right panels show intensity line profiles across the diameter of representative diffraction-limited spots. (n=3).

4.2.3 Prevalence of discordance between RNA and protein expressions of Syp targets

Using this imaging strategy, I performed smFISH experiments in 23 protein-traps to ask where protein and mRNAs are expressed relative to each other throughout the CNS. To systematically analyse the dataset, I made a manual scoring system to assess the presence or absence of RNA/protein in each compartment. The full binary scoring table is given in Appendix D. In this Chapter, I mainly focus on the central brain (CB) region, which harbours Type I NB lineages that make up 60% of all NBs in the L3 brain.

Surprisingly, I found that 57% (13/23) of the genes tested displayed a lack of correlation between protein and mRNA expression in the NB lineage. Here, two main types of discordance were observed. The first category was the discordance between the level of primary transcription

(nascent transcription foci) and the amount of cytoplasmic mRNA, indicating regulation via mRNA stability (Figure 4.4). For example, *smooth* transcription is evident in neural stem cells (NSC, NB) and ganglion mother cells (GMCs), but mRNA does not accumulate, so protein expression is absent. Cytoplasmic mRNA and protein expressions were only visible in neurons. Therefore, this control mechanism ensures cell-type specific expression of protein despite the ubiquitous transcription. This pattern of RNA stability control is reminiscent of *pros*, so it would be interesting to follow up on whether different 3'UTR isoforms of *smooth* are regulated differently in the progeny. The other category was regulation through translational control, where protein signal is absent despite cytoplasmic mRNA accumulation (Figure 4.4). As described in the earlier Chapter, *e93* transcripts were present in both NBs and progeny cells, but, the protein was only detected in the stem cell. This suggests that either the translation of *e93* is repressed in the progeny, or it is highly up-regulated in the NBs, giving rise to cell-type specificity. However, I note that differential regulation of protein stability could also give rise to such discordant cell type-specific mRNA and protein distributions. While the current approach aims to discover evidence of post-transcriptional control, the precise mechanism of regulation should be dissected with techniques such as puromycin proximity ligation assay (Dieck et al., 2015). When all neuropil compartments were considered, 18/23 or 78% of genes showed discordance in at least one cellular compartment. In summary, these results demonstrate the discordant relationship between mRNA and protein distribution is very common and may suggest a wider prevalence of post-transcriptional regulation underlying gene expression in the CNS.

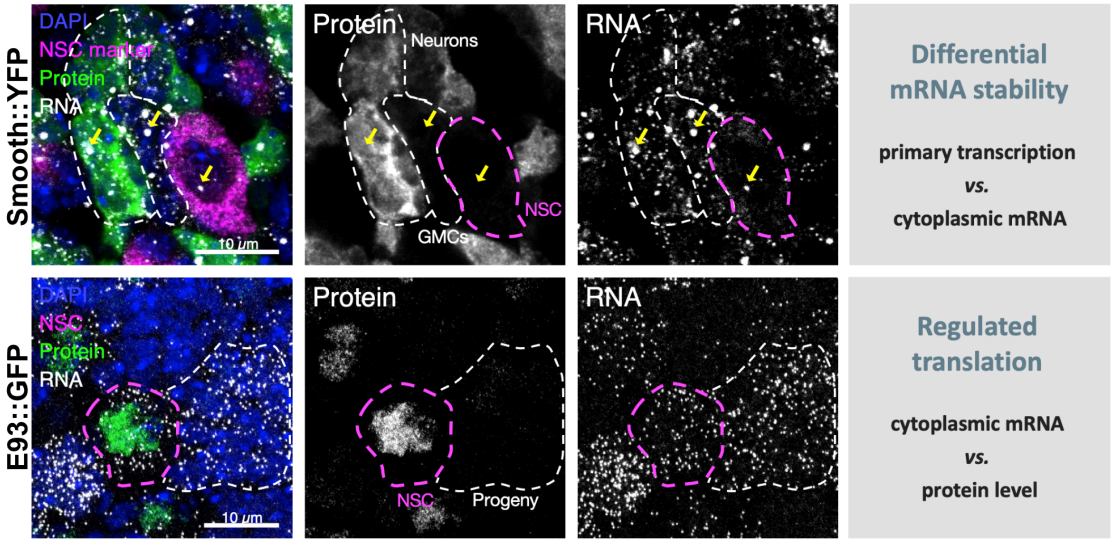


Figure 4.4: Examples of discordant mRNA and protein distributions. Representative smFISH confocal images (5 µm max projection) showing two types of discordant profiles. Neuroblasts (NSCs) are labelled in magenta and their progeny cells are denoted with white dashed lines. Yellow arrows represent the transcription foci in the nuclei. (n=3, 5 field of views per replicate). Note that E93::GFP images are duplicated from *Figure 3.26A* in Chapter 3 for illustration purposes. Please see the Appendix D for the full list of protein and mRNA distribution scoring in 23 protein-trap lines.

4.3 Systematic analysis of post-transcriptional regulation across the nervous system

I then asked whether the discordant profile of mRNA and protein expression is a common feature in the *D. melanogaster* nervous system. To achieve this, I became part of a greater effort from the Davis lab to systematically analyse protein and mRNA expression across a larger set of genes. The project was led by Dr Joshua Titlow, Dr Ana Palanca, Maria Kiourlappou, and myself (Titlow et al., 2022). This work produced a wealth of data composed of more than 1,300 figures covering many aspects of the larval nervous system. In the process, I carried out smFISH and fluorescent protein imaging on 65 distinct genes. However, for brevity and to highlight my primary contributions, I present the works leading to the methodology development and analysis of the results in the NB lineages.

4.3.1 Selection of 200 YFP-trap collections

To test whether my earlier findings with the Syp targets are a general phenomenon across *D. melanogaster* transcriptome, we utilised the Cambridge Protein Trap Insertion (CPTI) collection.

This collection features more than 500 Venus-YFP (referred to as 'YFP' hereafter) protein-trap lines originally created with an engineered *PiggyBac* transposon inserted semi-randomly into the *Drosophila* genome (Lowe et al., 2014). As with the filtered Syp targets, we retained 200 CPTI lines that recapitulated constitutive YFP-traps between mRNA isoforms. To assess whether these 200 genes can represent the diversity of gene expression in the nervous system, we analysed the functional annotation, expression level and structure of the selected genes.

For the functional annotation, Dr Titlow compared the diversity of Gene Ontology (GO) terms between the list of 200 CPTI genes and all protein-coding genes. Unique parent GOSlim terms were used to take the hierarchical nature of GO into account. Dr Titlow found that our selected genes represent 89.9% of GOSlim terms in all three GO categories (Titlow et al., 2022), indicating representative functional diversity. Next, we compared the distribution of mRNA expression levels with the published L3 brain transcriptome (Samuels, Järvelin, et al., 2020). We found that the 200 genes covered a broad range of RNA abundance levels (Figure 4.5A), however, we note that our selection of genes was on average more abundant than the transcriptome average. Finally, Maria Kiourlappu analysed gene structure. She found that the CPTI genes tend to have longer introns, consistent with the original report (Figure 4.5B) (Lowe et al., 2014), but notably, the mean lengths of 3'UTR extensions were comparable to that of all protein-coding genes. Considering the significant regulatory potentials of 3'UTR in mRNA localisation (Tushev et al., 2018), we propose that the mRNA localisation results and conclusions from the 200 genes are a fair representation of the protein-coding genome. Taken together, these results show that our selection of genes is moderately representative of global gene expression and does not significantly undersample the complexity of gene functions.

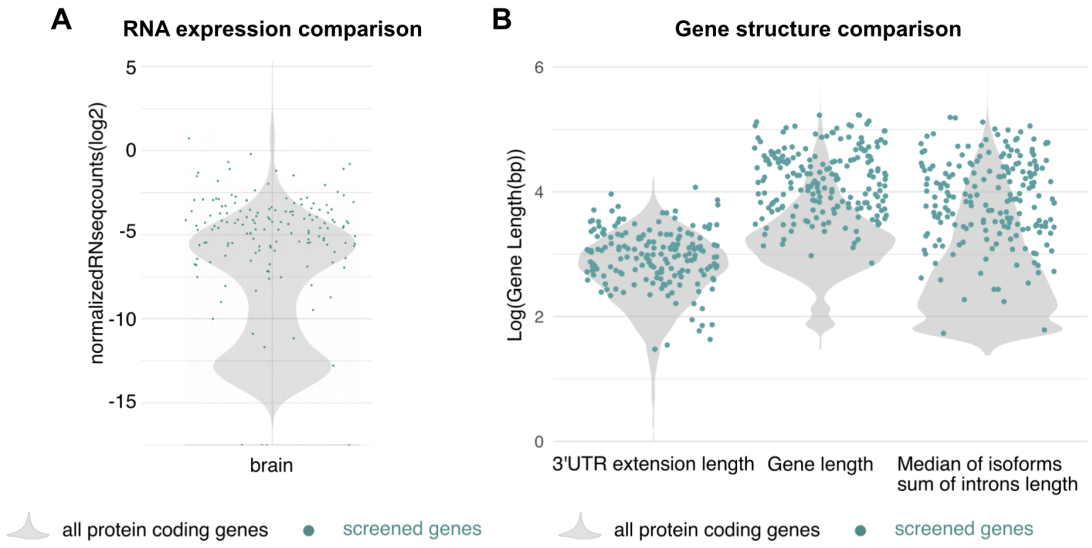


Figure 4.5: Properties of the selected 200 YFP-trap genes. (**A**) The transcript expression levels of the screened genes (*green symbol*) compared to the L3 brain background transcriptome (*gray violin plot*). The background transcriptome was obtained from (Samuels, Järvelin, et al., 2020). (**B**) Comparison of gene structures of the screened genes to all protein-coding genes. The screened genes have longer introns and overall gene length, but the 3'UTR extension lengths are comparable to all protein-coding gene average. Dr Aino Jarvelin and Maria Kiourlappou created the visuals.

4.3.2 Systematic image analysis pipeline

The smFISH experiments I performed on the 200 YFP-trap brains were very similar to the experiments I carried out with the Syp targets (Figure 4.2A-B). In addition to the four brain compartments, larval neuromuscular junction (NMJ) and motoneuron nerves were imaged. In summary, six compartments were imaged per gene, and ~1,300 associated figures were generated (Figure 4.6). Analysis of such large microscopy datasets can be a challenging task, therefore we devised a generalisable and scalable workflow to facilitate image analysis, annotation and data exploration (Titlow et al., 2022). This approach consists of raw image storage in OMERO (Goldberg et al., 2005) and figure generation for each compartment using OMERO.Figure web application (Allan et al., 2012). In our repository, we provided confocal images that are contrast-normalised between datasets to view smFISH spots with ease. In addition, the whole dataset is stored in the OMERO server for any user to browse the full 3D volumes with user-defined contrast. Then, Maria Kiourlappou developed a Python application (with some technical assistance with Python graphical user interface (GUI) programming from Dr

David Pinto) to systematically annotate OMERO.Figure images, providing a GUI to facilitate scoring (Kiourlappou et al., 2022). This strategy is conceptually similar to my manual scoring presented previously (Figure 4.4). Here, the GUI supplies relevant questions for each compartment, such as "is RNA present in neuroblasts?". Three scorers with expert knowledge of fly nervous system anatomy annotated all ~ 1,300 figures independently, and the conflicts were resolved by majority voting (Titlow et al., 2022).

The full set of figures and the annotations were uploaded to the Zegami web application to facilitate data exploration. Zegami is an AI-enabled image analysis platform that allows the incorporation of bioinformatics metadata. Users can view images associated with a specific gene from our repository with easy access to public transcriptomic and proteomic databases (Figure 4.7). In summary, the Zegami platform places our image dataset in a rich bioinformatic context to deploy machine-learning algorithms and facilitate gene candidate selections for future studies. The Zegami repository was created by Maria Kiourlappou and is available from this link: `https://bit.ly/3uDtOrU`.

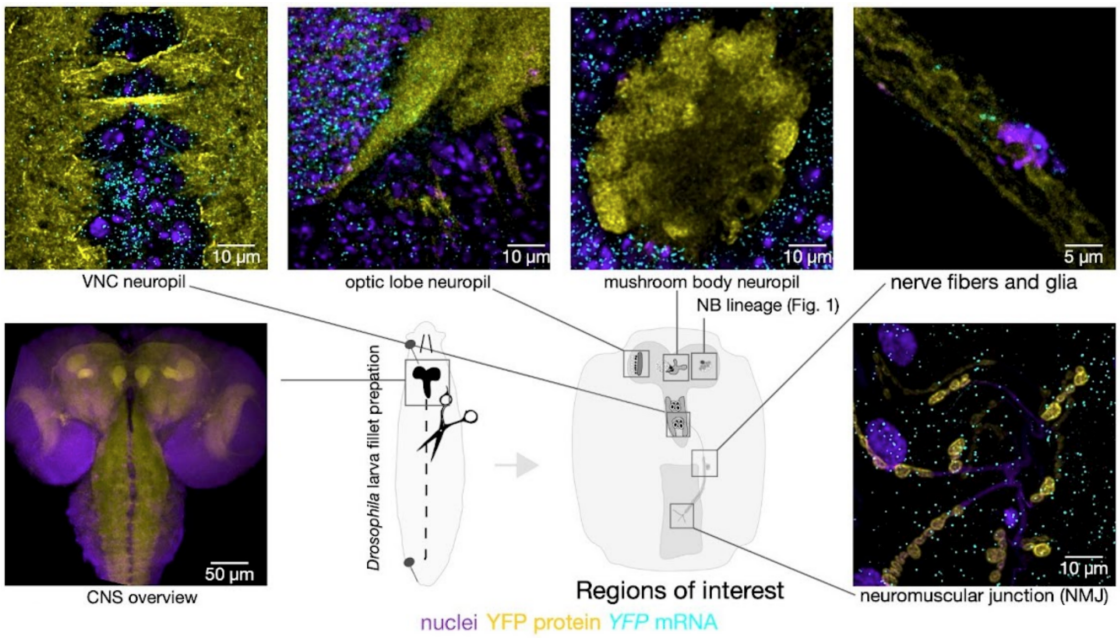


Figure 4.6: Overview of nervous system compartments imaged for the YFP-trap screen. Images were obtained from multiple nervous compartments focusing on neuronal soma (cell bodies) and neuropil (axon and dendrite bundles) structures, as well as NB lineges. This figure is adapted from Figure 2 of (Titlow et al., 2022). Representative images from n=3 field of views are shown.

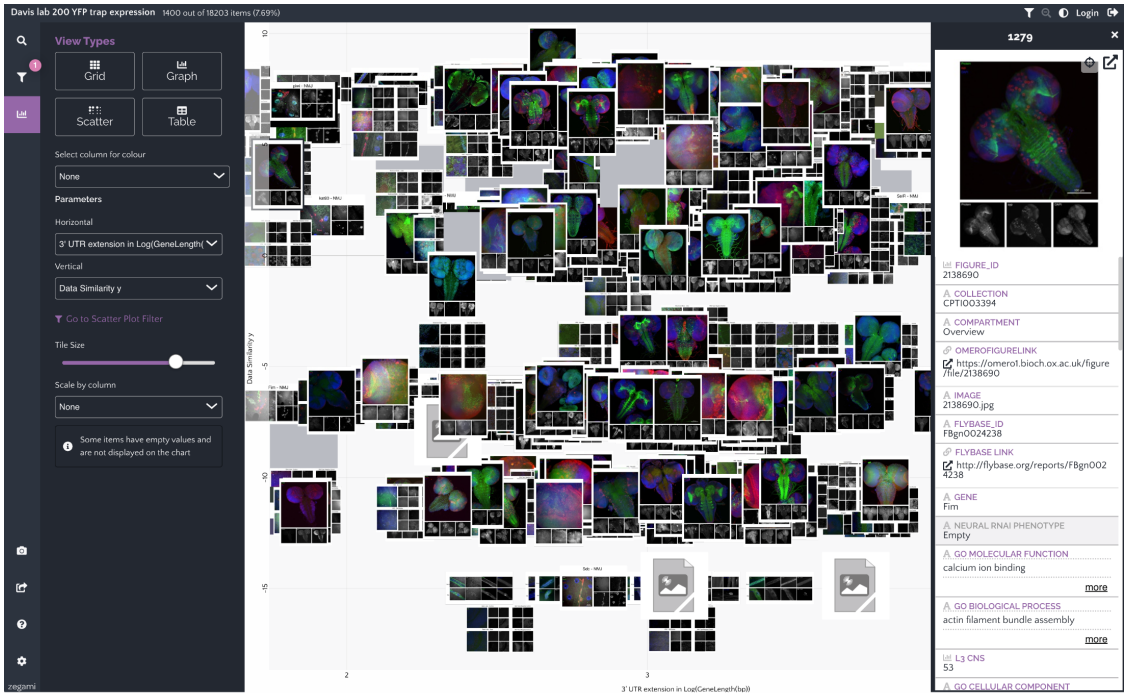


Figure 4.7: Zegami as a machine-learning enabled data exploration tool. Example screenshot of the Davis lab Zegami repository featuring ~1,300 figures generated from the screen. Zegami web application allows easy exploration of data and incorporation of additional databases for machine-learning algorithms. This Zegami repository can be accessed through this link: `https://bit.ly/3uDtOrU`.

4.3.3 Post-transcriptional control of NB differentiation

To assess the prevalence of post-transcriptional regulation in NB differentiation, mRNA and protein expression were examined in the larval central brain. Three experts scored the expression in a binary manner within the population of asymmetrically dividing NBs and their progeny cells. Given the absolute sensitivity of smFISH, I defined the number of mRNA molecules should be more than two to be considered 'present', ruling out the potential effect of transcriptional noise (M. M. K. Hansen et al., 2018; Raj & van Oudenaarden, 2009). Protein abundance was scored relative to *wild-type* brains' YFP channel fluorescence intensity. I used *syp exon* smFISH as an NB marker, which is highly expressed almost exclusively in central brain NBs (L. Yang et al., 2017).

Using Zegami, I was able to classify smFISH results into four gene expression categories: (i) 'No expression' (21.5%); (ii) 'Global expression' of protein and RNA across all cell types (57%); (iii) 'Transcriptional regulation', where protein level correlated with mRNA (15.5%); and (iv) 'Post-transcriptional regulation', showing discordant mRNA and protein distributions (6%)

(Figure 4.8A). As expected, 'Transcriptional' and 'Post-transcriptional' regulation categories gave rise to cell-specific expression patterns, contributing to 21.5% (43/200) of the genes screened (Figure 4.8B). Therefore, this cohort may represent the spectrum of dynamically expressed genes in response to NB differentiation. Notably, post-transcriptional regulation was unexpectedly common (28%, 12/43) amongst genes that were expressed in a cell-type specific manner. This category of genes exhibited obvious discordance throughout the NB lineage, where protein levels were uncorrelated to the relatively equivalent rates of transcription between cells. This result suggests a significant role of mRNA metabolism in cell-specific functions. Consistently, 4/12 genes (*cno*, *pbl*, *qkr58E-1*, *Rm62*) have been associated with NB division phenotypes from a genome-wide RNAi screen (Neumuller et al., 2011). In contrast to Syp targets, I did not find particular enrichment of TFs or RBPs in this cohort.

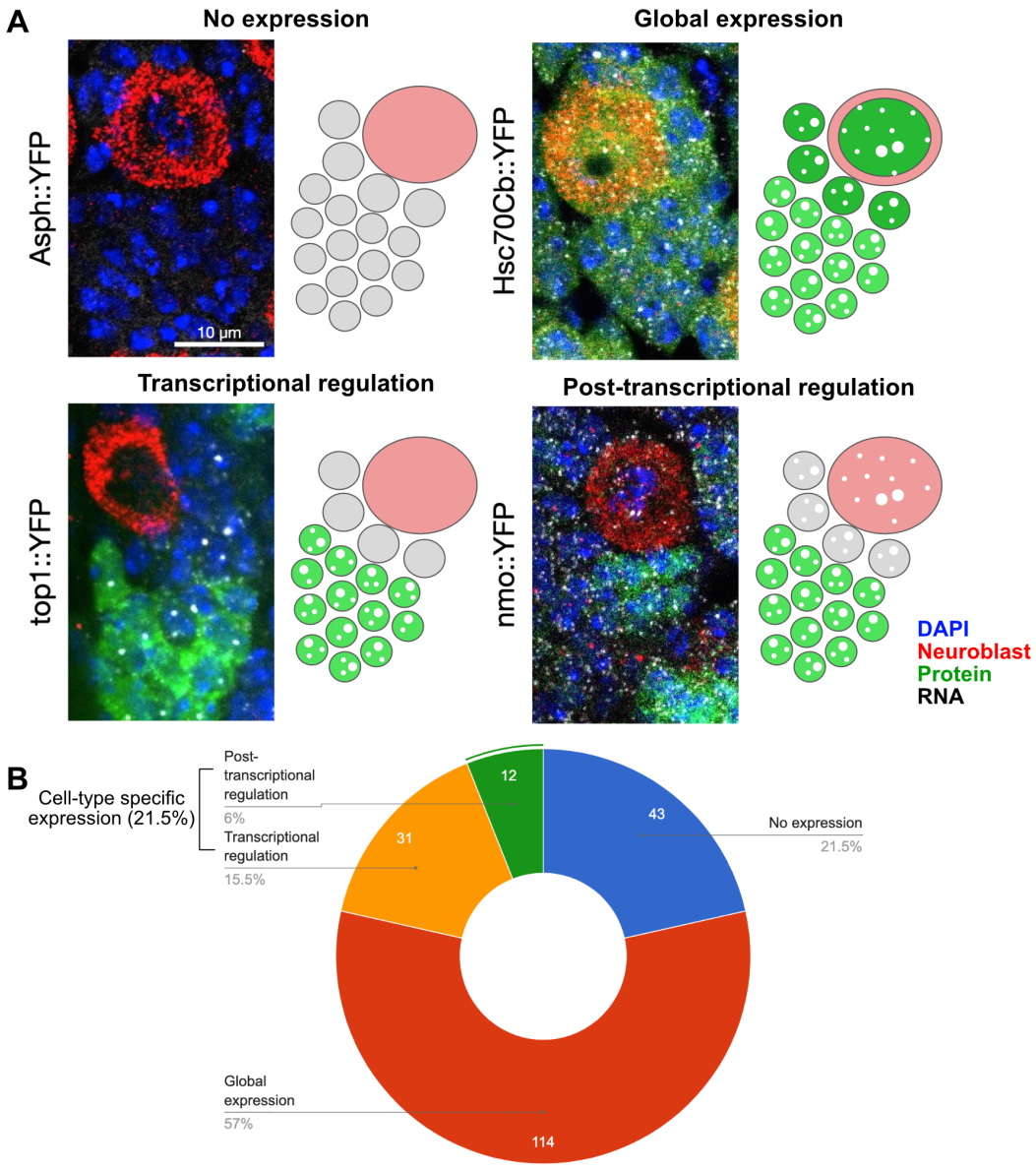


Figure 4.8: Systematic analysis of protein and mRNA expression pattern in the NB lineage. (**A**) Four distinct patterns of protein and mRNA expression in the central brain NB lineage. 'Transcriptional regulation' and 'Post-transcriptional regulation' categories together give rise to cell-type specific expression patterns (43/200 genes). Representative NB lineage from n=2 replicates are shown. (**B**) Pie chart showing the relative distribution of the four gene expression categories.

4.3.4 Summary of screen results

In addition to the NB lineage, we asked where proteins and mRNAs are expressed relative to each other throughout the six nervous system compartments. Neuropil compartments (OL, VNC, MB) were further divided into 'Soma' and 'Neuropil' regions to assess the proportion of transcripts showing localised mRNA in the neuron periphery. Compilation of image annotations from the

three expert scorers revealed that 196/200 or 98% of the screened genes show discordant mRNA and protein expression in at least one compartment. Furthermore, 137/200 (68.5%) of genes showed RNA localisation in at least one synaptic compartment (Figure 4.9). The full heatmap illustrating the frequency of discordance for each gene at different compartments is provided in Appendix E. These results illustrate mechanisms involving post-transcriptional regulation and localised mRNA are very common for the genes expressed in the brain, which may include many new potential regulators of neuronal function. It is also notable that our data suggest protein transport significantly influences protein distribution in neuropil compartments (indicated dark green in Figure 4.9) where 40-60% of proteins detected in the synaptic compartment localised without their mRNAs. This 'protein only' observations without the mRNA suggest the translated protein to be transported from soma to axons, whole mechanism will also be important for shaping the local axonal proteome in addition to the locally translated proteins (Holt et al., 2019). In conclusion, we provided a unique resource describing discordance between protein and mRNA levels for 200 genes in various nervous system compartments with high sensitivity. We also demonstrated the utility of systematic image analysis and a rich data exploration environment (Titlow et al., 2022).

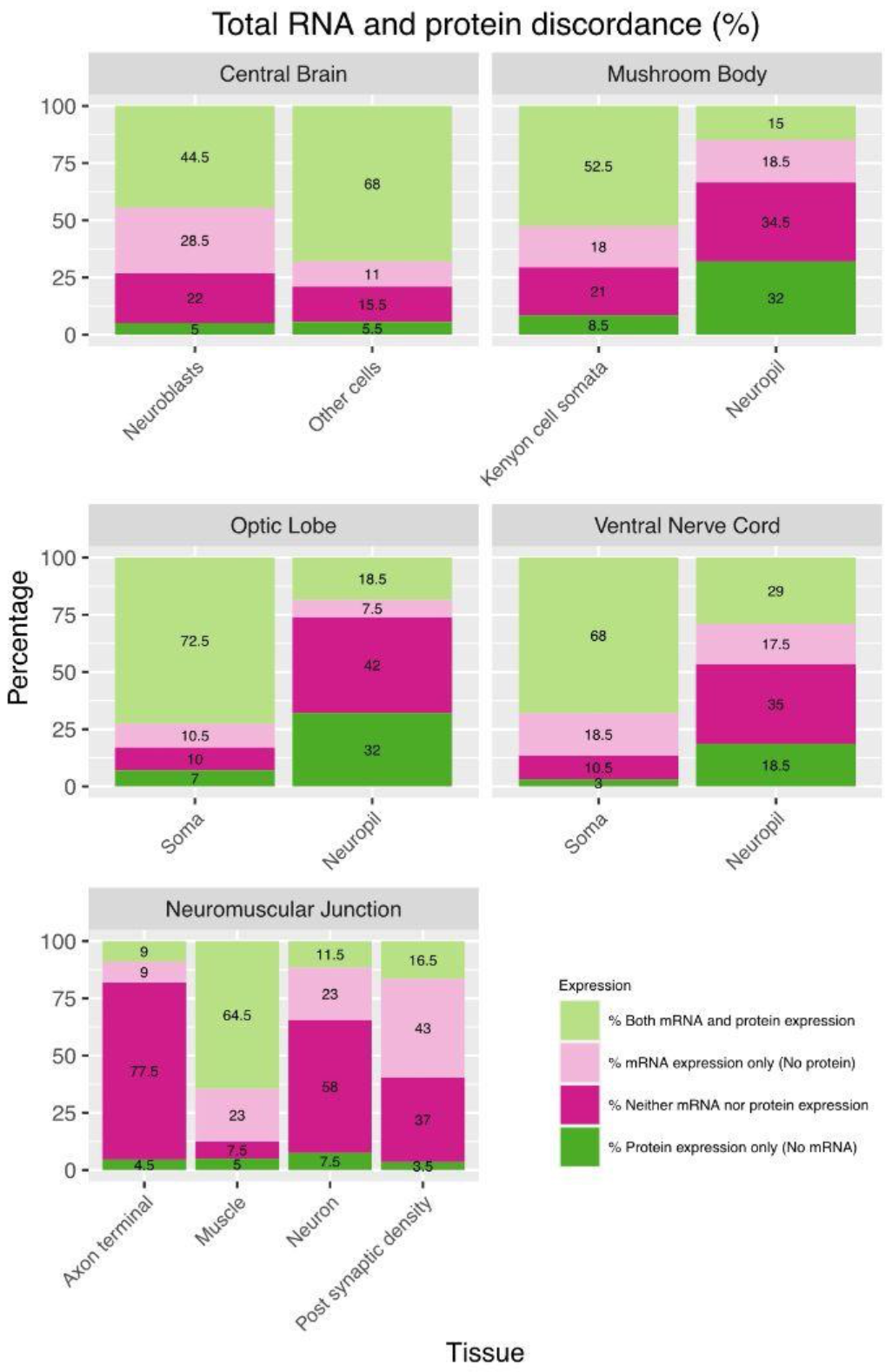


Figure 4.9: Summary annotations from the scoring of the 200 YFP-trap datasets. Three expert scorers annotated images from the screen for the presence or absence of protein and mRNA at each nervous system compartments. The plots show the percentage of genes showing each expression pattern in different compartments. The visual was created by Maria Kiourlappou.

4.4 Conclusion and discussion

Gene expression encompasses transcription, translation and turnover of protein and mRNA. The extent to which protein abundance correlates with mRNA or transcription level is a highly debated topic and probably is context-dependent in many cases (Buccitelli & Selbach, 2020). Post-transcriptional regulation governs many aspects of RNA metabolism and often manifests in the discordant distribution of mRNA with respect to its encoded protein. However, gene expression is rarely studied end-to-end (from transcription in the nucleus to degradation in its final cytoplasmic destination), making it difficult to address the relative contribution of post-transcriptional control in cellular homeostasis (Buccitelli & Selbach, 2020; Y. Liu et al., 2016). In this Chapter, I described a highly scalable approach to systematically investigate the mechanism of gene expression in complex tissue with high sensitivity and spatial resolution. This approach features a novel usage of existing protein-trap lines (Lowe et al., 2014; Nagarkar-Jaiswal et al., 2015), allowing visualisation of mRNA with *YFP* or *GFP* probes and protein detection with the tagged fluorescent protein. Coupled to automated imaging and a guided scoring system, this method can interrogate gene expression profiles at a much higher throughput compared to gene-by-gene approaches. This approach was first applied to a cohort of Syp targets with long 3'UTR extensions and subsequently expanded to a collection of 200 genes. Furthermore, with Dr Joshua Titlow, Dr Ana Palanca and Maria Kiourlappou, I also presented a generalisable workflow to facilitate image analysis and enhance the data exploration experience (Titlow et al., 2022). With the ever-growing resource of protein-traps and genetic tools, our systematic approach may benefit from a wider audience working with a variety of genetically manipulatable model organisms (Lee et al., 2018b; Nagarkar-Jaiswal et al., 2015; Sarov et al., 2016).

A smaller-scale experiment with Syp targets revealed the prevalence of discordant protein and mRNA distributions in the NB lineage, demonstrating a widespread manifestation of post-transcriptional regulation. An important follow-up of this result would be to investigate whether the discordant distribution of mRNA and protein is due to Syp-mediated post-transcriptional regulations. It would be interesting to examine protein-trap expressions in *syp* mutant background

and score for any altered protein and mRNA distributions. The frequent level of discordance was also observed in an equivalent analysis performed with the 200 genes, albeit at a lower scale than Syp targets. I interpret my results as indicating that targets of RBPs are more likely to be regulated post-transcriptionally. Furthermore, because Syp targets are enriched in transcripts with longer 3'UTR isoforms than the 200 genes or to the transcriptome, perhaps it is not surprising to see higher discordance rates within the RBP targets. Nevertheless, both datasets highlight the general need to improve our understanding of how these sets of genes are regulated to impact neural differentiation. To achieve this goal would require the identification of trans-acting RBPs and elucidation of how they interact with their targets. However, out of 523 canonical RBPs (Sysoev et al., 2016) 307 are expressed in NBs and their progeny (Berger et al., 2012), so characterising the function of so many RBPs is a very a daunting task indeed. Nevertheless, Chapter 3 can be considered one case study, utilising iCLIP and smFISH. This workflow could be a useful guide to design future high-throughput experiments.

Although our recently submitted systematic study has also described the prevalence of synaptic mRNA localisation and local translations, I have not had the space to describe the work here (see Figure 5 of (Titlow et al., 2022)). Nevertheless, it is worth noting that our description of numerous axonally localised transcripts is consistent with transcriptomic studies that demonstrated the presence of thousands of locally present RNAs in neuron extremities (von Kügelgen & Chekulaeva, 2020) and their functions in axon development and synaptic plasticity (Cioni et al., 2018; Holt et al., 2019; Holt & Schuman, 2013; Shigeoka et al., 2016). Notably the transcripts detected with smFISH in at least one of the *D. melanogaster* neuropil compartments were also present in, on average, 8 synaptic transcriptome studies as mouse orthologues (Titlow et al., 2022; von Kügelgen & Chekulaeva, 2020). Therefore, our dataset also serves as single-molecule validation of these reports.

In summary, I have introduced a high-resolution approach to systematically co-visualise the distribution of mRNAs and proteins in the intact nervous system. I also highlighted the utility of the single-molecule detection method for gaining insight into cell-specific stability

regulation or the transport of rare mRNA molecules to synaptic compartments. The ability to detect and quantify rare RNAs can provide critical insight into how gene expression is regulated, and this benefit is not limited to nervous system biology. In the next Chapter, I present the applicability of single-molecule FISH method to characterise rare replicating viral particles in the early stages of the virus infection.

5

Single-molecule detection of SARS-CoV-2 RNA provides new insights into infection dynamics and variant differences

5.1 Introduction

Transcription, or in some cases replication, of RNA is a fundamental process underlying numerous life processes, and it must be tightly controlled to ensure the correct biological outcome. In the previous Chapters, I demonstrated how a sensitive detection of individual RNA molecules could provide new information on post-transcriptional gene regulation in the context of neurobiology. Techniques such as smFISH offer highly specific and spatially resolved information on gene expression even in heterogeneous systems, but more importantly, it is applicable widely to multifold systems. In 2020, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused the COVID-19 pandemic. The unprecedented public health impact of the virus has been met with an equally unprecedented response from scientists. Much of this effort has focused, appropriately, on the mechanisms of virus entry into host cells. However, the early replication events of the newly internalised virus, despite their essential role in establishing productive infection (Boersma et al., 2020; Singer et al., 2021), remain poorly understood.

Typically, most current assays to study viral replication dynamics include engineered viruses containing reporter-genes, immunofluorescence, or western blots of viral antigens. However, such

methods are not sufficiently sensitive enough to pick up the rare virus in the early hours of the infection (Boersma et al., 2020; Dhir et al., 2018; Kamel et al., 2021). Finely-controlled RT-qPCR and RNA-seq can be sensitive and highly informative, but such ensemble methods are unable to assess spatial information and mask cell-to-cell heterogeneity in response to virus infection (Fraser & Kærn, 2009; Raj & van Oudenaarden, 2009). Alternatively, intracellular localisations of the early SARS-CoV-2 replication events were described using electron microscopy (Cortese et al., 2020; Klein et al., 2020; Laue et al., 2021) but with limited throughput for kinetic measurements.

Exploiting the single-cell analyses in virology has offered incredible insights into the significant variability of individual cells in their susceptibility to viral infection and the stochastic aspects of gene expression and infection processes (Doganay et al., 2017; Patil et al., 2015). Here, many new discoveries were made using of fluorescence *in situ* hybridisation (FISH)-based methods (Billman et al., 2017; Boersma et al., 2020; Chou & Lionnet, 2018; Ramanan et al., 2016; Shulla & Randall, 2015; Singer et al., 2021). However, this approach has been applied to SARS-CoV-2 in a limited capacity (Acheampong et al., 2021; Burke et al., 2021; Rensen et al., 2022; Wing et al., 2021), with most studies utilising amplification-based signal detection methods to visualise virus-derived RNAs (Best Rocha et al., 2020; Carossino et al., 2020; Guerini-Rocco et al., 2020; Jiao et al., 2020; Kusmartseva et al., 2020; Lean et al., 2020; X. Liu et al., 2020). While informative, the use of chromogenic histochemical detection or enzyme amplification presents limitations to the sensitivity and specificity to detect individual RNA molecules.

During the nationwide lockdown, I was offered an opportunity to collaborate with multiple virology labs in Oxford to address this gap in the knowledge of the SARS-CoV-2 lifecycle. Therefore, in this Chapter, I aimed to develop a smFISH method, based on earlier published protocols (Femino et al., 1998; Raj et al., 2008; Singer et al., 2021; Titlow et al., 2018), to visualise SARS-CoV-2 RNAs with high sensitivity, specificity and spatial precision. I examined the utility of the method for detecting and quantifying various SARS-CoV-2 RNA species in infected cells and tissues. I then organised a computational pipeline to batch process large smFISH image datasets. Next, I profiled viral RNA replication kinetics at the single-molecule and cell level in the

first hours of infection. Finally, I demonstrated the application of the new powerful approach to compare early replication kinetics of the ancestral VIC strain to the B.1.1.7 Variant of Concern (VOC), which was the dominant variant circulating in England when the study was conducted.

5.1.1 Specific aims of this chapter

1. To establish single-molecule detection of SARS-CoV-2 RNA
2. To characterise early replication kinetics of SARS-CoV-2 RNA
3. To compare replication kinetics of SARS-CoV-2 strains

5.1.2 Publication

Contents of this Chapter have been published as follows (J. Y. Lee et al., 2022):

Lee, Jeffrey Y, Peter AC Wing, Dalia S Gala, Marko Noerenberg, Aino I Järvelin, Joshua Titlow, Xiaodong Zhuang, et al. '**Absolute Quantitation of Individual SARS-CoV-2 RNA Molecules Provides a New Paradigm for Infection Dynamics and Variant Differences**'. Edited by Miles P Davenport. *ELife* 11 (20 January 2022): e74153. `https://doi.org/10.7554/eLife.74153`.

Multiple co-authors contributed to this publication and where appropriate, their contributions are acknowledged in the text. SARS-CoV-2 remains a significant threat to human health, requiring dedicated containment facilities to study their virology. Unless otherwise stated, virus propagation and viral infection protocols were performed by Dr Peter Wing, Dr Xiaodong Zhang (Jane McKeating Lab; Nuffield Department of Medicine, Oxford), or Dr Marko Noerenberg (Alfredo Castello Lab; Centre for Virus Research, Glasgow) in Category 3 containment labs. Only deactivated samples were brought into the Category 2 laboratory for experimentation and imaging.

5.2 Visualising SARS-CoV-2 RNA at the single-molecule resolution

5.2.1 Sensitive detection of SARS-CoV-2 RNA

To analyse early events during SARS-CoV-2 replication, I designed smFISH experiments using fluorescently labelled DNA antisense oligonucleotide probes. Using the Stellaris Probe Designer, ~150 candidate sequences were obtained against the viral ORF1a region, which were subsequently BLAST filtered against the human transcriptome. Of the filtered ~90 sequences, the 5' most 48 sequences were chosen, which target the viral ORF1a to detect positive-sense gRNA (+gRNA), referred to as the '+ORF1a' probe set (Figure 5.1A). To test +ORF1a probes, Vero E6 cells were inoculated with SARS-CoV-2 (VIC strain) at a multiplicity of infection (MOI) of 1 for 2 h, and non-internalised viruses were removed by trypsin digestion to synchronise the infection. Cells were fixed straight after trypsin treatment, then hybridised with +ORF1a probes, and imaged using a spinning disk confocal microscope. smFISH detected virus-specific signals in the infected Vero E6 cells, visible as well-resolved diffraction-limited single spots with consistent fluorescence intensity and shape (Figure 5.1B). The unimodal distribution of fluorescence intensity suggests single-molecule resolution was achieved (Raj et al., 2008) (Figure 5.1C).

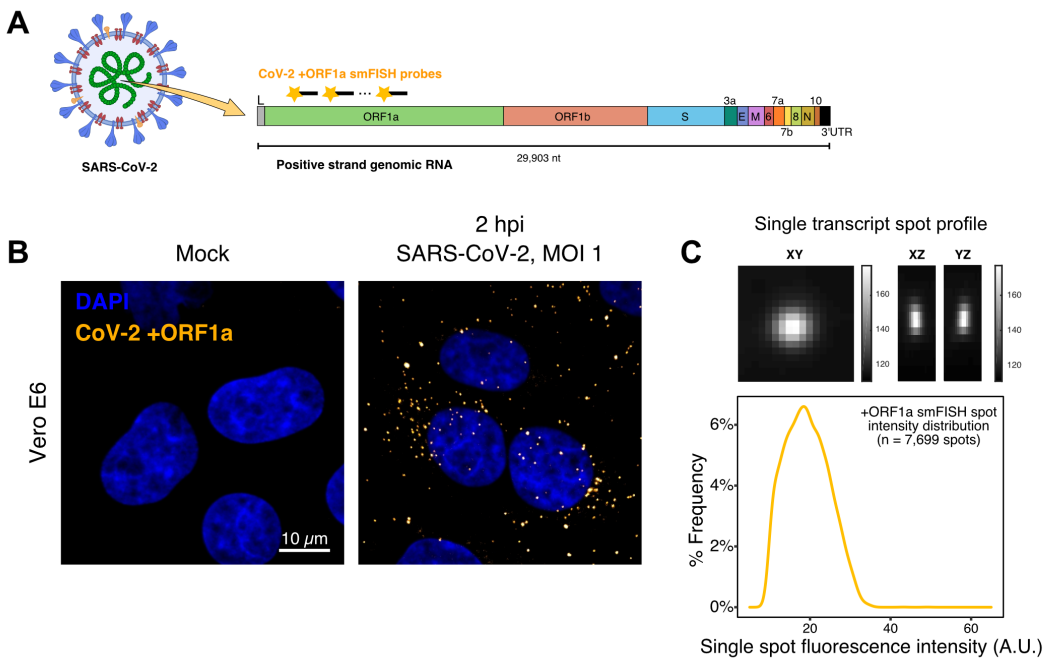


Figure 5.1: Single-molecule detection of SARS-CoV-2 genomic RNA. (**A**) Schematic of single-molecule fluorescence *in situ* hybridisation (smFISH) for detecting SARS-CoV-2 positive strand genomic RNA (gRNA). (**B**) Confocal image (4 µm maximum intensity projection) of smFISH against SARS-CoV-2 (VIC) infected Vero E6 cells at 2 hours post infection (hpi). Uninfected 'Mock' cells were used as a control of probe specificity against the virus. A representative field of view from three replicate experiments is shown. (**C**) Reference spatial profile of a diffraction-limited +ORF1a smFISH spot calculated from 7699 spots. The spots were counted from three biological replicate experiments.

Next, I assessed the utility of the technique to spatially locate and quantify SARS-CoV-2 gRNA across different infection timepoints (Figure 5.2A). After the virus inoculation (MOI 1) and trypsin treatment, infected Vero E6 cells were incubated with fresh culture media for designated hours post infection (hpi). At 2 hpi, most of the smFISH signal appeared as diffraction-limited single gRNA spots corresponding to the incoming virus. Additionally there were a small number of brighter and spatially extended foci containing multiple gRNAs. These foci likely represent early RNA replication events. Notably, each cell contained many intracelluar gRNA foci despite using the MOI of 1 (63% of the cells receiving one or more infective particles). Considering the virus titre was calculated with the plaque assay in the same Vero E6 cell line, this result suggests my virus stock contains many defective viral particles, albeit fusogenic. At 8 hpi, I noted a significant expansion in the number of single and multi-gRNA signals. By 24 hpi, a sharp increase in viral gRNA density spatially saturated the cell. Still, there was further

enrichment of multi-RNA foci in the proximity of the perinuclear region, consistent with the association of viral replication organelles with membranous structures derived from the ER-Golgi Intermediate compartment (ERGIC) (V'Kovski et al., 2021). Despite the spatial saturation of signals, individual gRNA molecules could be observed at the periphery of cells, consistent with individual particles being exocytosed, as detected with electron microscopy (Cortese et al., 2020; Klein et al., 2020). The SARS-CoV-2 virion tightly compacts its gRNA with hundreds of viral structural proteins, which may hamper its detection via antisense probes (Iserman, Roden, Boerneke, Sealfon, McLaughlin, Jungreis, Fritch, et al., 2020). Therefore, I assessed whether encapsulated gRNA is accessible to the probes. SARS-CoV-2 particles from the viral stock were immobilised on poly-L-lysine-coated coverglass and subsequently hybridised with +ORF1a probes. In this experiment, I observed a large number of smFISH spots displaying an intensity-matched profile of individual RNA molecules, suggesting that detection of gRNA within viral particles was achieved (Figure 5.2B). In contrast, no signal was observed in the Mock condition where coverglass was incubated with PBS only. Therefore, the smFISH method detected +gRNA molecules in all the expected subcellular locations (free in the cytoplasm, in clusters at the periphery of the nucleus, and in virions), reflecting different steps of the viral life cycle (Cortese et al., 2020; Hackstadt et al., 2021; Mendonca et al., 2021).

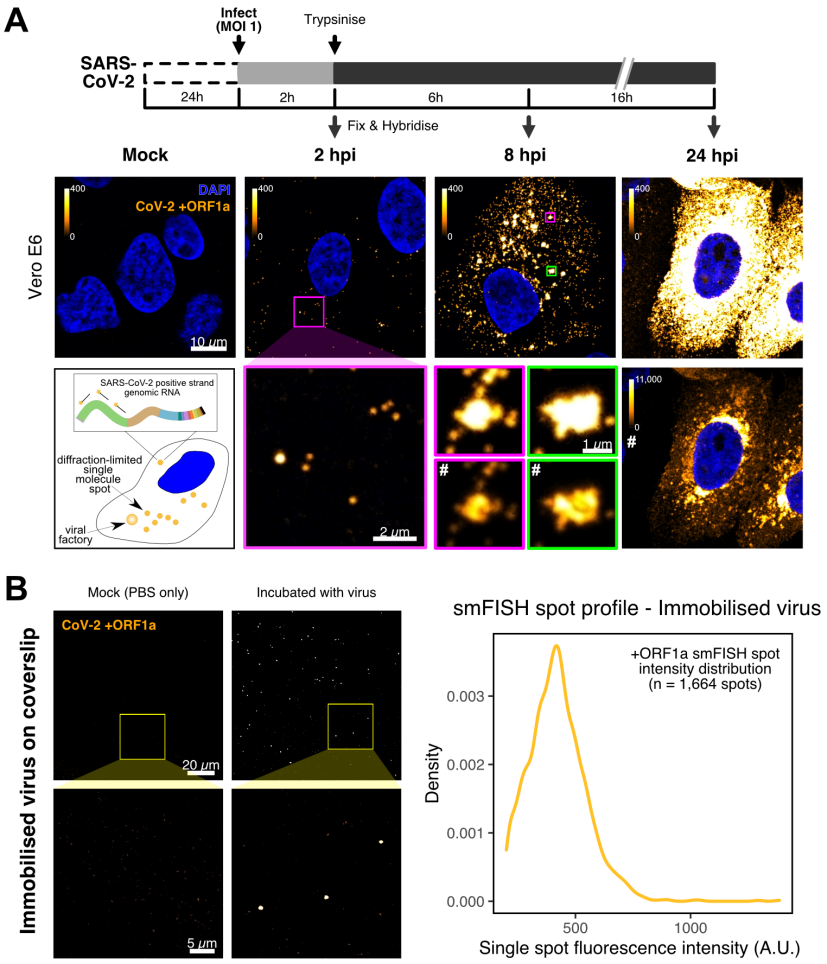


Figure 5.2: Detection of SARS-CoV-2 genomic RNA across the infection timecourse. (**A**) Experimental design for visualising SARS-CoV-2 gRNA with smFISH at different timepoints after infection of Vero E6 cells with SARS-CoV-2 VIC strain. Confocal images (4 µm maximum intensity projection) of the representative cells are shown. (n=2). The calibration bar indicates the intensity range (0-400). The image panels with hashtag (#) symbol are adjusted to show wider intensity range (0-11,000). (**B**) Visualisation of encapsidated viral RNA immobilised on a coverslip with smFISH. (n=3). The fluorescence intensity distribution of detected spots from encapsidated viral RNAs is plotted.

I further confirmed that the +ORF1a probe signal is of viral RNA origin by adding RNA-dependent RNA polymerase (RdRP) inhibitor remdesivir (RDV). The drug treatment resulted in a significant reduction of smFISH signals (Figure 5.3A). Treating fixed cells with a cocktail of RNaseT1 and RNaseIII (ssRNA and dsRNA substrate specificity, respectively) also ablated the signal, confirming the specific detection of RNA. To assess the efficiency of detection, I employed a well-established dual-colour method (Mueller et al., 2013; Raj et al., 2008; Ramanan et al., 2016). Here, the probe set is divided into two groups of 24 alternating oligonucleotides ('ODD'

and 'EVEN') (Figure 5.3B). The two groups were labelled with ATTO565 or ATTO633 and co-hybridised to detect the same transcript in two fluorescence channels. Due to the interlaced strategy, the chromatic aberration between the two channels is minimised. The average distance between the two channels for each spot was below 250 nm (Figure 5.3B), showing precise spatial registration within the diffraction limit of the microscope. Moreover, on average 95% of the diffraction-limited spots were labelled with both colours, demonstrating highly efficient detection of single viral gRNA molecules.

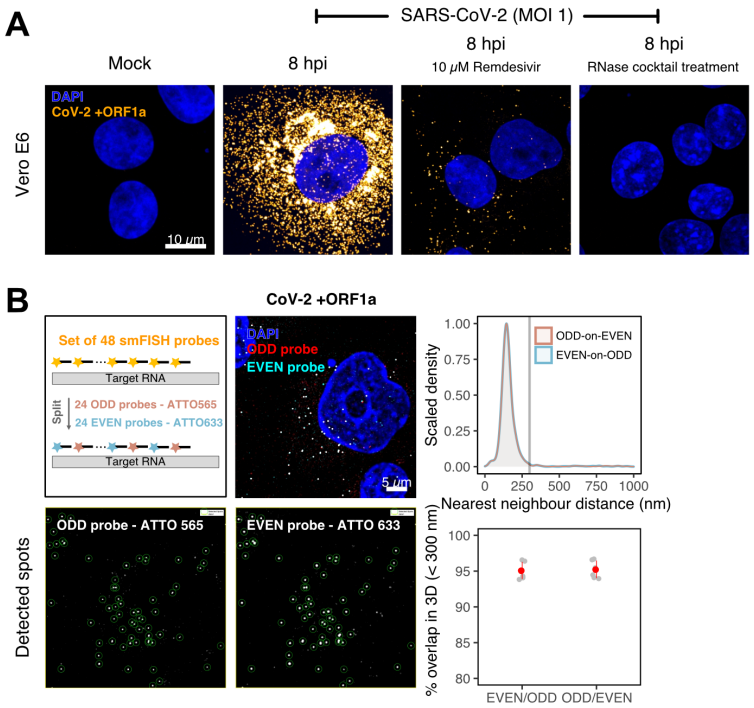


Figure 5.3: Sensitive detection of SARS-CoV-2 RNA. (**A**) The specificity of +ORF1a probes to SARS-CoV-2 RNA was tested in cells treated with Remdesivir or a cocktail of RNAseT1 and RNAseIII. Confocal images (4 µm maximum intensity projection) of representative cells are shown. (n=3). (**B**) Assessment of smFISH detection sensitivity by a dual-colour detection method. A maximum intensity projection of infected Vero E6 cells are shown with the FISH-quant graphical user interface displaying detected spots in 'ODD' and 'EVEN' probe channels. The density histogram of the nearest-neighbour distance from one fluorescent channel to another is plotted with the vertical line indicating 300 nm distance. The percentage of spots detected in both 'ODD' and 'EVEN' probe channels are plotted. (n=5).

5.2.2 Specific detection of SARS-CoV-2 RNA

It is possible that the +ORF1a probes can potentially cross-hybridise with RNAs derived from other related coronavirus species. The potential lack of specificity would be a particular problem with

clinical samples that might have been exposed to other circulating human coronaviruses (e.g. cold viruses). To verify the specificity of the +ORF1a probes to SARS-CoV-2, I computationally aligned probe sequences against other coronaviruses and the transcriptomes of human and African green monkeys using bowtie2 (Langmead & Salzberg, 2012) (see the Methods for details). The majority of the probe sequences showed significant mismatches (edit distances) with SARS-CoV-1, MERS-CoV, and other coronaviruses along with human and green monkey RNAs (Figure5.4A). The high mismatch rate of +ORF1a probe highlights its specificity to SARS-CoV-2 and suggests that the probes are unlikely to hybridise with RNAs of other coronaviruses. As a proof-of-principle, I evaluated this experimentally by incubating +ORF1a probes with Huh-7.5 hepatoma cells infected with the common cold virus, HCoV-229E. HCoV-229E-infected cells were provided by Louisa Iselin. As a positive control to identify heavily infected cells, I also stained infected cells with an antibody against dsRNA (J2 clone), which detects viral replication intermediates (Cortese et al., 2020; Kamel et al., 2021; Weber et al., 2006). As expected, J2-positive HCoV-229E infected cells were negative for +ORF1a probe-specific signal (Figure 5.4B). In contrast, J2-positive Calu-3 cells infected with SARS-CoV-2 produced bright smFISH signal with +ORF1a probes. In summary, these data demonstrate the high specificity of +ORF1a probes in detecting SARS-CoV-2 gRNA.

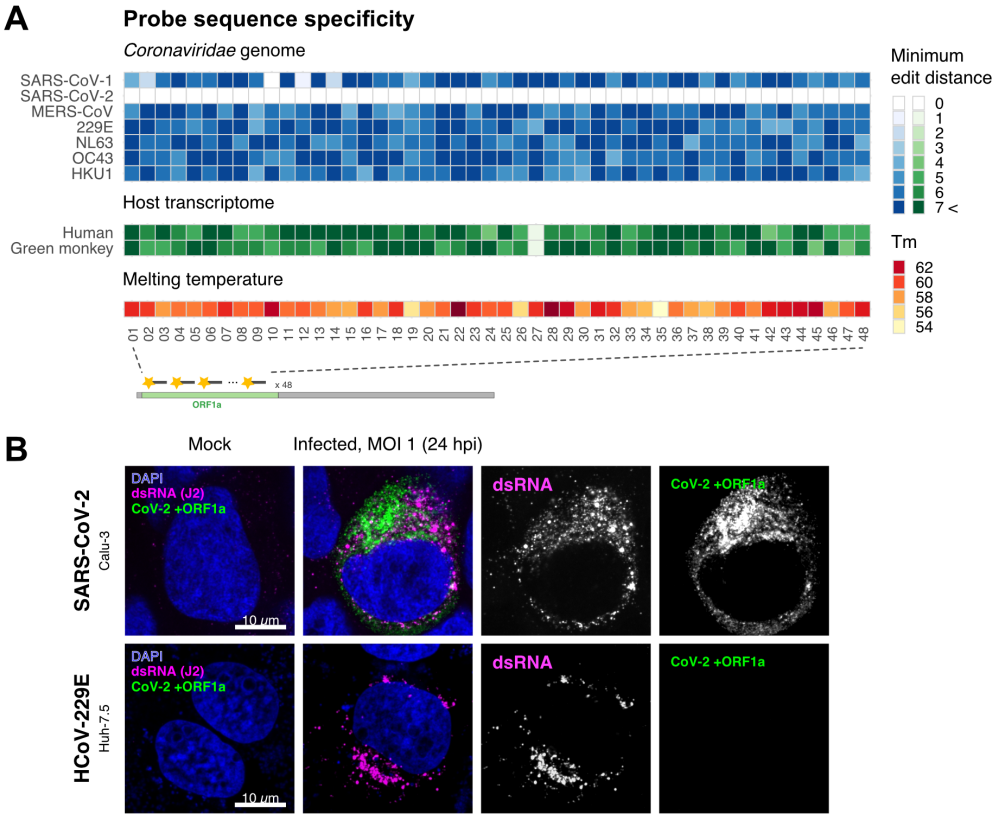


Figure 5.4: Specific detection of SARS-CoV-2 RNA. (**A**) A heatmap of +ORF1a probe sequence alignment against multiple coronavirus species and host transcriptomes. The minimum edit distance represents mismatch score (higher is better). (**B**) +ORF1a probe specificity was tested against SARS-CoV-2 (in Calu-3) and HCoV-229E (in Huh-7.5). J2 anti-dsRNA was used as a positive control to track infection of both virus species. Single slice confocal images are shown. (n=2).

5.2.3 SARS-CoV-2 detection in animal model samples

Having established the method in cell culture models, I next tested whether smFISH could be used to detect SARS-CoV-2 RNAs in formalin-fixed and paraffin-embedded (FFPE) lung tissue from experimentally infected Golden Syrian hamsters. I reasoned it would be essential to test the utility of the method in FFPE samples, as they represent the bulk of COVID-19 clinical and experimental specimens. The animal housing, infection and tissue preparations were performed by Dr Francisco Salguero's lab in the UK Health Security Agency (formerly Public Health England). The hamsters were inoculated with SARS-CoV-2 ($5X10^4$ plaque-forming unit [PFU]) by intranasal delivery and infection was assessed by qPCR measurement of viral RNA in throat swabs at four days post infection. The mean titre of infectious virus was $5.3X10^3$ PFU/ml (Dowall et al., 2021). Due to

the extended fixation period and dense tissue structure of FFPE samples, the lung sections were reverse-crosslinked in antigen retrieval solution for 1 h at 80°C, and tissue clearing was performed with 8% SDS (Rouhanifard et al., 2018) prior to probe hybridisation. The standard proteinase K step was deemed unsuitable in these samples due to excessive loss of tissue morphology after the enzymatic digestion. +ORF1a probes detected SARS-CoV-2 gRNA in a representative section of the infected lung tissue, showing variable intracellular RNA levels (Figure5.5). Typically, viral RNA detection utilises amplification-based *in situ* methods to enhance the signal and shorten imaging durations (Carossino et al., 2020; Lean et al., 2020; X. Liu et al., 2020). However, I found smFISH is a suitable alternative to these expensive methods as it offers a signal bright enough to scan the entire lung section in under 15-20 minutes. Moreover, there is an added benefit to zoom in to a user-defined region of interest for precise single-molecule analysis. This result demonstrates single-molecule detection of SARS-CoV-2 gRNA within infected tissue samples fixed and treated similarly to clinically archived tissues. I conclude that smFISH is likely to work well on materials derived from infected patients.

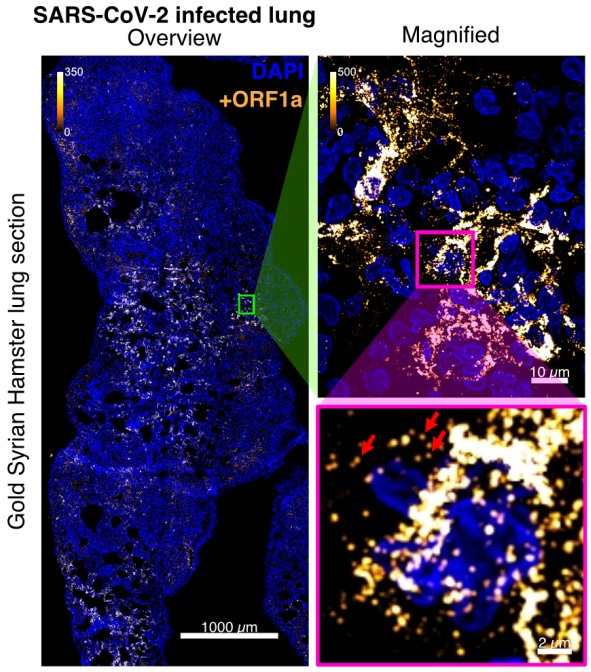


Figure 5.5: SARS-CoV-2 detection in infected hamster lungs. smFISH using +ORF1a probes on formalin-fixed paraffin-embedded (FFPE) lung tissue sourced from SARS-CoV-2-infected Golden Syrian hamster at 4 days post infection. The tissue specimen was pre-processed by Dr Francisco J Salguero. 20x objective image and a 100x zoom inset are shown (maximum intensity projection of 4 µm). A representative lung section from 4 replicate hybridisation experiments is shown.

5.2.4 dsRNA antibody underestimates SARS-CoV-2 replication

In the early stages of the virus infection, SARS-CoV-2 forms replication factories using host membrane structures (Eymieux et al., 2021). These factories include double-stranded viral replication intermediates. Previously, these replication structures have been described using J2 dsRNA antibody stains, which are commonly used to track and identify infected cells (Burgess & Mohr, 2015; Cortese et al., 2020; Targett-Adams et al., 2008; Weber et al., 2006). However, soon after virus entry, rare virus-derived dsRNA antigens are indistinguishable from endogenous dsRNAs naturally expressed in host cells (Dhir et al., 2018; Kimura et al., 2018). To systematically assess the ability of the J2 antibody to track SARS-CoV-2 replication in comparison to smFISH, I combined immunofluorescence and smFISH in the infected Vero E6 cells at 2 and 6 hpi. J2 antibody staining was performed first and subsequently fixed with 1% PFA to prevent the possibility of dsRNA denaturation by formamide in the hybridisation solution (Figure 5.6A). With

the J2 stain, I was not able to detect a virus-specific increase in signal at 2 hpi compared to the Mock condition. A 6 hpi, 10% of the infected cells displayed J2 fluorescence above the threshold. However, the average J2 signal intensity across the population of infected cells at both 2 and 6 hpi was comparable to that of Mock infected cells (Figure 5.6B). These results are in agreement with previous reports demonstrating a lack of virus-specific signals before 6 hpi (Cortese et al., 2020; Eymieux et al., 2021; Kamel et al., 2021). In contrast, smFISH was able to identify more than 85% of the cells containing diffraction-limited smFISH signals at both 2 and 6 hpi. Furthermore, it was able to detect a significant increase in the copy number of gRNA from 2-6 hpi. Therefore, this data demonstrates that the J2 antibody, although widely used, significantly underestimates the viral replication and the frequency of infection in the early stages.

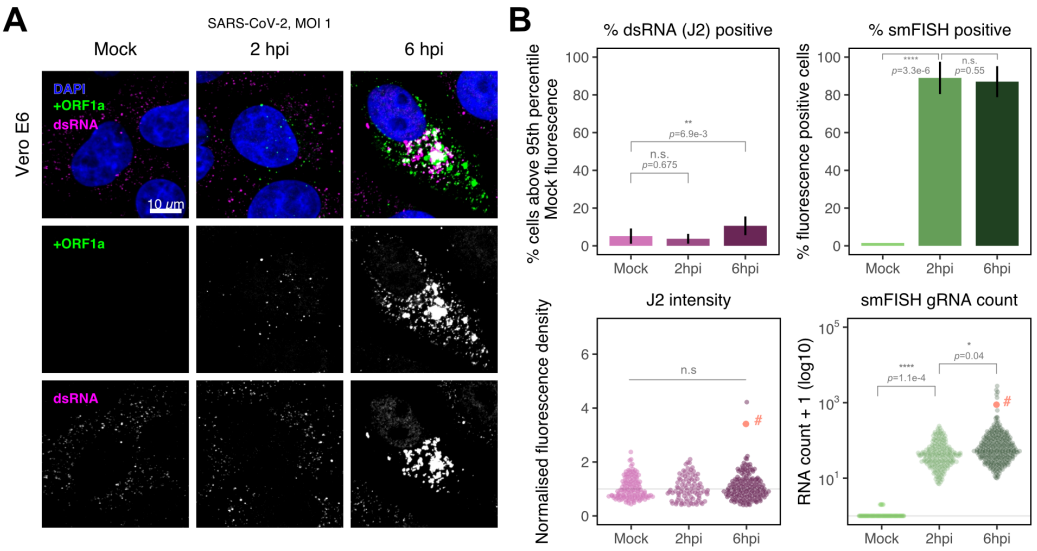


Figure 5.6: J2 anti-dsRNA underestimates SARS-CoV-2 replication. (**A**) Comparison of anti-dsRNA (J2) and +ORF1a smFISH. Full z-projected images are shown. (**B**) Percentage of infected cells detected with J2 or smFISH (*top panels*). The J2 positivity threshold was defined as the 95th percentile fluorescent signal of uninfected cells due to the presence of endogenous signals. Data are presented as mean±SD. Comparison of quantification results between J2 and smFISH (*bottom panels*). Each symbol represents one cell. J2 signal was quantified by integrated fluorescence density of cellular volume, normalised over Mock average. gRNA count represents the sum of single-molecule spots and decomposed spots within viral factories. The cell denoted with the '#' symbol is the infected cell shown in (A). (n=3). *p<0.05; **p<0.01; ****p<0.0001. see the Materials and methods Chapter for the details on statistics).

5.3 Co-visualising SARS-CoV-2 genomes and sub-genomes

In addition to the positive-sense gRNA, SARS-CoV-2 produces a plethora of RNA species, namely subgenomic (sg)RNAs and negative-sense RNAs. These RNA species are critical to express the full scope of viral proteins at the right time and stoichiometry. However, their detection is often challenging due to sequence overlap or its complementarity to gRNA. Therefore, I asked whether our smFISH strategy could be used to visualise and characterise the subcellular localisation of different viral RNA species.

5.3.1 Multiplexed smFISH strategy reliably discriminates gRNA and sgRNA

Quantification of sgRNAs is challenging due to their sequence overlap with the 3' end portion of the gRNA. Sequencing-based methods often rely on a minority of reads that span the junctions between the common 5' Leader and subgenomic open reading frames (ORFs). Therefore, I designed a nested smFISH strategy where two additional probe sets targeting the 3' subgenomic ORFs were synthesised with different fluorophores. First, the +ORFN probe set was designed to hybridise to all canonical positive sense viral RNAs. Second, the +ORFS probe set is complementary to Spike ORF, so it can detect sgRNA encoding S (S-sgRNA) and gRNA (Kim et al., 2020) (Figure5.7A). In this scheme, sgRNAs will show fluorescence only for the +ORFN or +ORFS probe but gRNA will be dual-labelled with the +ORF1a and sgRNA probes. When infected Vero E6 cells (6 hpi) were hybridised with +ORF1a and +ORFN probes, I was able to observe a substantial level of signal derived from the +ORFN probe compared to +ORF1a, indicating a considerably higher abundance of sgRNA over gRNA (Alexandersen et al., 2020; Kim et al., 2020) (Figure 5.7B). Using the single-molecule resolution and spot-colocalisation, the nested dual-colour smFISH method was also able to reliably distinguish gRNA and S-sgRNA, detected as dual-colour and single-colour spots, respectively (Figure 5.7B, *lower panels*).

As explored in the previous Chapter, subcellular localisation of RNAs can reveal insight into their roles in biological processes. Hence, I used the nested smFISH strategy to examine the subcellular localisations of gRNA and sgRNA. I extracted coordinates of RNA spots at 6 hpi and

related them to nuclear and plasma membrane borders using the RNA Distribution Index calculator (Stueland et al., 2019). Infected cells with spatially saturated smFISH signals were excluded from the analysis to prevent biases. The analysis revealed that the +ORFN single-labelled sgRNAs were more uniformly dispersed throughout the cytoplasm than dual-labelled gRNA, consistent with their predominant role as mRNAs to direct protein synthesis (Figure 5.7C). However, gRNAs were enriched near the perinuclear region often in aggregated puncta, which may represent their localisation in membranous replication organelles (Klein et al., 2020).

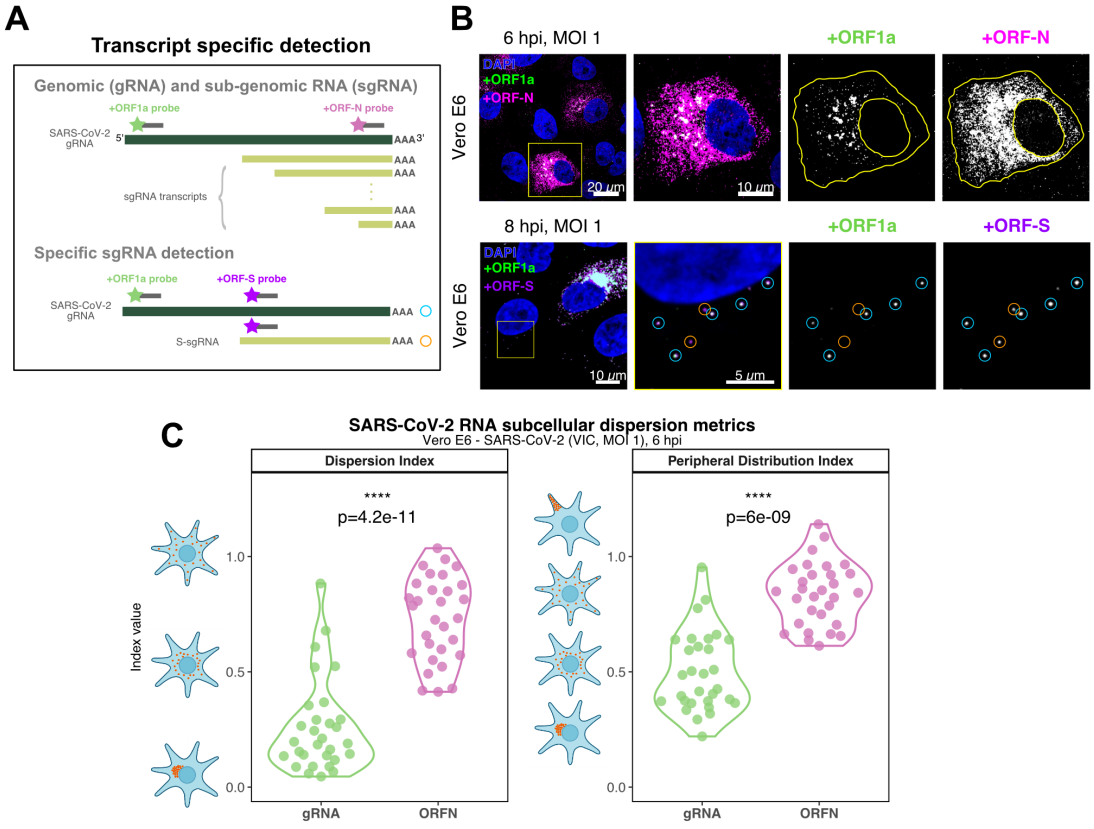


Figure 5.7: Multiplexed detection of SARS-CoV-2 gRNA and sgRNA. (**A**) Schematics of transcript-specific targeting of SARS-CoV-2 gRNA and sgRNA using a nested smFISH strategy. (**B**) Discrimination of gRNA and sgRNA in infected (VIC strain) Vero E6 cells. Infected cells were hybridised with probes against +ORF1a and +ORFN at 6 hpi (*top panels*) or +ORF1a and +ORFS at 8 hpi (*bottom panels*). Representative 3 µm maximum intensity projected confocal images are shown. Orange circular regions of interest (ROIs) indicate S-sgRNA encoding Spike, whereas dual-colour spots (teal-coloured ROIs) represent gRNA. (n=2). (**C**) RNA dispersion and peripheral distribution indices were calculated using the RNA distribution index (RDI) calculator on +ORF1a and +ORFN channels. Schematic diagrams of subcellular RNA distribution of select indices are shown next to the corresponding plots. ****p<0.0001 (n=3, five field of views per replicate were analysed).

5.3.2 gRNA and sgRNA differentially associate with the viral nucleocapsid protein

The association of gRNA with the viral nucleocapsid (N) protein is essential for the assembly of coronavirus particles, where the liquid-liquid phase separation of N protein and gRNA drives the virion formation (Carlson et al., 2020; Dinesh et al., 2020; Iserman, Roden, Boerneke, Sealfon, McLaughlin, Jungreis, Fritch, et al., 2020). However, the extent of N protein-sgRNA associations is not known. To assess this, I combined the nested smFISH strategy with immunofluorescent detection of viral N protein (Figure 5.8A-B). Single-molecule coordinates of gRNA and sgRNAs were extracted and the colocalised anti-N intensity was analysed in infected Vero E6 cells (8 hpi). Cells with spatially saturated viral RNA levels were excluded from the analysis to avoid artefactual colocalisation by chance. Colocalisation thresholds were determined by randomly positioning the same density of spots 10 times for each image and sampling 2 standard deviation values of 'random' anti-N fluorescence intensity. I found that gRNA spots displayed a bimodal N fluorescence intensity distribution, consistent with the existence of two populations of gRNAs: one that is N-associated and the one "free" (Figure 5.8C). Furthermore, the data showed that N protein primarily co-localises with gRNA, while displaying a limited overlap with sgRNAs (Figure 5.8D). Together, these data demonstrate the specificity of the probe sets to discriminate between the gRNA and sgRNAs, and the utility of the nested smFISH strategy to study viral RNA localisations as well as their interaction with other factors.

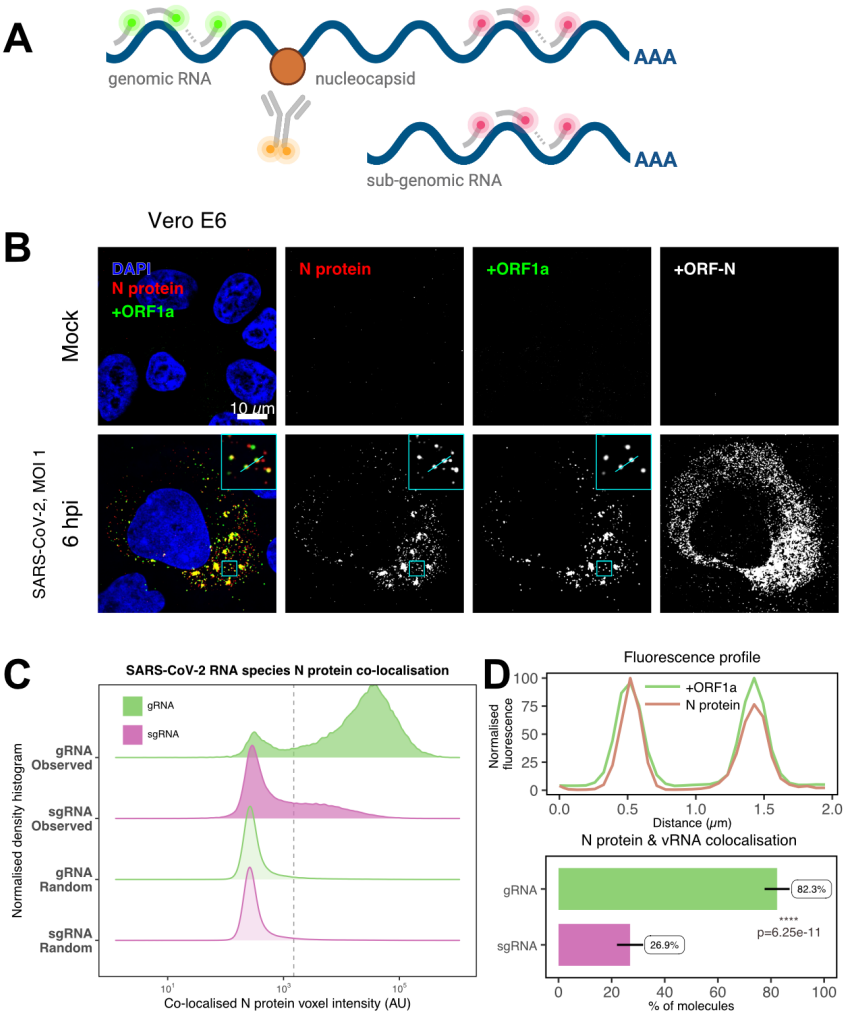


Figure 5.8: Differential association of gRNA and sgRNA with viral Nucleocapsid. (**A**) Schematics of combined smFISH and immunofluorescence detections of the viral RNA and proteins. (**B**) Co-detection of viral nucleocapsid (N) with gRNA and sgRNA. Representative 3 µm z-projected confocal images are shown. The inset shows a magnified view of co-localised N and gRNA. (**C**) Density histogram of N protein voxel intensities co-localised to gRNA or sgRNA spots. A threshold was determined using a 'random' simulation to place the same density of RNA spots within infected cells and calculating the voxel intensities that correspond to chance co-localisation. The vertical dotted line corresponds to the threshold of co-detection defined as 2 standard deviation values from the random distribution analysis (n=7 field of views). (**D**) Fluorescence line profiles of anti-N and gRNA smFISH intensity across a 2 µm distance shown in the magnified inset of (B, cyan box) (*top*). Percentages of co-localised gRNA or sgRNA molecules with N protein at 6 hpi. (n=7) (*bottom*). ****p<0.0001.

5.3.3 Negative-strand gRNA requires denaturation for proper detection

Positive-sense RNA viruses also produce complementary negative-strand (antisense) gRNA and sgRNAs that act as templates for the synthesis of positive-sense RNAs (Kim et al., 2020). These

antisense RNAs are thought to localise to viral replication factories as replication intermediates. However, their detection is often challenging with RT-qPCR or RNA-seq protocols using oligo(dT) selection methods due to the primer binding to dsRNA structures and the lack of polyA tail in these species (Ramanan et al., 2016; Sethna et al., 1991). Exploiting the direct hybridisation nature of smFISH, I decided to use this technique to detect SARS-CoV-2 negative gRNA. Surprisingly, my initial attempts at negative-strand smFISH failed to detect diffraction limited spots, and designing probes against other regions of the -gRNA did not yield improvement. I reasoned that the tight dsRNA structure of gRNA hybrid could make the probe inaccessible for hybridisation. To address this, I denatured dsRNA complexes using either formamide, DMSO, or sodium hydroxide treatment of fixed infected cells (Singer et al., 2021; Wilcox et al., 2019). I assessed the extent of denaturation using a J2 dsRNA antibody stain (Figure 5.9A). I found the combination of DMSO with heat treatment at 80°C resulted in a significant reduction of J2 signal, with minimal effect on cell morphology. Formamide (70%) and heat treatment at 80°C was also able to disrupt dsRNA hybrids but it caused excessive cell detachment from the coverslip. The use of NaOH has been previously suggested for the negative strand detection of Sindbis virus (Singer et al., 2021), however, its effect on dsRNA hybrids of SARS-CoV-2 was negligible. Using the DMSO and heat treatment, smFISH probe specific for the -ORF1b antisense sequence was able to produce intense diffraction-limited spots in the infected Vero E6 cells (Figure 5.9B). These -gRNA spots were detected at a much lower level than the positive-sense counterpart, consistent with its low abundance from RNA-sequencing reports (Kim et al., 2020; Thorne et al., 2021). Notably, -gRNA spots displayed a substantial overlap with the +ORF1a signal, especially within the perinuclear region in a clustered fashion. This result is consistent with active sites of viral replication in membranous structures derived from the ER.

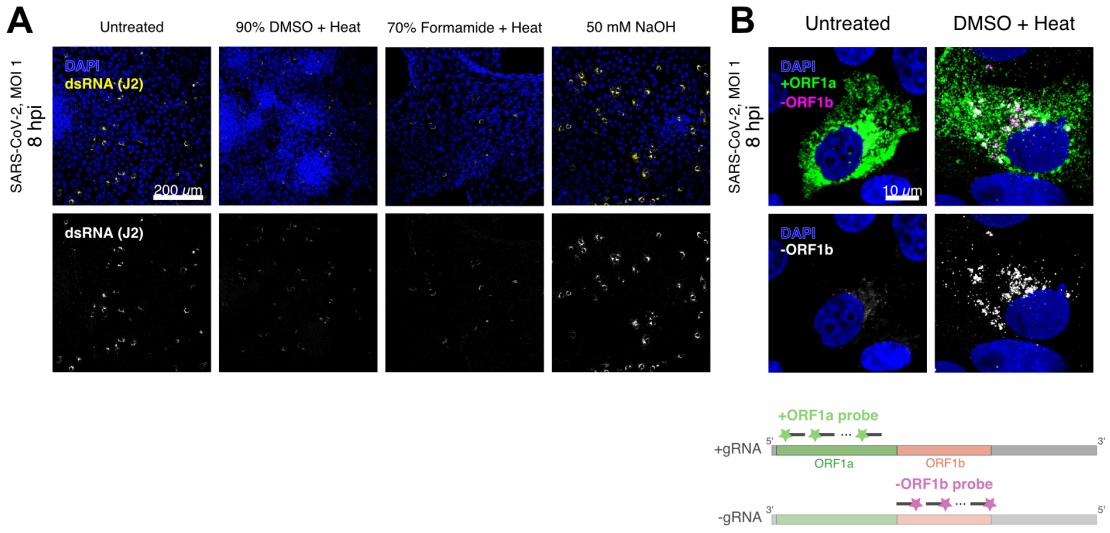


Figure 5.9: SARS-CoV-2 negative-strand detection requires chemical denaturation. (**A**) Comparison of dsRNA denaturation efficiency assessed by the reduction of anti-dsRNA (J2) immunofluorescence. Fixed cells were treated with DMSO, formamide, or NaOH prior to immunostaining. Representative low-magnification single-slice confocal images are shown. (n=4). (**B**) Detection of both positive and negative genomic RNA by denaturing viral dsRNA with DMSO and heat treatment at 80°C (*top panels*). 3 µm z-projected confocal images are shown. (n=4). Schematic illustration of +ORF1a and -ORF1b probe targeting regions (*bottom panel*). The -ORF1b probe target region does not overlap with the +ORF1a target sequences to prevent probe duplex formation.

I further confirmed the specificity of my probes against the dsRNA hybrids by treating the fixed cells with RNaseT1 or RNaseIII, which are nucleases specific to single-stranded RNA (ssRNA) or dsRNA, respectively (Figure 5.10A). As expected, RNaseT1 reduced the +ORF1a signal while RNaseIII diminished the J2 signal. However, to fully remove the +ORF1a signal, a cocktail of RNaseT1 and RNaseIII was required, which suggests that the +ORF1a probe set can hybridise to both single and double stranded RNA in my experimental conditions (Figure 5.10A). Additionally, when the cells were treated with DMSO and heat prior to the RNaseT1 digestion, both +ORF1a and -ORF1b signals were removed. This indicates that disrupting the dsRNA hybrid with DMSO makes it accessible for RNaseT1 degradation (Figure 5.10B). Therefore, these results demonstrate that hybridisation to negative-strand requires chemical disruption of viral dsRNA.

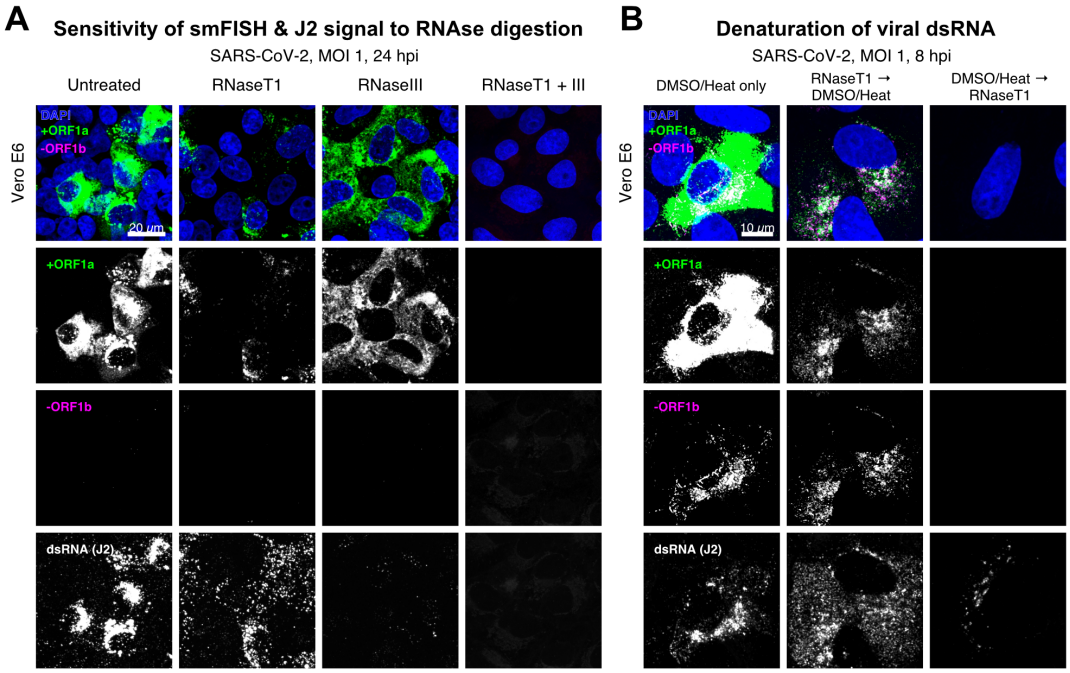


Figure 5.10: Assessment of probe specificity viral ssRNA and dsRNA. (**A**) Sensitivity of the +ORF1a probe and the J2 antibody signal to RNase digestions. Fixed infected cells (24 hpi) were treated with RNaseT1 and/or RNaseIII to digest single-stranded RNA and/or dsRNA, respectively. Representative full z-projected confocal images are shown, which are single-molecule contrast matched. (n=3). (**B**) RNaseT1 digestion of denatured dsRNA. Fixed infected cells (8 hpi) were treated as follows: (i) DMSO/heat only (*left panels*); (ii) RNaseT1 then DMSO/heat (*centre panels*); or (iii) in the reverse order of DMSO/heat and then RNaseT1 (*right panels*). Treated cells were hybridised with +ORF1a and -ORF1b probes and stained with J2. Full z-projected confocal images are shown. (n=2).

5.4 Profiling early-stage SARS-CoV-2 replication kinetics

Observation of viral RNA replication, accumulation and localisation to viral factories provide insight into how the virus establishes infection and differs between mutant strains. Therefore, I aimed to profile the expression kinetics of both gRNA and sgRNA at the single-molecule and cell levels, providing unprecedented insight into viral RNA replication, accumulation and localisation to viral factories.

5.4.1 Developing an analysis pipeline to automate quantification of smFISH datasets

One of the main technical highlights of smFISH is that individual transcripts appear as diffraction-limited spots of uniform shape and fluorescence intensity. This allows accurate transcript counting and quantification of viral factories whose fluorescence intensity can be decomposed into single

RNA units (Mueller et al., 2013; Raj et al., 2008). However, exploiting this feature presents two challenges: i) properly sampled images (sufficiently small pixel size) have a large file size (~200 GB per replicate) that require extensive computational resources and ii) the analysis tool must be able to take advantage of the point-spread function (PSF) approximation of each spot to model what a representative single-molecule transcript would look like. Sufficient sampling is particularly important so that each detected RNA spots are appropriately 3D Gaussian approximated, and ideally the pixel size should be set to be smaller than the 50% of the optical resolution to ensure sufficient image sampling (i.e. Nyquist sampling) (S. Wang, 2019). Most smFISH/FISH quantification tools rely on spot-detection algorithms with simple fluorescence intensity cut-offs or provide limited modelisation of smFISH spots to 3D Gaussian shapes. FISH-quant, written in MATLAB, provides a wealth of tools for smFISH quantification with PSF approximation capabilities. However, the program is mainly handled via a graphical user interface (GUI) and performs poorly when batch processing large datasets (Mueller et al., 2013).

Recently, the developers of FISH-quant deployed BigFISH, a python-implementation of FISH-quant, offering faster processing and compatibility with existing python image analysis libraries. Therefore, in collaboration with Dr Joshua Titlow and Danail Stoychev, we developed an automated smFISH analysis pipeline. We seamlessly integrated Cellpose (a machine learning-based cell segmentation python package) with BigFISH to analyse large smFISH datasets with multi-core processing capabilities. A schematic of the BigFISH pipeline is presented in Figure 5.11. The workflow is as follows: (1) the hybridised cells are imaged automatically using Olympus spinning-disk confocal microscope, where Z-Drift correction was aided by forcing refractive index (RI) mismatch by mounting cells in diluted mounting media; (2) the cell boundaries are segmented using Cellpose with CellMask, which stains lysine residues of cellular proteins; (3) candidate single-molecules and viral factories are detected above a fluorescence threshold set with Mock infected samples; (4) 3D Gaussians are fitted to the detected spots to obtain a reference single-molecule image; (5) viral factories (larger and brighter spots) are detected and resolved to enumerate RNA content within each factory. The pipeline is able to handle large smFISH datasets, which takes 3-4 h of hands-off processing time for a ~200 GB dataset. The output

consists of the number and coordinates of cytoplasmic transcripts and the number, coordinates and RNA composition of viral factories for each cell. The pipeline also outputs images of reference single-molecules for quality control purposes. The BigFISH automated analysis pipeline is also applicable to any other smFISH datasets and requires minimal knowledge in Python. The full scripts are available in this GitHub repository (`https://github.com/jefflee1103/Lee_Wing-SARS2/tree/main/smFISH/analaysis-codes`).

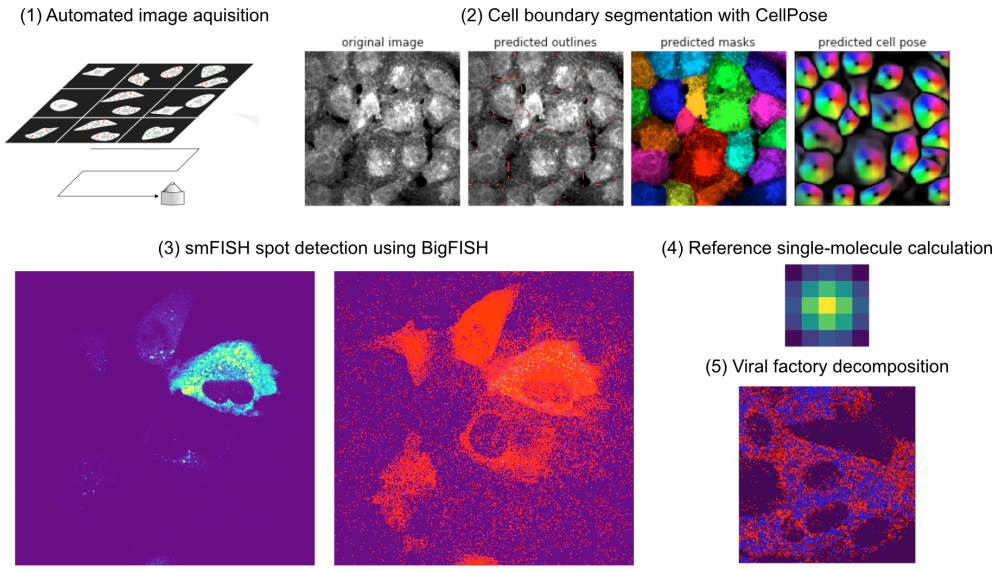


Figure 5.11: BigFISH automated smFISH analysis pipeline. Schematics of an automated computational pipeline to analyse large smFISH datasets. (1) Images are automatically acquired; (2) Cell boundaries are segmented using CellPose python library; (3) smFISH spots are pre-detected using BigFISH, the left panel shows pseudo-colour smFISH channel and the right panel shows detected spots in red; (4) Reference single-molecule spot is calculated from the pre-detected spots; (5) Viral factories are clustered and decomposed using the reference single-molecule as a unit.

5.4.2 SARS-CoV-2 replication is heterogeneous

Having established the experimental and computational methods to quantify single molecules of SARS-CoV-2, I decided to profile high-resolution viral replication dynamics in the first 10 h of the infection. For this kinetics assay, Vero E6 cells were infected with SARS-CoV-2 (MOI 1) with or without the replication inhibitor RDV. Fixed cells were then hybridised with +ORF1a and +ORFN probes to investigate the replication kinetics of both gRNA and sgRNA (Figure 5.12A-B). At 2 hpi, the +ORF1a probe set detected mainly single molecules of +gRNA with

a median value of ~30 molecules per cell. As discussed previously, the virus stock contains many defective particles (Figure 5.2A). Considering the MOI of one used in this experiment, the particle-to-infectivity ratio of my virus stock is therefore likely to be above 30. The RDV treatment had a negligible effect on the distribution of gRNA content between cells, indicating that most of these RNAs are of incoming virus origin (Figure 5.12C). Beyond 4 hpi, I found the RNA replication occurs in non-homogeneous fashion for each cell, despite the infection protocol maximising synchrony. Between the hours 4-10 hpi I was able to classify the infected cell population into three groups based on their varying gRNA burden: (i) 'partially resistant' cells with $<10^2$ gRNA copies that showed minimal increase in gRNA burden between 2 and 10 hpi (60% of the population); (ii) 'permissive' cells with 10^2–10^5 copies per cell showing a modest increase over time (~30%); and (iii) 'super-permissive' cells with $>10^5$ copies per cell showing a sharp increase in gRNA copies (~10%). The increase in gRNA was significantly inhibited by the RDV treatment, indicating active viral replication.

Note that due to the high spatial saturation of smFISH spots in 'super-permissive' cells, RNA counts could not be quantified using the spot counting method. Instead, the RNA counts were estimated by correlating the fluorescence intensity of a single molecule to that of total cellular volume. To confirm the veracity of this method, I plotted a calibration curve comparing RNA counts quantified via intensity the method versus the spot counting method (Figure 5.12D). Reassuringly, the analysis showed a linear relationship between the two methods, suggesting integrated intensity analysis is a suitable alternative to enumerate RNA counts in 'super-permissive' cells.

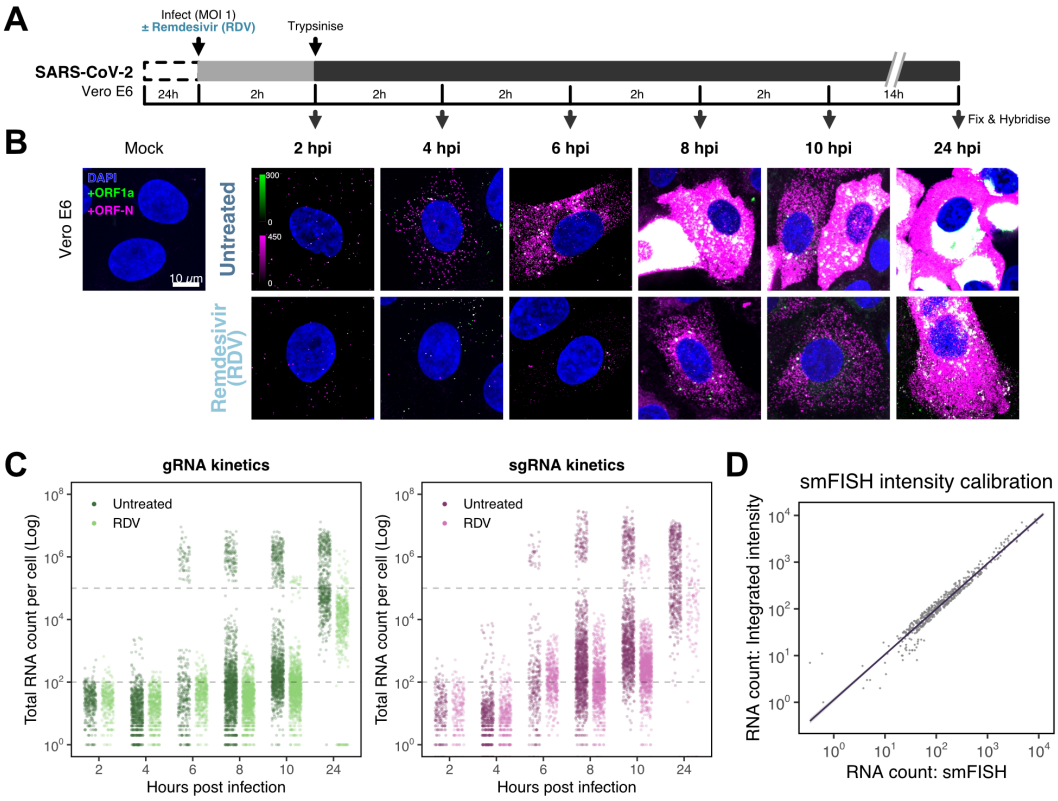


Figure 5.12: Replication kinetics of SARS-CoV-2 in Vero E6 cells. (**A**) Experimental design to profile SARS-CoV-2 replication kinetics using smFISH. Vero E6 cells were infected with SARS-CoV-2 (MOI 1, VIC strain) and incubated for the designated times post infection. In the RDV condition, the drug was added to cells at 10 μM together with the virus. (**B**) Maximum z-projected confocal images of infected cells. Intensity calibration bars are shown for +ORF1a in green and +ORFN in magenta. Representative 'super-permissive' cells are shown where applicable. (**C**) Bigfish quantification of +gRNA or +sgRNA counts per cell. sgRNA counts were calculated by subtracting +ORF1a counts from +ORFN counts per cell. Horizontal dashed lines indicate the 10^2 or 10^5 molecule marks. 24 hr post infection (hpi) samples and the cells harbouring $>10^6$ RNA counts were quantified by extrapolating single-molecule intensity. Quantified cells from all replicates are plotted (2–8 hpi, n=4; 10 and 24 hpi, n=3). (**D**) Correlation of smFISH RNA counts and cellular smFISH fluorescence intensity per cell. A subset of +ORF1a channels from non-'super-permissive' cells were used for the analysis. RNA counts were quantified using Bigfish (absolute RNA count) or via estimation using integrated intensity. A linear correlation was fitted using geom_smooth() function of 'ggplot2' R package (n=383 cells).

Next, I simulated an ensemble of total viral RNA (vRNA) content in the culture from the single-cell smFISH data. This analysis discovered that 'super-permissive' cells are the dominant source of gRNA from the culture (Figure 5.13A). This suggests that bulk analysis methods such as RT-qPCR are biased towards this high gRNA group. Consequently, the bulk RNA method may not be able to distinguish between the enhanced viral replication versus the increased frequency of 'super-permissive' cells. Interestingly, I found that this replication heterogeneity persists beyond

the initial hours of infection. Even at 24 hpi, 40% of the cells did not reach the 'super-permissive' state, and they formed a distinct population with approximately 10-fold less gRNA (Figure 5.13B and 5.12C). Therefore, these results highlight a wide heterogeneity in cell susceptibility to support SARS-CoV-2 replication present in both early and late stages of the infection.

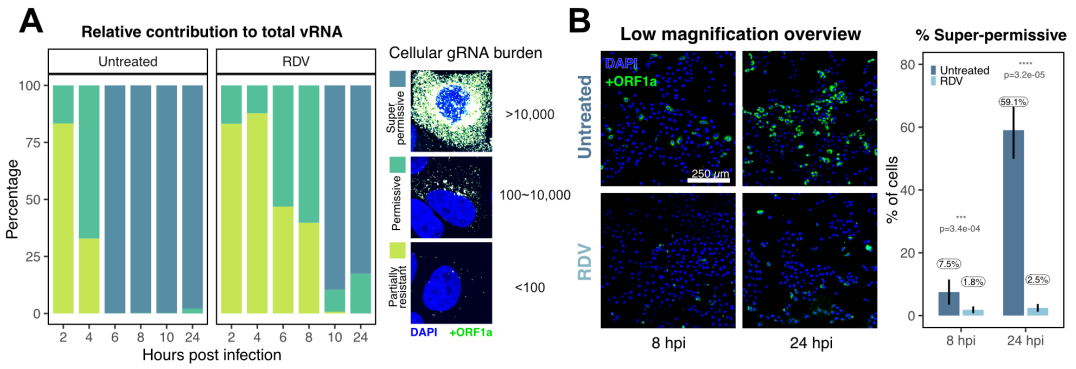


Figure 5.13: Heterogeneous replication of SARS-CoV-2 RNA. (**A**) The relative contribution of viral gRNA within the infected culture population. The infected cells were classified into three groups based on intracelullar gRNA counts: (i) 'partially resistant'=gRNA<100; 'permissive'=100<gRNA<10^5; 'super-permissive'=gRNA>10^5. The total gRNA within the culture population was obtained by summing gRNA counts in all cells, and the figure shows the relative fraction from each classification. Representative max-projected images of cells in each category are shown. (**B**) Identification of super-permissive cells from a low-magnification (×20) high-throughput smFISH assay (*left panels*). Full z-projected (9 µm) images of infected Vero E6 cells hybridised with +ORF1a probes are shown. At this magnification, only the 'super-permissive' cells display sufficient fluorescence above the background. Percentage of super-permissive cells in untreated and RDV-treated conditions at 8 and 24 hpi (*right panel*). Labels represent average values. Data are represented as mean±SD (n=3). ***p<0.001; ****p<0.0001.

SARS-CoV-2 utilises membranes derived from host ER and ERGIC to generate viral factories, which offers a protected compartment for RNA replication and virion assembly (Wolff et al., 2020). Despite their essential role, the genesis and dynamics of these factories are poorly explored. Therefore, I exploited the spatial resolution of smFISH to study these structures. I defined viral factories as spatially extended smFISH foci containing multiple gRNA molecules that are compatible, in terms of their diameter and subcellular location, with the double membrane vesicles (DMVs), which have been previously described using EM (Figure 5.14A) (Cortese et al., 2020; Mendonca et al., 2021). A 150 nm diameter threshold was used before 8 hpi and 200 nm beyond 8 hpi. Therefore, I refer to these bright foci as 'factories' capable of replicating RNA and assembling viral particles. Using the kinetics assay, I observed 1–2 factories per cell at 2 hpi, which rapidly

expanded to ~30 factories/cell by 10 hpi (Figure 5.14B). Note that 'super-permissive' cells are excluded from this analysis due to a lack of spatial separation between viral factories, therefore these quantification values are likely to be underestimated at 6-10 hpi. Using the reference single-molecule I was also able to quantify the number of gRNA within these bright foci by decomposing their structures based on their 3D shape and fluorescence intensity. The average number of gRNA molecules within these factories, although variable, increased over time. RDV treatment reduced both the number of viral factories per cell and their RNA content (Figure 5.14C). Together, these data show the capability of smFISH to provide measurements of active SARS-CoV-2 replication, providing a highly sensitive platform to test drugs or compare mutant viruses.

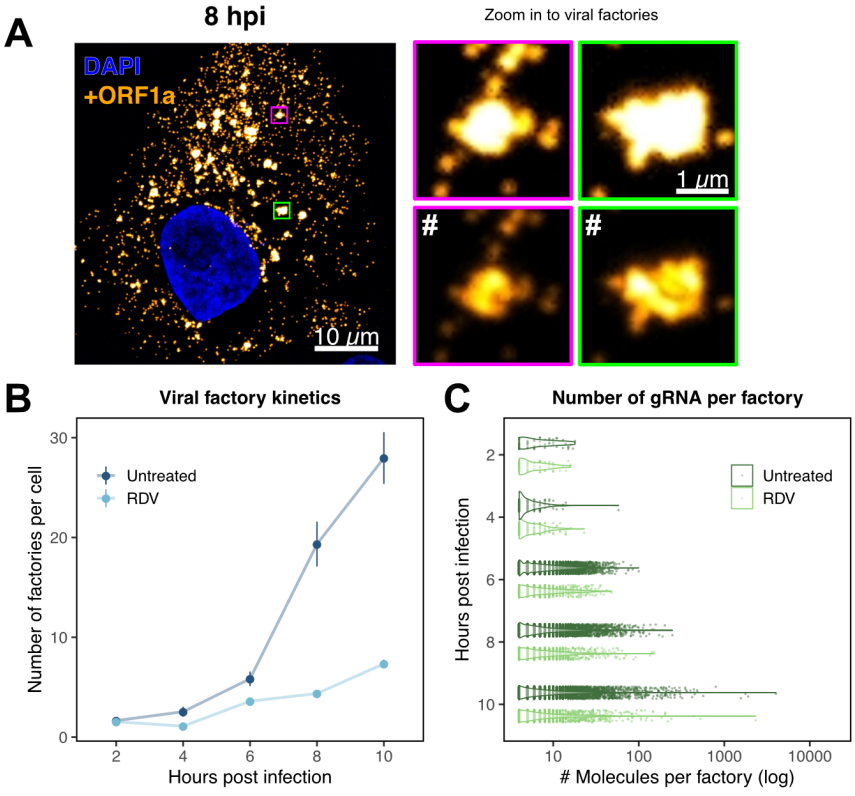


Figure 5.14: Replication kinetics of SARS-CoV-2 viral factories. (**A**) An example of a SARS-CoV-2 (VIC strain) infected Vero E6 cell at 8 hpi hybridised with +ORF1a probes. Note that these image panels are duplicates of *Figure 5.2A*, presented again for illustration purpose only. Two viral factories are highlighted with magenta and green insets. Insets denoted with '#' show a wider contrast range to highlight the 3D shape of the factories. (**B**) The number of viral factories per cell increases over time as assessed by smFISH cluster detection. Cells harbouring $>10^7$ copies of RNA, less than 10 molecules of RNA, cells with no viral factories, and cells from 24 hpi timepoints were excluded from this analysis. Data are presented as mean±SEM. (n=3). (**C**) The kinetics of gRNA copies within viral factories. Spatially extended viral factories were resolved by cluster decomposition to obtain single-molecule counts within these structures. (n=3).

5.4.3 Relative level of gRNA/sgRNA changes with the infection progression

Next, I assessed the replication kinetics of sgRNA via simultaneous analysis of +ORF1a and +ORFN probe channels. The smFISH data revealed similar kinetics and heterogeneity of sgRNA replication but at higher abundance compared to gRNA. Interestingly, already at 2 hpi, 63% of the infected cells expressed on average 11 copies of sgRNA (Figure 5.15A). Because the synthesis of +sgRNA requires replication of its complementary -sgRNA, this result suggests multiple rounds of replication occur rapidly following virus entry. Surprisingly, this process was insensitive to RDV treatment. By 6 hpi, most cells in the culture population contained sgRNA indicating active viral replication across the culture. I also examined the relative levels of sgRNA and gRNA, which has been shown to differ between patients, disease severity and infection stage (Alexandersen et al., 2020; Thorne et al., 2021). smFISH analysis revealed that the ratio of sgRNA/gRNA ranged from 0.5 to 8 over time, consistent with the dynamic range report from diagnostic samples (Alexandersen et al., 2020) (Figure 5.15B). Notably, the sgRNA/gRNA ratio followed a particular peak pattern, where it increased between 2 and 10 hpi but then started to decline towards 24 hpi, suggesting a clear shift in preference to produce gRNA over sgRNA at later stages of the infection. Consistently, sgRNA synthesis was favoured in 'partially resistant' and 'permissive' cells, whereas 'super-permissive' cells exhibited lower sgRNA/gRNA ratios (Figure 5.15C). A similar pattern was observed with the RDV-treated cells albeit with the lower sgRNA/gRNA values. In conclusion, these results indicate that gRNA synthesis is favoured in the late phase of infection, which may reflect a shift toward gRNA production to meet the demand for virion assembly and egress.

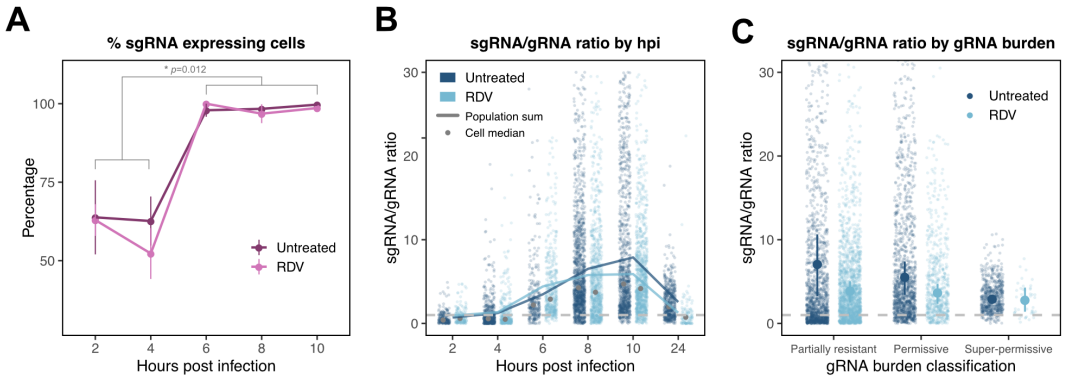


Figure 5.15: Dynamic changes of SARS-CoV-2 gRNA and sgRNA levels. (**A**) Percentage of infected cells expressing sgRNA. sgRNA-expressing cells were identified by those having an (ORF-N – ORF1a) probe count of more than 1. Data are represented as mean±SEM. (n=3). (**B**) Per cell ratio of sgRNA/gRNA ratio across the timeseries. Grey symbols represent cell-to-cell median values, whereas the line plot represents the ratio calculated from the population sum of gRNA and sgRNA. The horizontal dashed line represents value of 1. (n=3). (**C**) Per cell ratio of sgRNA/gRNA counts grouped by gRNA burden classification ('partially resistant', 'permissive', 'super-permissive'). Data are represented as median±SEM. The horizontal dashed line represents value of 1. (n=3).

5.4.4 SARS-CoV-2 gRNA is highly stable

Viral RNA stability can influence the ability of the virus to initiate and maintain a productive infection. In the previous kinetics assay, I found an incredible persistence of gRNA in the cells treated with 10 µM RDV in the first 8 hours of infection (Figure 5.12C). This observation suggests the incoming viral gRNA is long-lived. To explore the stability of SARS-CoV-2 gRNA from a different angle, I designed an experiment where RDV was added to infected cells at 24 hpi, and the changes in gRNA abundance were chased for 24 h thereafter (Figure 5.16A). In this strategy, RDV acts in a similar capacity as actinomycin D on eukaryotic RNA polymerase to model RNA half-lives (Lam et al., 2001). smFISH revealed RDV-treatment in this regime was able to reduce the proportion of 'super-permissive' and non-viable (condensed nuclei) cells at 48 hpi, suggesting the expected mechanism of action of the drug to inhibit viral replication and virus-induced cytopathy (Figure 5.16B). However, even after 24 h of RDV treatment, a considerable level of gRNA was observed in these cells, indicating that gRNA is also stable in the late stages of the infection (Figure 5.16C). To estimate the half-life of the gRNA from these measurements, I fitted a decay curve using the last three timepoints (28, 32 and 48 hpi). The approximation yielded an unusually

long half-life of 6–8 h (Figure 5.16C). I expect this half-life estimate to be an underestimate as gRNA loss can occur not only by decay but also via virus egress. I also note that I cannot rule out the possibility that the incomplete inhibition by RDV could affect my estimations, however, the 10 μM dosage used in this assay exceeded the 90% inhibitory concentration (Figure 5.16D).

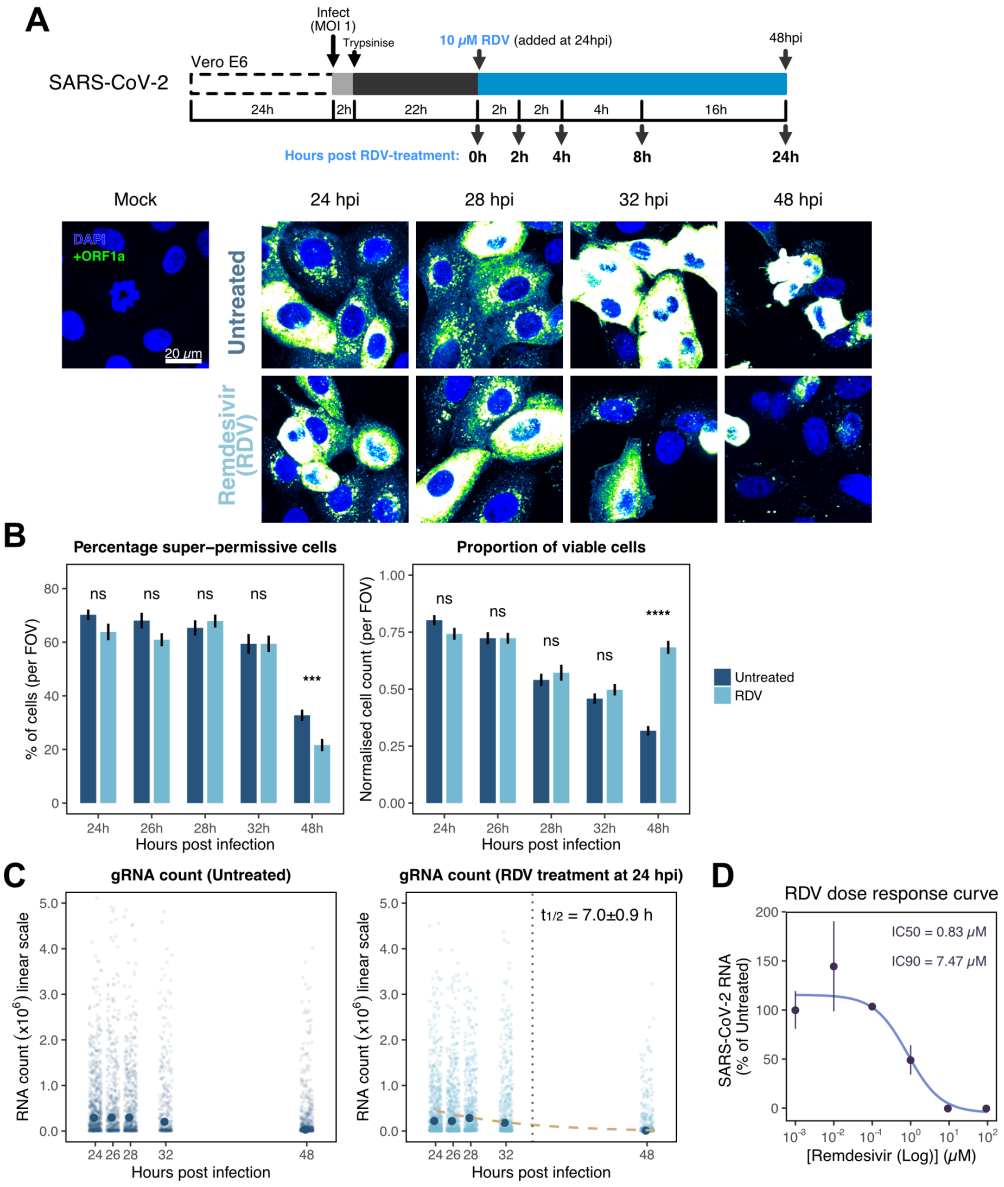


Figure 5.16: Analysis of SARS-CoV-2 gRNA stability. (**A**) Experimental design to profile SARS-CoV-2 gRNA stability in the late-stage infection. Vero E6 cells were infected with SARS-CoV-2 (VIC strain) for 24-48 hpi. In the RDV condition, the drug (10 μM) was added to cells at 24 hpi and maintained for the times shown. Representative full z-projected confocal images of infected cells from the time series. (**B**) The proportion of 'super-permissive' and viable cells per field of view (FOV) across the time series. Viable cells were quantified by counting non-condensed nuclei in randomly sampled FOVs. Non-condensed nuclei counts were normalised to the average of the 'Mock' condition. ***p<0.001; ****p<0.0001. Data are presented as mean±SEM. (n=2). (**C**) Quantification of viral gRNA counts in untreated and RDV-treated cells. Viral gRNA counts were quantified by extrapolating single-molecule intensity to the total fluorescence intensity of the cell. The decay curve was fitted with the median RNA count values using the last three timepoints. Data are presented as median±SEM. (n=2). (**D**) Dose-response curve of RDV versus SARS-CoV-2 RNA replication. Viral RNA was quantified using RT-qPCR in infected Vero E6 cells at 24 hpi. The concentrations of drug required for 50% inhibition (inhibitory concentration 50; IC_{50}) and 90% inhibition (IC_{90}) were estimated by fitting a non-linear (weighted) least-squares model (n=3).

5.5 Investigating the emergence of super-permissive cells

The kinetics analysis of infected Vero E6 cells identified a minor population of 'super-permissive' cells containing high gRNA copies. This result is surprising because (i) the infected cells were homogeneous monolayers of clones, (ii) I used an infection protocol maximising infection rate and synchronisation, and (iii) virus replication is evident in almost all cells, as depicted by the presence of sgRNAs, regardless of their development into 'super-permissivity'. The variable susceptibility to viral replication perhaps reflects the existence of cell-to-cell immune signalling, which could be an important feature of SARS-CoV-2 infection. Therefore, I decided to investigate their appearances more closely.

5.5.1 Super-permissive cells are commonly observed in multiple cell lines

'Super-permissive' cells are easily detectable with smFISH using low magnification air objectives (10-40x), making it amenable to scan the entire culture well and characterise the 2D localisation of such cells. For the spatial sampling of 'super-permissive' cells, I chose the 8 hpi timepoint as it represents the end-stage of primary infection, where extracellular viruses become detectable (Cortese et al., 2020; Kamel et al., 2021; Wing et al., 2021). I reasoned this timing should allow investigation of the initial emergence of 'super-permissive' cells, barring the effect of secondary infections. As characterised in the earlier kinetics assay, a typical infected population of Vero E6 cells consists of ~10% 'super-permissive' cells while the majority (85-90%) of the cells display low levels of viral RNA at 8 hpi (MOI 1) (Figure 5.17A-B). To extend these observations, I examined the viral RNAs in two human lung epithelial cell lines, A549-ACE2 and Calu-3, which are widely used as *in vitro* models of SARS-CoV-2 infection (Chu et al., 2020; Hoffmann et al., 2020). In agreement with the earlier observations with Vero E6, 3–5% of A549-ACE2 and Calu-3 cells showed a 'super-permissive' phenotype (Figure 5.17C-D). The percentage of 'super-permissive' cells was lower in human cell lines compared to Vero E6, which may be attributable to immunocompetence of these lines or species specificity. The overall infection rate was slightly lower in Calu-3 cells (~65%) probably because of the mild clumping morphology of this cell

line. Nevertheless, this result suggests the emergence of 'super-permissive' cells is a common feature during SARS-CoV-2 infection for multiple cell lines.

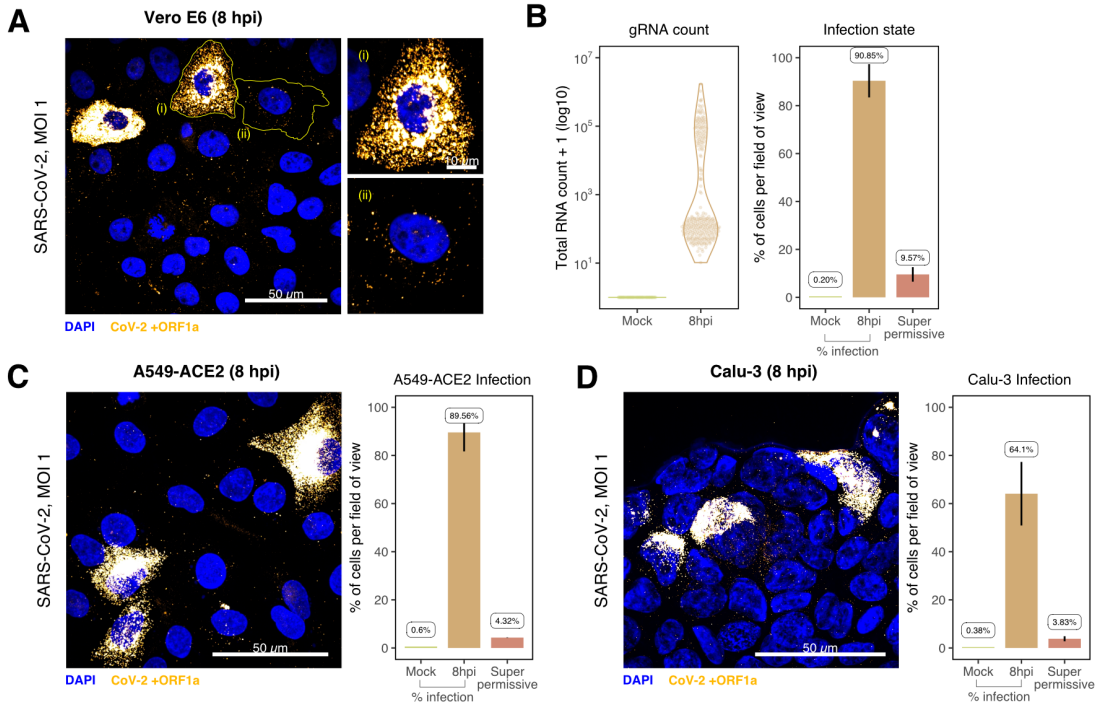


Figure 5.17: Identification of super-permissive phenotype in multiple cell lines. (**A**) Representative ×60 magnified field of view of SARS-CoV-2-infected (VIC-strain) Vero E6 cells at 8 hpi. +ORF1a probe was used to visualise cellular heterogeneity in viral RNA counts. Magnified panels show (i) a 'super-permissive' cell and (ii) a cell with discrete viral RNA copies. (**B**) Discrete separation of gRNA count distribution among infected cells randomly sampled at 8 hpi, where each symbol represents a cell (*left panel*). Statistics for the percentage of infected cells and frequency of 'super-permissive' cells at 8 hpi (*right panel*). The number labels represent average values. Cells with $>10^5$ gRNA copies were considered to be 'super-permissive' (n=4). (**C**-**D**) Heterogeneous SARS-CoV-2 replication in lung epithelial A549-ACE2 (**C**) and Calu-3 (**D**) cells. (n=3 for both A549-ACE2 and Calu-3 cell lines). The percentage of infected and super-permissive cells was quantified as with Vero E6 cells.

5.5.2 Spatial distribution of super-permissive cells

Understanding how the 'super-permissive' cells are distributed in culture can be informative because such patterns can highlight potential drivers for susceptibility (Healy et al., 2020). For example, viral infection can cause innate immune signalling leading to the secretion of soluble factors (e.g. interferons) (Belkowski & Sen, 1987; Schoggins & Rice, 2011). This type of signalling can induce an enhanced antiviral state in the vicinity of the susceptible cell. Such negative association can give rise to evenly spaced 'super-permissive' cells (Figure 5.18). In

contrast, 'super-permissive' cells could be the source of molecular factors promoting susceptibility to infection (positive association), in which a clustered pattern would be expected. Alternatively, the 'super-permissivity' phenotype could be defined by the cell-intrinsic state and subsequent random dispersion of these cells.

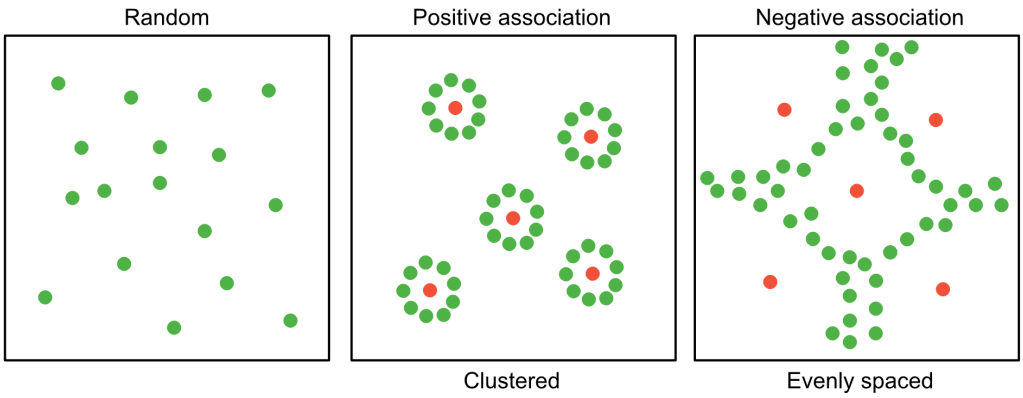


Figure 5.18: Proposed simulation schemes of 'super-permissive' cell spatial distribution. Schematics of the spatial distribution of cells in response to random, positive or negative associations. Red symbols represent the seed point of cell-to-cell interaction and green symbols represent resultant distribution in relation to the red symbol given the positive or negative associations.

Considering these three scenarios, I compared the average nearest-neighbour distances between 'super-permissive' cells (*Observed*) and simulated points of the same density of spots that were distributed either randomly, evenly, or in clusters (Figure 5.18). In this strategy, each of the three proposed patterns manifests a distinct range of nearest-neighbour distances that can be compared to the *Observed* distribution. I found that the 'super-permissive' infected Calu-3 cells exhibited significantly different spatial scatter from even and clustered distributions, but it could be modelled as randomly distributed (Figure 5.17A-C). This result suggests a lack of positive or negative association in their distributions. Similar random distribution patterns were found with 'super-permissive' Vero E6 and A549-ACE2 cells (Figure 5.19D). In summary, these data point toward the idea that susceptibility to 'super-permissivity' likely depends on cell-intrinsic properties rather than cell-to-cell signalling mechanisms.

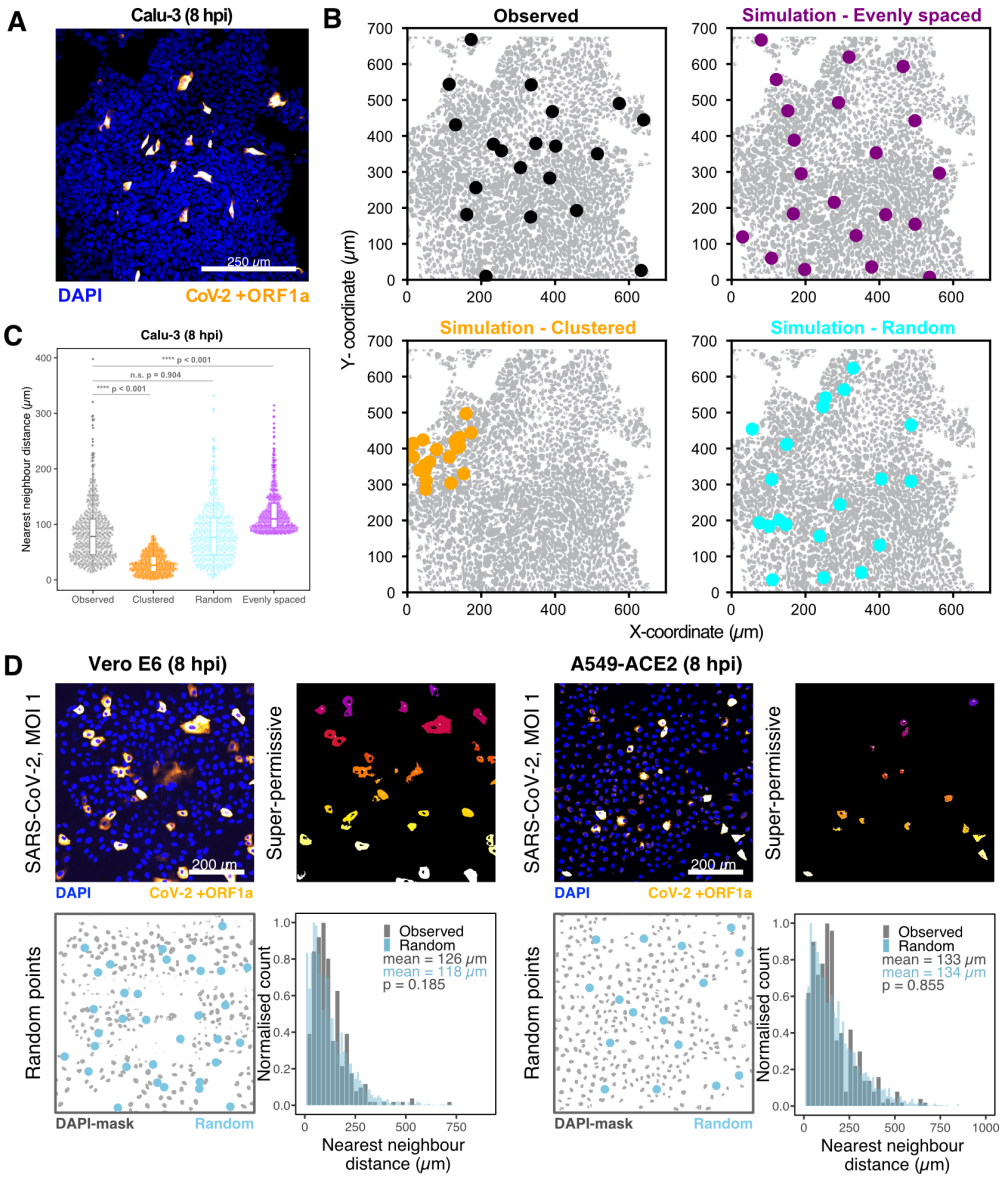


Figure 5.19: Spatial distribution of super-permissive SARS-CoV-2-infected cells. Spatial distribution analysis of 'super-permissive' SARS-CoV-2 (VIC strain) infected Calu-3 cells. +ORF1a probe was used to visualise viral RNA. (**A**) Low-magnification image (full z-projection) of infected Calu-3 cells showing a minor population of 'super-permissive' cells visible with smFISH. (**B**) Ground truth distribution of 'super-permissive' cells (*Observed*) and the example coordinates of the three modes of spatial distributions: (i) evenly spaced, (ii) clustered, and (iii) random. The simulations were confined to DAPI-positive masks (gray). (**C**) Beeswarm plot of nearest-neighbour distances obtained from the spatial distribution analysis. (n=3). (**D**) Spatial distribution analysis of super-permissive Vero E6 (*left panels*) and A549-ACE2 (*right panels*) cells at 8 hpi. Overview smFISH image of infected cells (*top left*). 2D mask of 'super-permissive' cells (*top right*). Simulated points of random distributions (cyan) within the DAPI mask (gray) (*bottom left*). Histogram of nearest-neighbour distances calculated from the 'Observed' and 'Random' distributions (bottom right).

5.6 Comparative replication kinetics of the SARS-CoV-2 strains

When the current study was underway, the virus pool in the United Kingdom became heavily dominated by the B.1.1.7 variant (also known as the UK Kent strain), which has been later renamed Alpha by the World Health Organisation. The B.1.1.7 variant is associated with higher transmission (Davies et al., 2021; Galloway et al., 2021; Volz et al., 2021) and has 17 coding changes mapping to both non-structural (ORF1ab, ORF3a, ORF8) and structural (Spike and N) proteins. Many studies have focused on the mutations in the Spike protein and whether these alter particle uptake into cells and provide resistance to vaccines or naturally acquired antibodies (Collier et al., 2021; Dicken et al., 2021; Planas et al., 2021). However, the potential alterations in their replicative fitness have not been explored in detail. Mutations within the non-structural genes could affect virus replication, independent of Spike-mediated entry, thus I used smFISH to compare the replication kinetics of the B.1.1.7 and the ancestral VIC strain.

5.6.1 B.1.1.7 replicates slower than the ancestral VIC strain

To perform a comparative replication kinetics assay, Vero E6 cells were infected with B.1.1.7 or VIC strains (MOI 1), fixed at 2, 6 and 8 hpi, and hybridised with +ORF1a and +ORFN probes (Figure 5.20A). I discovered that the number of gRNA molecules at 2 hpi and the percentage of infected cells were similar for both viruses, reflecting comparable entry efficiencies (Figure 5.20B). However, the growth of gRNA and sgRNA expressions were significantly delayed in B.1.1.7-infected cells compared to the VIC strain at 6-8 hpi (Figure 5.20C). Given the comparable level of virus entry at 2 hpi, this data suggests the reduced B.1.1.7 RNA burden is due to the differential replication efficiency rather than infection rate. Furthermore, treatment with the viral replication inhibitor RDV significantly closed the gap between the two strains, indicating that this phenotype is replication-dependent. Consistent with the delay in replication, the B.1.1.7-infected cells displayed a reduced number of replication factories per cell, and each factory contained on average fewer number of gRNA molecules than the VIC strain (Figure 5.21A-B). Despite the significant delay in the replication, the ‘super-permissive’ phenotype was observable at 6-8 hpi

with B.1.1.7, although at a much lower frequency than VIC. Interestingly, the vRNA burden in 'super-permissive' cells were comparable between the two strains.

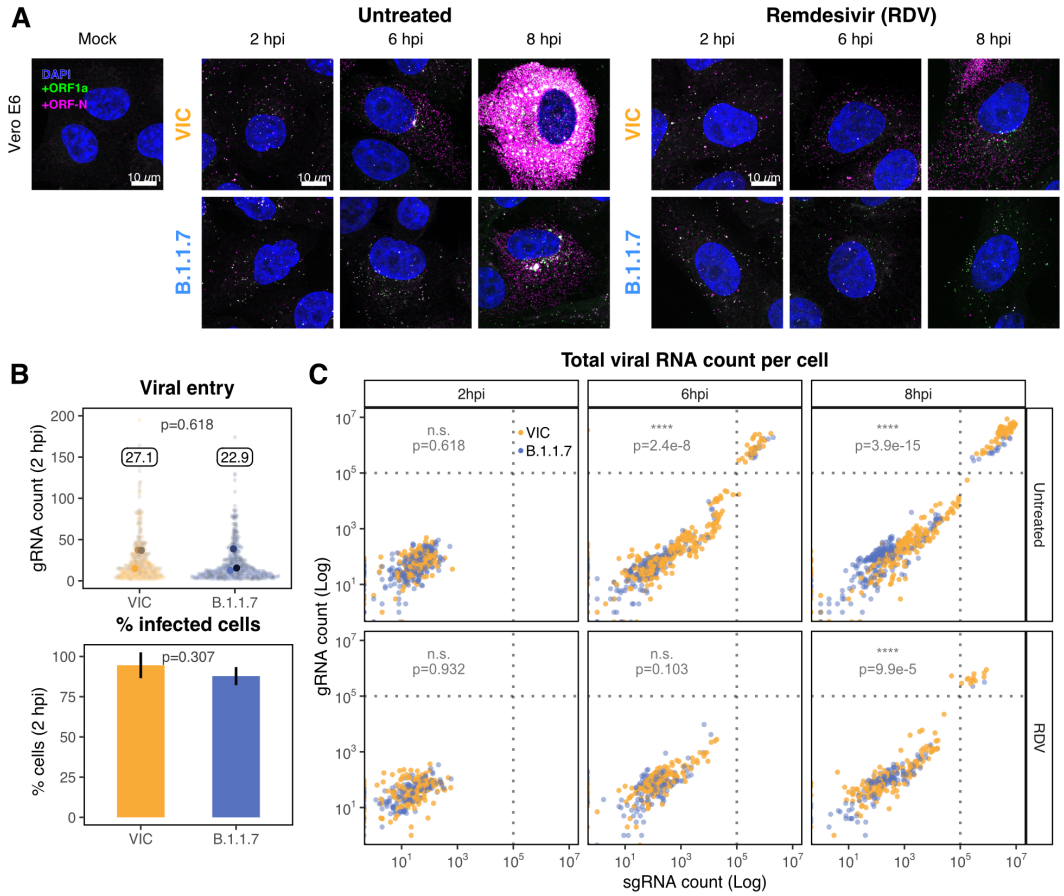


Figure 5.20: Delayed replication kinetics of B.1.1.7 variant. Vero E6 cells were infected with VIC or B.1.1.7 SARS-CoV-2 strains (MOI 1) for 2 h. Non-internalised viruses were removed by trypsin digestion, and cells were fixed at designated times. In the RDV-treated conditions, 10 μM drug was added at the point of virus inoculation. (**A**) Maximum z-projected confocal images of Vero E6 cells infected with VIC or B.1.1.7 strains. Representative 'super-permissive cells' are shown. (**B**) Comparison of gRNA counts at 2 hpi between VIC and B.1.1.7 infected cells (*top panel*). Each symbol represents a cell. The text label indicates average counts. (n=3). Comparison of percentage of infected cells between the two viral strains at 2 hpi (*bottom panel*). Infected cells were determined by +ORF1a probe fluorescence. Data are represented as mean±SD (n=3). (**C**) Bigfish quantification of gRNA and sgRNA smFISH counts per cell in VIC and B.1.1.7 infected Vero E6 cells. Due to the bimodality of the data, statistical significance was determined using a two-sample Kolmogorov-Smirnov test to compare the cumulative distribution of +ORF1a spot counts between the two strains. (n=3). ****p<0.0001.

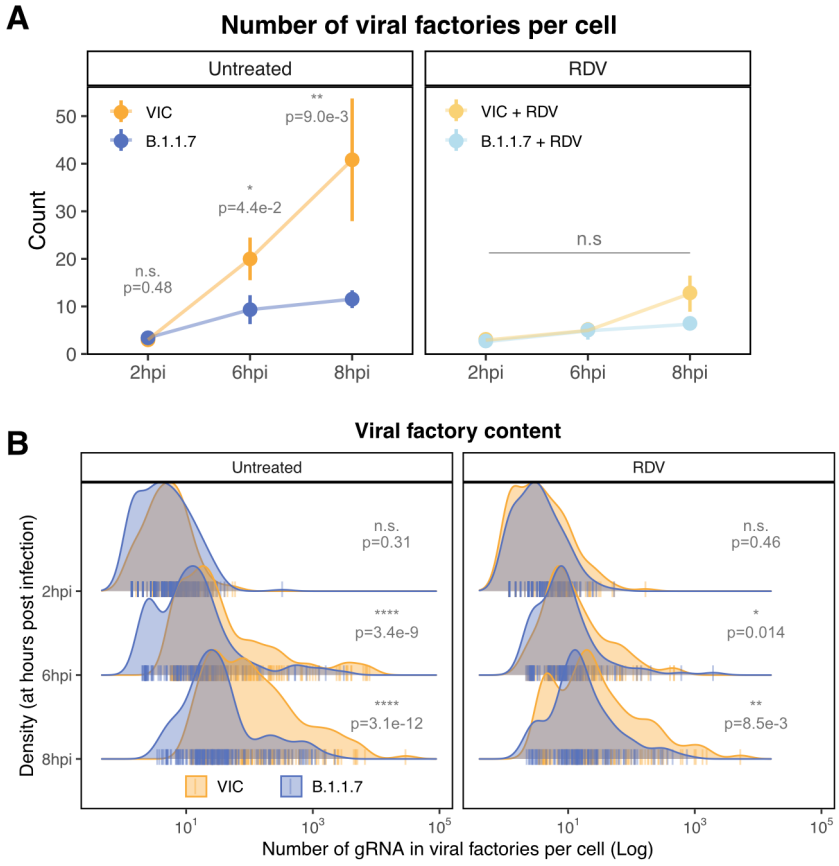


Figure 5.21: Comparison of viral factory dynamics between VIC and B.1.1.7. Vero E6 cells were infected with VIC or B.1.1.7 SARS-CoV-2 strains (MOI 1) for 2 h. Non-internalised viruses were removed by trypsin digestion, and cells were fixed at designated times. In the RDV-treated conditions, 10 µM drug was added at the point of virus inoculation. (**A**) A comparison of the number of viral factories per cell between the two viral strains. Cells harbouring >10^7 copies of vRNA were excluded from the analysis. Viral factories were identified using Bigfish cluster detection. Data are presented as mean±SEM. (n=3). (**B**) Density ridge plot showing the number of gRNA copies within each viral factory between the two strains. The density distribution represents the number of molecules per viral factory per cell. Each vertical segment symbol represents a cell. (n=3). *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

In agreement with the slower replication, B.1.1.7-infected cells demonstrated a shallower increase in the sgRNA/gRNA ratio between 2-8 hpi compared to VIC (Figure 5.22A). When the sgRNA/gRNA ratio was compared per gRNA burden classification basis (Figure 5.13A), I found that B.1.1.7-infected 'partially resistant' and 'permissive' cells displayed lower sgRNA/gRNA ratios compared to the VIC counterparts (Figure 5.22B). On the contrary, B.1.1.7-infected 'super-permissive' cells showed a 1.5-fold higher ratio compared to VIC, suggesting that sgRNA production is maintained at a slightly higher level in these cells following B.1.1.7-specific

mutations. In summary, these results illustrate differences in the replication kinetics of B.1.1.7 profiled at the molecular resolution.

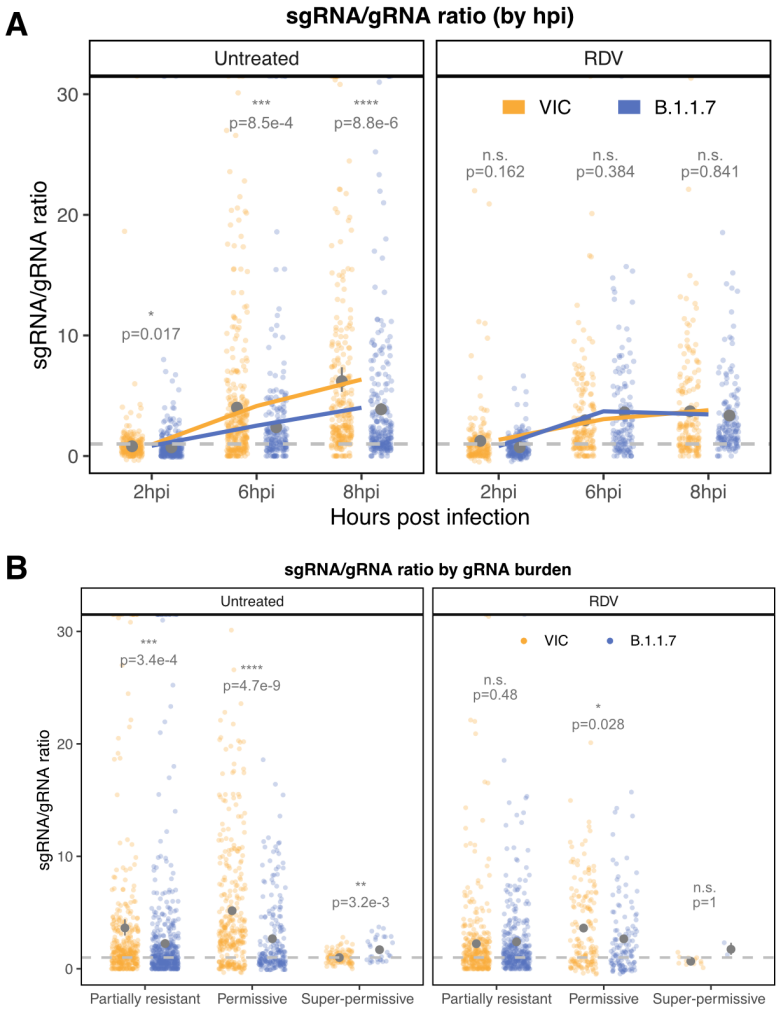


Figure 5.22: Comparison of sgRNA expression dynamics between VIC and B.1.1.7. Vero E6 cells were infected with VIC or B.1.1.7 SARS-CoV-2 strains (MOI 1) for 2 h. Non-internalised viruses were removed by trypsin digestion, and cells were fixed at designated times. In the RDV-treated conditions, 10 µM drug was added at the point of virus inoculation. (**A**) Per cell ratio of sgRNA/gRNA counts in VIC or B.1.1.7 infected cells. The horizontal dashed line represents the value of 1. (n=3). (**B**) Per cell ratio of sgRNA/gRNA counts grouped by gRNA burden classifications ('partially resistant', 'permissive', 'super-permissive'). Grey symbols represent mean ± SEM. The horizontal dashed line represents a ratio of 1. (n=3). *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

5.6.2 Super-permissive B.1.1.7 cells become clustered at 24 hpi

Using high-throughput smFISH intensity analysis, I found the slow-replicating B.1.1.7 also exhibit a lower frequency of 'super-permissive' cells at 6 and 8 hpi. (Figure 5.23A-B) To investigate whether 'super-permissive' cells of the two strains arise via different mechanisms, I performed spatial distribution analysis of B.1.1.7 or VIC infected cells at 8 hpi. Additionally, I also performed the same analysis at 24 hpi to characterise the potential effects of secondary infections on their development. At 8 hpi, the distribution of 'super-permissive' cells with B.1.1.7 was random (Figure 5.24), which showed a comparable distribution pattern to VIC. Nearest neighbour indices (*Rn*, see Methods section for details) of the two strains at 8 hpi approximated to 1, indicating random emergence of 'super-permissive' cells. In contrast, at 24 hpi, the distribution changed to a non-random pattern for B.1.1.7-infected cells, but it remained random for VIC. *Rn* indices at this stage with B.1.1.7 were skewed towards 0.4-0.5, indicating these cells display clusteredness. This result suggests B.1.1.7 potentially has a higher capacity to spread locally between adjacent cells than VIC.

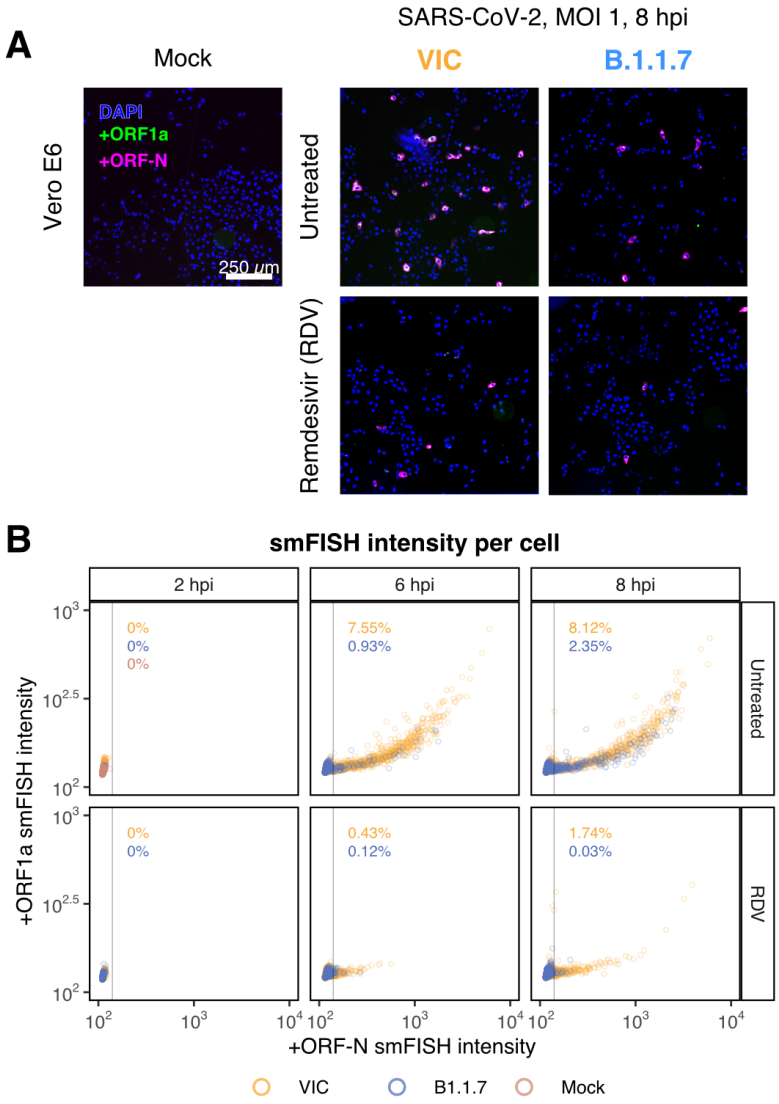


Figure 5.23: Frequency of B.1.1.7-infected 'super-permissive' cells. Vero E6 cells were infected with VIC or B.1.1.7 SARS-CoV-2 strains (MOI 1) for 2 h. Non-internalised viruses were removed by trypsin digestion, and cells were fixed at designated times. In the RDV-treated conditions, 10 µM drug was added at the point of virus inoculation. (**A**) Low-magnification (x20) z-projected (6 µm) images of +ORF1a and +ORFN smFISH in infected Vero E6 cells. (**B**) Scatter plot showing high-throughput smFISH intensity quantification of +ORF1a and +ORFN probe intensity in VIC and B.1.1.7-infected cells. Each symbol represents a cell. At this magnification, smFISH fluorescence is only detectable in 'super-permissive' cells above the background. The percentage of super-permissive cells was calculated based on a gate that was set with a +ORFN signal using the Mock condition as a threshold shown as a vertical line. (n=3).

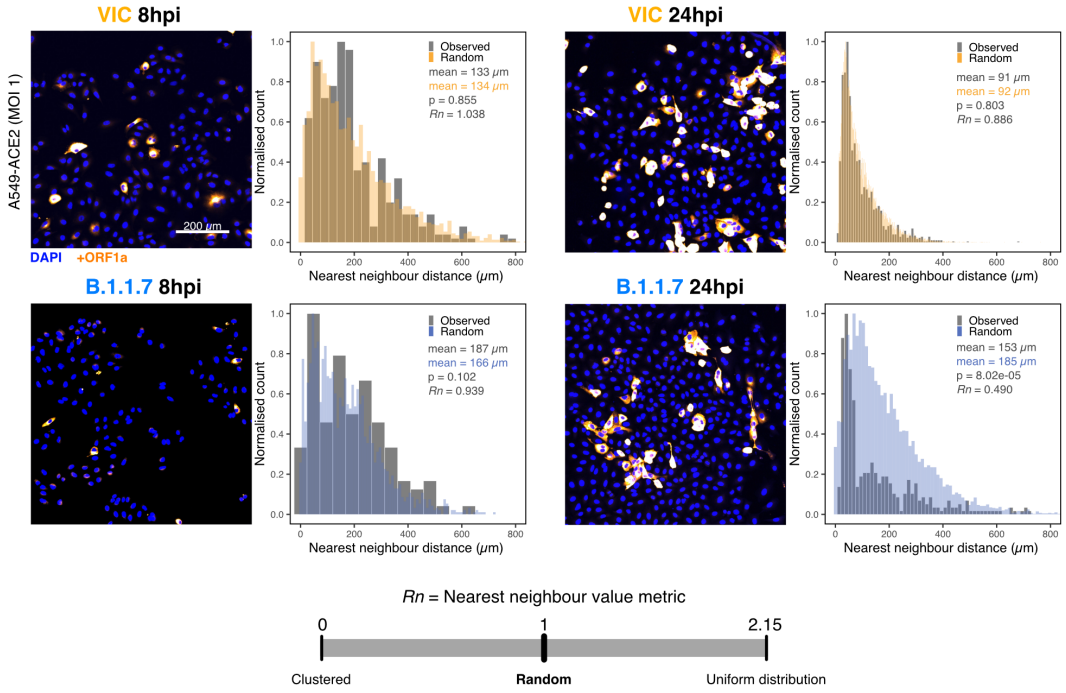


Figure 5.24: Comparison of 'super-permissive' cell spatial distribution between VIC and B.1.1.7. Spatial distribution analysis of super-permissive Vero E6 cells infected with VIC or B.1.1.7 strains at 8 and 24 hpi. Low-magnification overviews of infected cells hybridised with +ORF1a probes are shown. To compare the '*Observed*' distribution of super-permissive cells and '*Random*' distributions, the same number of random points as super-permissive cells was simulated 10 times per field of view. Nearest-neighbour distances calculated from super-permissive cells and simulated points are presented as histogram plots. Nearest-neighbour values (*Rn*) were calculated for each condition to assess the distribution of cell clusters (see Materials and methods). *Rn* values below 1 indicate a degree of clustering, and an *Rn* values >1 indicate uniform placements. (n=2).

5.6.3 B.1.1.7 phenotype is also observed in A549 cells

To rule out the possibility of a monkey cell-specific phenotype, I performed a similar kinetic smFISH assay using A549-ACE2 cells infected with B.1.1.7 or VIC (Figure 5.25A). Similar to Vero E6, VIC and B.1.1.7 showed a comparable frequency of infected cells and the number of gRNA molecules per cell at 2 hpi. These results show a similar degree of viral particle entry by the two strains (Figure 5.25B). However, infection with the B.1.1.7 variant led to a consistent reduction of gRNA and sgRNA burden at 8 and 24 hpi (Figure 5.25C). Moreover, fewer 'super-permissive' cells were detected at these timepoints. To evaluate whether the slower replication kinetics of B.1.1.7 leads to reduced secretion of new viral particles, Dr Peter Wing compared the level of infectious virus from the cell supernatant using plaque-forming assays. Dr Wing found a

modest but significant reduction in the infectious titre of B.1.1.7 compared to VIC at 8 and 24 hpi (Figure 5.26), consistent with the reduced cellular RNA burden. Considering these results together, I conclude that RNA replication and virion secretion rates of B.1.1.7 are slower than VIC in contrast to its fitness advantage to spread in the human population.

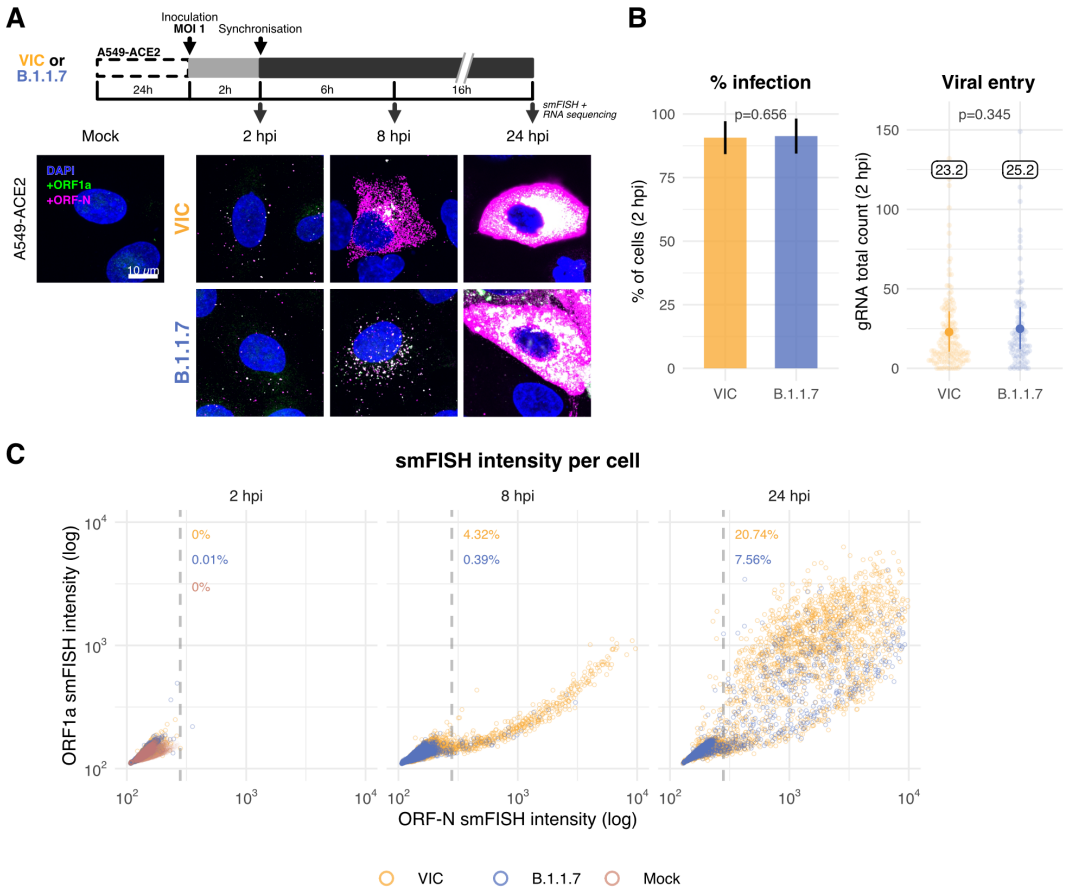


Figure 5.25: Comparative replication kinetics of VIC and B.1.17 in A549-ACE2. (**A**) Experimental design to compare replication kinetics and transcriptomic landscapes of VIC and B.1.1.7 strains. A549-ACE2 cells were inoculated with VIC or B.1.1.7 strains (MOI 1) for 2h. Non-internalised viruses were removed by trypsin digestion, and cells were fixed at 2, 8 and 24 hpi. Maximum z-projected confocal images of A549-ACE2 cells infected with VIC or B.1.1.7. Representative super-permissive cells from the time series are shown. (**B**) Comparing the percentage of infected A549-ACE2 cells between VIC and B.1.1.7 at 2 hpi (*left panel*). Infected cells were determined by +ORF1a smFISH fluorescence positivity. Data are represented as mean±SD. Comparison of gRNA counts at 2 hpi between the two strains (*right panel*). Each dot represents a cell. Data are represented as mean±SEM, and the labels represent average values. (**C**) Scatter plot showing high-throughput smFISH intensity quantification of +ORF1a and +ORFN channels in the infected cells. Each symbol is a cell. The threshold gate was set with a +ORFN signal using the Mock condition (dotted line). (n=3).

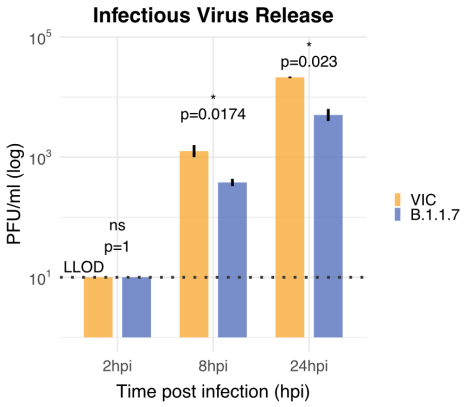


Figure 5.26: Comparison of infectious particle release between VIC and B.1.1.7. Comparison of VIC and B.1.1.7 infection on the secretion of infectious viruses. This experiment was performed by Dr Peter Wing. Extracellular media from infected A549-ACE2 cells were collected at the indicated timepoints and infectious virus was quantified using a plaque forming assay. Data are presented as mean±SEM. No apparent plaques were found at 2 hpi. Pseudocounts of 10 PFU/ml were added at 2 hpi for plotting purposes (LLOD, lower limit of detection). (n=3). *p<0.05.

5.6.4 Orthogonal RNA-seq confirms smFISH results

To validate my smFISH observations using an independent method, Dr Peter Wing and Natasha Palmalux sequenced ribo-depleted total RNA libraries of matching A549-ACE2 cells infected with B.1.1.7 or VIC at 2, 8 and 24 hpi (Figure 5.25A). Dr Aino Jarvelin performed pre-processing of the raw RNA-seq output (J. Y. Lee et al., 2022). As expected, the number of reads mapping to the SARS-CoV-2 genome increased over time (Figure5.27A), reflecting active replication. Between 2-24 hpi, reads mapping to the 3' end of the viral genome increased relative to the 5' portion, indicating more frequent synthesis of sgRNAs over gRNA (Kim et al., 2020). Consistent with the similar internalisation rates assessed using smFISH, RNA-seq also revealed a similar portion of reads mapping to SARS-CoV-2 at 2 hpi between B.1.1.7 and VIC (Figure 5.27B).

Reassuringly, the sequencing analysis also revealed a reduced abundance of vRNAs in B.1.1.7-infected cells at 8-24 hpi. The level of B.1.1.7 vRNA was almost unaltered between 2 and 8 hpi and then increased dramatically at 24 hpi. This contrasts with VIC-infected samples, which showed a continuous increase over time. It is also notable that the RNA-seq data suggest closing of the vRNA abundance gap between the two strains at 24 hpi, which is consistent

with my spatial analysis showing the enhanced ability of B.1.1.7 to spread between cells in secondary infections (Figure 5.24).

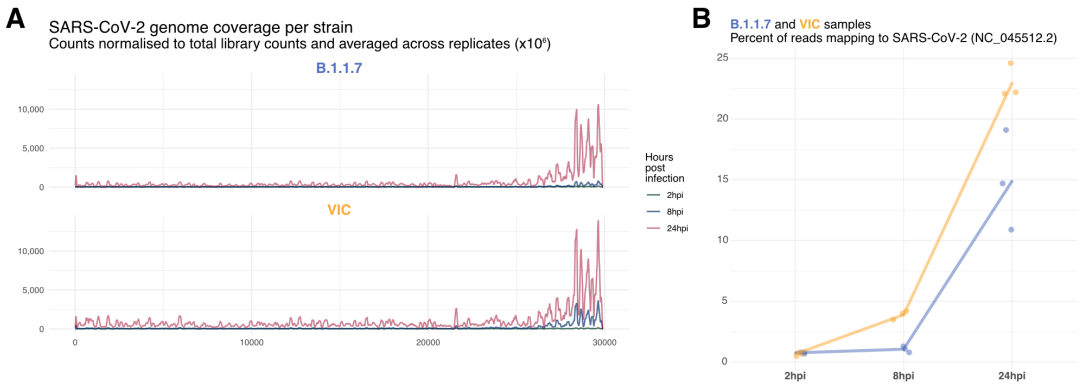


Figure 5.27: RNA-seq analysis of replication kinetics comparing VIC and B.1.1.7. A549-ACE2 cells were inoculated with VIC or B.1.1.7 strains (MOI 1) for 2h. Non-internalised viruses were removed by trypsin digestion, and cells were harvested at 2, 8 and 24 hpi for RNA extraction and sequencing. (**A**) Read coverage of the positive-strand SARS-CoV-2 genome. Read counts were normalised to total library sizes and averaged across replicates. (n=3). (**B**) Percentage of reads mapping to the SARS-CoV-2 genome of total mapped reads, shown separately for the two strains. Each symbol represents an experimental replicate (n=3).

To further explore the differences in gene expression between the B.1.1.7 and VIC strains, I assessed the abundance of different viral RNAs in the infected A549-ACE2 cells from the RNA-seq libraries. From these libraries, a very small fraction of vRNA reads were mappable to the negative strand, however, these transcripts are detectable as early as 2 hpi (Figure 5.28A). These data add further support to my earlier conclusion that viral replication events occur rapidly upon cell entry (Figure 5.15A). The ratio between negative and positive sense vRNAs increased gradually over time for both strains. Although I observed a small reduction in the ratio for B.1.1.7 at 24 hpi, it was not statistically different from VIC. To assess the expression of sgRNAs, Dr Jarvelin quantified the reads mapping to the discontinuous transcription junctions derived from viral RdRP (Kim et al., 2020; J. Y. Lee et al., 2022; Thorne et al., 2021; V'Kovski et al., 2021). In agreement with smFISH data, the sgRNA/gRNA ratio peaked at 8 hpi (Figure 5.28B), followed by a decline at 24 hpi. B.1.1.7 displayed a significantly lower sgRNA/gRNA ratio at 8 hpi compared to VIC, but the ratio remained stable between 8-24 hpi, and surpassed VIC at 24 hpi. This result agrees with my earlier observation of the differential sgRNA expression kinetics of B.1.1.7 in

'super-permissive' cells (Figure 5.22A-B). Together, these RNA sequencing data confirm that B.1.1.7 variant exhibits delayed replication kinetics, complementing my smFISH results.

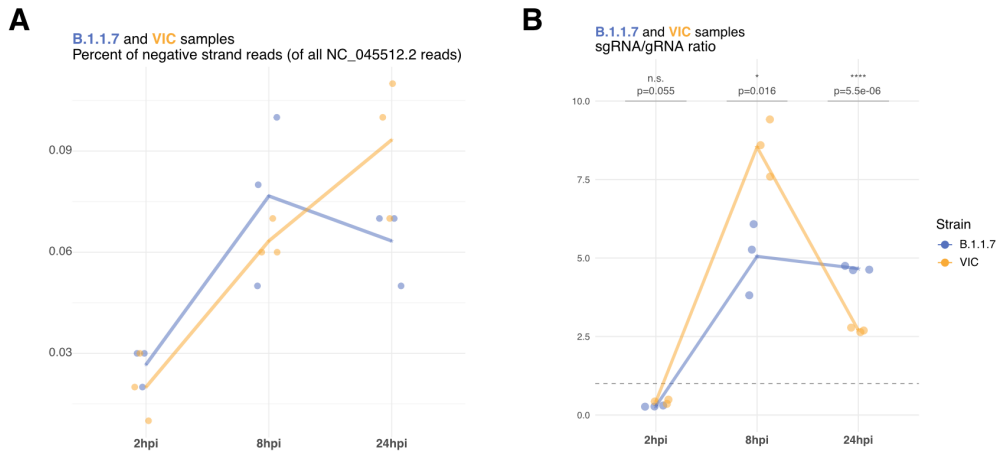


Figure 5.28: Comparing viral negative-strand RNA and sgRNA expression between VIC and B.1.1.7. A549-ACE2 cells were inoculated with VIC or B.1.1.7 strains (MOI 1) for 2h. Non-internalised viruses were removed by trypsin digestion, and cells were harvested at 2, 8 and 24 hpi for RNA extraction and sequencing. (**A**) Percentage of reads mapping to SARS-CoV-2-negative (antisense) strand relative to all SARS-CoV-2 reads, shown separately for the two strains. Each symbol represents an experimental replicate. (n=3). (**B**) Estimated ratios of SARS-CoV-2 subgenomic to genomic RNA for the two virus strains at the three sampled timepoints. sgRNA abundance estimation was analysed by Dr Aino Jarvelin. (n=3). *p<0.05; ****p<0.001.

5.7 Application of smFISH in Hepatitis B virus infection

So far, I have demonstrated the utility of smFISH workflow in the RNA virus, SARS-CoV-2. Next, I explored whether my sensitive imaging method is applicable to detect and measure DNA virus infections. To test this, I chose the Hepatitis B virus (HBV) as a model. Chronic infection of HBV is driven by the persistence of covalently closed circular DNA (cccDNA) in the nucleus, which can be integrated into the host genome (Nassal, 2015). Replication of HBV involves transcription of pre-genomic RNA (pgRNA), which undergoes subsequent reverse transcription to replenish cccDNA or produce a viral genomic DNA (Yuen et al., 2018). Therefore, pgRNA is one of the key molecules governing the HBV infection lifecycle.

To detect HBV pgRNA, I designed a smFISH probe set against the unique pgRNA sequence. HepG2 cells expressing NTCP receptor (HepG2-NTCP) were inoculated with HBV (1,000 genome

copies per cell) and fixed at 48 hpi. HBV *de novo* infection was performed by Dr Senko Tsukada in the Category 3 laboratory. Hybridisation with pgRNA probes revealed distinct diffraction-limited spots in the cytoplasm of the infected cells (Figure 5.29A). Additionally, brighter and larger puncta were visible within the nucleus, representing pgRNA transcription foci. These transcription foci could be decomposed into individual RNA units using the same method to resolve SARS-CoV-2 virus factories. Importantly, RNase treatment of infected cells resulted in ablation of the signal, confirming specificity. I then tested whether pgRNA probes could be used to detect viral RNA in archived liver biopsy FFPE specimens. smFISH showed a bright signal in liver cells known to be infected with HBV (Figure 5.29B), and it produced clear diffraction-limited spots. Collectively, these results represent an exciting opportunity to study HBV infection from the point of nascent and mature pgRNA dynamics.

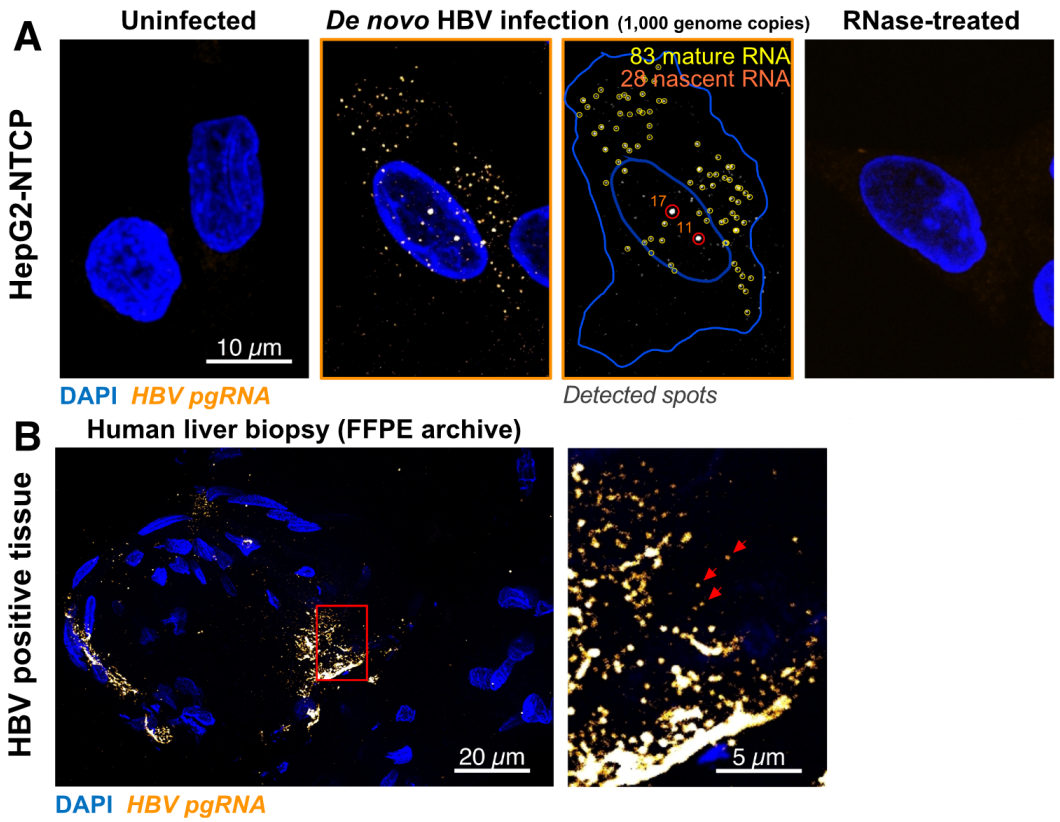


Figure 5.29: Sensitive detection of Hepatitis B virus pgRNA. (**A**) smFISH against pre-genomic RNA (pgRNA) on HepG2-NTCP cells infected with HBV (1,000 genome copies per cell) at 48 hpi. FISH-quant spot detection overlay shows the number of detected cytoplasmic pgRNAs and nascent pgRNAs within the nucleus. RNase treatment was used as a control to confirm RNA detection. Infected cells were provided by Dr Senko Tsukada. (n=2). (**B**) smFISH of HBV pgRNA in a formalin-fixed paraffin-embedded liver biopsy sample from an HBV-positive patient. Red arrows represent diffraction-limited single-molecule spots. Maximum intensity projected confocal images are shown. A representative image from five field of views is shown.

5.8 Conclusion and discussion

5.8.1 New insight into SARS-CoV-2 early replication events

In this Chapter, I presented the utility of using an imaging-based method to provide new insights into the early-stage RNA dynamics of SARS-CoV-2 infections. This work complements approaches like bulk RT-qPCR and conventional RNA-seq as it allows absolute quantification of viral RNAs at single-molecule and cell resolutions. My results highlighted the detection of individual gRNA molecules within the first 2 h of infection, which most likely reflects incoming viral particles (Eymieux et al., 2021; Mendonca et al., 2021). More importantly, I also observed

a small number of viral factories comprising several gRNAs, that were sensitive to the RDV treatment, as well as the presence of sgRNA in these cells. These data provide the first evidence that SARS-CoV-2 replication events are initiated within the first 2 h of infection. Therefore, highly sensitive tools are absolutely necessary to interrogate these early replication events. I have also demonstrated the disadvantage of the commonly used J2 anti-dsRNA for surveying these early stages and the advantage of using +ORF1a probes over J2. I also noted that co-staining infected cells with J2 antibody and +ORF1a showed a partial overlap, suggesting that the virus may also induce changes in the cellular pool of dsRNA (Shang et al., 2022).

A key finding of the study is that cells vary in their susceptibility to SARS-CoV-2 infection. During primary infections (~6-8 hpi), most cells had low vRNA levels ($<10^2$ copies/cell), but a minor population (4-10%) had a much higher vRNA burden ($>10^5$ copies/cell). Notably, the number of intracellular vRNA copies at 2 hpi was similar across the culture, suggesting the 'super-permissivity' phenotype is not explained by differences in virus entry. The amount of vRNA in these 'super-permissive' cells account for the bulk of the vRNA within the culture, and they can easily mask the dominant cell population. Hence, care should be taken when analysing bulk extracted RNA in similar experimental settings. The 'Super-permissive' phenotype was a common feature of multiple cell lines including Vero E6, A549-ACE2 and Calu-3. As both A549-ACE2 and Calu-3 have intact innate sensing pathways (Cao et al., 2021; Y. Li et al., 2021), unlike Vero E6 cells (Desmyter et al., 1968), this variable susceptibility is unlikely to reflect differential immune signalling and is consistent with their random emergence within the culture. The cause of this heterogeneity in infection fitness is not clear but it may rely on the intrinsic properties of each cell, such as the stage of the cell cycle, metabolic state or the expression of individual antiviral genes. However, it should be noted that the virus stock used in this study had a high particle-to-infectivity ratio (Figure 5.2A & Figure 5.12A). Although the majority of the cells received replication-competent viruses, as indicated by the presence of +sgRNA in 98% of cells at 4 hpi (Figure 5.15A), the effect of high particle-to-infectivity ratio on the variability of virus replication is yet to be established. An interesting experiment to confirm the intrinsic variability of the host cell towards virus replication would be to co-infect using two SARS-CoV-2

strains encoding different fluorescent proteins. Here, the degree of overlap of the two colours in 'super-permissive' cells will be able to explain the relative contribution of the intrinsic cellular state versus the variability in the virus inoculum leading to heterogeneous SARS-CoV-2 replications.

Several single-cell RNA sequencing studies of SARS-CoV-2 infected cultures elucidated ciliated cells as the primary target of the virus, however, only a minority of these cells were shown to have viral reads (Fiege et al., 2021; Ravindra et al., 2021). These results also point to the idea of cell-to-cell variation in susceptibility, however, I cannot rule out the effect of low sequencing depth in these experiments. Applying *in situ* hybridisation methods in concert with cell-type-specific labels to clinical biopsies and experimentally infected animal samples (Salguero et al., 2021) will be an important next step to understand how specific cell types and viral gRNA burdens are related.

I found that SARS-CoV-2 gRNA persisted in the cytoplasm of infected cells in the presence of RDV, suggesting a long half-life that may reflect the secondary structure of the RNA genome resisting host nuclease activities (Huston et al., 2021; Simmonds et al., 2021). Furthermore, smFISH revealed complex dynamics of gRNA and sgRNA expression patterns where rapid transcription of sgRNA (peaking at 8 hpi) is followed by a shift towards gRNA production at 24 hpi. Because the viral particle stoichiometry heavily favours structural proteins and only a single copy of gRNA, I hypothesise the initial preferential synthesis of sgRNA aims to fulfil the protein demand for viral structural proteins. Once the structural proteins are available in sufficient quantities, the late shift towards gRNA synthesis may ensure sufficient copies of gRNA to generate enough virions.

5.8.2 Differential replication kinetics of SARS-CoV-2 variants

B.1.1.7 has been reported to encompass an enhanced ability to transmit compared to the ancestral Wuhan isolates (Caly et al., 2020; Davies et al., 2021; Volz et al., 2021). Multiple mutations in the Spike sequence have been demonstrated to improve human ACE2 binding, but the consequences of non-Spike mutations leading to enhanced viral fitness are not well understood. Interestingly, my comparative smFISH kinetics assay revealed slower replication dynamics of the B.1.1.7 variant compared to the ancestral strain. The consensus on the replication behaviour of B.1.1.7 is not

clear, where occurrences of both faster and slower replication have been reported (Dicken et al., 2021; Thorne et al., 2021; Zhang et al., 2021). However, these studies were performed either at late timepoints post-infection (24-48 hpi) or in virus-free *in vitro* replication systems. It is notable that while B.1.1.7 still lags vRNA replication compared to VIC at 24 hpi, it displays a significant recovery compared to 8 hpi. Therefore, it is possible that towards the late-stage infection, B.1.1.7 can yield a similar, or even higher, viral RNA load compared to VIC. Other mutations introduced in B.1.1.7 are also known to produce a novel accessory protein ORF9b, which has been shown to suppress the host immune response (Parker et al., 2021; Thorne et al., 2021). Therefore, it is tempting to speculate that the initial delay in vRNA replication synergises with ORF9b to evade host antiviral response. Together with improved virus entry mediated by mutated Spike, these features may provide B.1.1.7 with a comparative advantage over the original Wuhan strain.

Finally, I demonstrated my imaging workflow could be used to detect and quantify RNA intermediates of HBV, providing new avenues to investigate RNA dynamics in DNA virus infections. In summary, these findings highlight the utility of smFISH to uncover new aspects of virus RNA metabolism that are worthy of further study.

6 Discussion

6.1 Post-transcriptional regulation of *Drosophila* neurogenesis

How is the vast diversity of neurons produced from neural stem cells? This question remains one of the greatest challenges of neurobiology. The developing brain of *Drosophila* has been used extensively to uncover the mechanisms and principles of neurogenesis. In the last few decades, the key genes that regulate neuronal identity have been identified and found to predominantly encode TFs, many of which are also conserved in mammalian systems. However, much is yet to be discovered for the upstream regulation of TF themselves. Cell-specific RNA sequencing experiments and large-scale RNAi screens have revealed RBPs as a key partner of TFs to finely regulate temporal and spatial gene expression (Berger et al., 2012; Konstantinides et al., 2022; Z. Liu et al., 2015; Neumuller et al., 2011; Stoiber et al., 2015). To date, only limited examples of post-transcriptional regulation in neurogenesis have been described using candidate approaches (Guan, Bellemin, et al., 2021; L.-Y. Liu et al., 2020; Y. Liu et al., 2016; Samuels, Järvelin, et al., 2020). Therefore, in the first two Chapters of this thesis, I explored the contribution and breadth of post-transcriptional control in *Drosophila* nervous system development.

In Chapter 3, I performed iCLIP to identify comprehensively the RNA targets of Syp in the developing brain. Syp binds to more than 1,000 transcripts in the brain, which includes both novel and previously characterised genes involved in different steps of brain development. I

found that Syp targets were highly enriched for transcripts encoding TFs and RBPs, and the target transcripts featured highly extended 3'UTR isoforms. In whole larvae, Syp is mainly bound to targets encoding cytoskeletal binding proteins, suggesting that Syp regulates different biological processes in a compartment-specific manner. Then, I focused on target TFs with known roles in neurogenesis, in particular with long 3'UTR extensions. 3'UTRs harbour regulatory sequences that are often recognised by RBPs to exert their functions (Mayr, 2017), so longer 3'UTRs may have higher regulatory potentials. Using a detailed imaging analysis, I characterised two examples of post-transcriptional regulation of TFs during neurogenesis. Pros has been shown to permanently silence self-renewal genes in mature neurons by epigenetic control of heterochromatin (X. Liu et al., 2020). I showed that H3K9me3 modification requires Syp-mediated stabilisation of the $pros^{long}$ isoform. Next, I studied how Syp influences the progression of the larval NB temporal TF cascade from Br to E93 (Syed et al., 2017). Using smFISH and immunofluorescence to quantify the *e93* transcription rate and mRNA half-life, I showed that Syp controls the rate of translation of *e93* but not its transcript stability. Furthermore, Syp was required to initiate the transcription of *e93*, suggesting Syp has a broader role that impacts multiple gene regulatory layers. Modulating the levels of *syp* or *e93* led to a precocious extension of the Br temporal window. The results of these experiments provide insight into the understanding of how Syp and post-transcriptional regulations can influence NB differentiation and temporal patterning.

In Chapter 4, I then asked how prevalent post-transcriptional regulation is in the NB lineage. Post-transcriptional regulation often manifests itself in the discordant distribution of mRNA in comparison to the distribution of the encoded protein (Buccitelli & Selbach, 2020; Titlow et al., 2022). I reasoned that examining the 'lifecycle' of gene expression from transcription to translation can provide greater insights into the complexity of gene expression control and its influence on neurogenesis and NB differentiation. While in the previous Chapter's examples of *pros* and *e93*, the complex gene regulation patterns were elucidated via a low-throughput method, Chapter 4 describes a medium-throughput method to achieve the same detailed analysis of gene expression. The key to the increased throughput was the use of GFP or YFP protein-trap fly libraries to systematically interrogate protein distributions, instead of using an antibody, and

mRNA distributions through smFISH with the same probe set against GFP or YFP expression (Lowe et al., 2014; Nagarkar-Jaiswal et al., 2015). This higher throughput strategy was first applied to Syp targets with long 3'UTR extensions. The results revealed that more than 50% of the tested genes show discordance between protein and mRNA expressions in different cell types of the NB lineage. The discordance profiles could be easily categorised into regulated mRNA stability or translation, based on the simple comparison of the levels of transcription, mRNA accumulation and protein output. Therefore, this cohort of genes is potentially regulated post-transcriptionally by Syp and may be important for specifying cell identity and regulating cell-type specific behaviours.

This approach was then applied to 200 pre-existing YFP-trap lines to estimate the global contribution of post-transcriptional regulation in brain development, in a project co-led by Dr Joshua Titlow, Maria Kiourlappou, Dr Ana Palanca, and myself (Titlow et al., 2022). With high spatial resolution and tissue-specific context, we found that 12/200 genes showed obvious evidence of post-transcriptional regulation in NB lineages. Simple extrapolation of this proportion suggests that over 450 transcripts could be regulated post-transcriptionally and exhibit a similar discordant patterns in the NB lineage. Our results provide insight into the widespread metabolism of mRNA *in vivo*, and highlight the need to improve our understanding of post-transcriptional gene expression control. Collectively, we have demonstrated a powerful approach that opens up future experimental possibilities in model organisms with pre-existing protein-trap libraries. The availability of protein-traps has exploded in the past few years with the advent of CRISPR methods, making our approach very widely applicable (Lee et al., 2018b; Van Nostrand et al., 2017).

6.1.1 Relevance to mammalian neurogenesis

Although the experiments I describe in this thesis were performed in *Drosophila* larval brains, it would be interesting to investigate whether similar gene expression principles are conserved in mammalian neurogenesis. Both Syp and Imp have known human orthologues that are expressed in the developing nervous tissues, which may regulate a similar cohort of transcripts in mammalian brains. The targets of human and murine SYNCRIP/HNRNPQ, indeed highlight a group of transcripts involved in axon guidance and neuron-specific mRNA localisation processes (Briese et

al., 2018; Khudayberdiev et al., 2021). However, the function of SYNCRIP in neural stem cells is yet to be established in mammals, and it is possible that multiple redundant factors, such as HNRNPR, could act together to ensure correct neurogenesis in mammals. Moreover, mammalian SYNCRIP has a protein domain, lacking in *Drosophila* Syp, at its C-terminal end that has been shown to be Synaptotagmin binding (Mizutani et al., 2000), which could mean that mammalian SYNCRIP functions a little differently from *Drosophila*. Interestingly, as in *Drosophila*, evidences exist to support the role of morphogenetic gradient to influence mammalian NSC behaviour and neurogenesis of different neuron types (Caviness et al., 2009; Sun et al., 2015). Given the similarity to the temporal gradient of RBPs in the fly brain, it would be worth investigating whether a similar temporal molecular gradient also regulates mammalian neurogenesis.

6.1.2 Future perspectives

Many key questions remain to be explored. What is the detailed molecular mechanism of Syp activity? I found that Syp promotes stability of $pros^{long}$ and translation of *e93*, with direct indications that Syp binds to the highly extended 3'UTR of these transcripts. Does Syp binding protect $pros^{long}$ from competing RBPs driving mRNA degradation? Does Syp recruit translational machinery to its targets? Many hnRNPs, such as Syp, do not harbour protein domains that can directly alter the fate of their target mRNA (Hobor et al., 2018). Therefore, the molecular mechanism of Syp will depend heavily on the interacting protein partners (Geuens et al., 2016). The varying proteome of stem cells across different developmental timescales adds further complexity to the Syp regulome, and may even lead to differential regulatory consequences on the same target mRNA (Van Nostrand, Pratt, et al., 2020). Precise multi-omic profiling of Syp interactors could provide greater insight into its function, but without such experimental evidence, one can only speculate on the molecular mechanisms of regulation by Syp.

Do Syp and Imp compete for the same transcript? The opposite regulation of *chinmo* by Syp and Imp (Dillard et al., 2018; Z. Liu et al., 2015) highlights that competition between Syp and Imp could be common, and this inter-RBP competition may contribute to the neural diversity and NB age determination (Doe, 2017a; Guan, Bellemin, et al., 2021; Guan, Bouchet,

et al., 2021). In Chapter 3, I introduced my future plans to identify time-resolved targets of Syp and Imp across multiple developmental stages. I anticipate this experiment will be able to profile dynamic RNA interactomes of Syp and Imp and determine whether they bind to the same sites on the shared targets.

RBPs constitute ~10% of the human genome, and they regulate the diverse landscapes of RNA metabolism in many biological processes (Castello et al., 2012; Van Nostrand, Freese, et al., 2020). The introduction of tagged-RBP cell-culture libraries and standardised CLIP procedures have significantly expanded the catalogue of functional *cis*-elements encoded in the human genome that function at the mRNA level through interaction with RBPs (Van Nostrand et al., 2017; Van Nostrand, Freese, et al., 2020). The systematic survey of mRNA and protein distribution presented in this thesis adds a spatial dimension to mRNA metabolism in a complex tissue. Together these efforts will be able to help improve our understanding of the principles of mRNA processing and their wider consequences in gene expression control.

6.2 RNA replication dynamics of SARS-CoV-2 infection

Since the onset of the COVID-19 pandemic, increasing attention has been given to the ongoing threat of novel zoonotic coronavirus spillover to the human population (Fung et al., 2021). Currently, therapeutic options to treat coronavirus infections are still limited (Frediansyah et al., 2021; Vegivinti et al., 2022). Thus more in-depth understanding of viral infection is essential to identify new targets for antivirals. Much of this effort has focused on the mechanisms of viral entry into host cells. However, our quantitative and kinetic knowledge of the early SARS-CoV-2 replicative lifecycle is poor, despite its essential role in the establishment of a productive infection. With this in mind, I developed an smFISH-based approach to detect and quantify SARS-CoV-2 RNA at single-molecule and cell resolutions, with 'absolute' sensitivity, which is able to detect most molecules in the cell. This work constitutes the second part of this thesis.

In Chapter 5, I presented the complete workflow of single-molecule sensitivity SARS-CoV-2 viral RNA detection from image acquisition to computational analysis. The novel workflow

highlighted the advantage of the *in situ* hybridisation method over anti-dsRNA detection and its compatibility with sequencing results. A key new finding of this study is that cells vary in their susceptibility to SARS-CoV-2 infection, giving rise to heterogeneous populations of 'partially-resistant', 'permissive', and 'super-permissive' cells. It is important to note that viral replication does occur in non-'super-permissive' cells as evidenced by the presence of sgRNAs and the slow-but-steady increase in gRNA counts over time. Therefore, methods with limited sensitivity may under-represent the rate of infection in similar *in vitro* models. Furthermore, due to exponentially high vRNA load, ensemble methods may not distinguish between the enhanced viral replication versus the increased frequency of 'super-permissive' cells.

How is 'super-permissiveness' of cells to infection induced? Spatial distribution analysis of these cells suggests the phenotype is dependent on the nature of each cell in a cell-autonomous manner. Such intrinsic variability may involve cell cycle status, local mRNA-binding proteome, or stochastic expression of innate immune mediating proteins. It would be interesting to purify 'super-permissive' cells and investigate their transcriptomic and proteomic features. Cross-referencing these results to the existing SARS-CoV-2 RNA interactome (Kamel et al., 2021; Lee, Lee, et al., 2021) may provide further insight. Interestingly, low oxygen tension or pharmacological induction of hypoxia down-regulate the frequency of 'super-permissive' cells (Wing et al., 2021), thus the hypoxia-inducible factor (HIF) pathway seems to be partially involved in allowing cells to be 'super-permissive'.

Next, I applied smFISH to compare the replication kinetics of SARS-CoV-2 strains, and I found that the replication of B.1.1.7 is significantly delayed compared to the ancestral VIC. B.1.1.7 has been reported to have an enhanced ability to transmit compared to the ancestral Wuhan isolates (Caly et al., 2020; Davies et al., 2021; Volz et al., 2021). Multiple mutations in the Spike sequence have been demonstrated to improve human ACE2 binding, but the consequences of non-Spike mutations leading to enhanced viral fitness are not well-understood. Interestingly, a comparative study of B.1.1.7 and VIC reported elevated levels of intracellular vRNA and N protein in B.1.1.7-infected cells at 24-48 hpi (Thorne et al., 2021). Although my results suggest

B.1.1.7 still lags behind in vRNA production at 24 hpi, it exhibits a clear recovery compared to 8 hpi. It is therefore possible that B.1.1.7 yields a similar, or even higher, total amount of viral RNA and protein than VIC, just in a prolonged and delayed temporal fashion. In line with this idea, an independent longitudinal study of nasopharyngeal swabs revealed that the B.1.1.7 variant was associated with longer infection times and yet showed similar peak viral loads to non-B.1.1.7 variants (Kissler et al., 2021). Here, the authors hypothesised the extended duration of 'virus shedding' may contribute to the increased transmissibility. Therefore, the mechanism for increased viral transmission could be in part attributable to the delayed replication of B.1.1.7 as shown in my data. A slower rate of replication may allow the virus to persist for longer periods before host antiviral sensors are triggered. However, 'in-bulk' RT-qPCR analysis does not provide information on the frequency of 'super-permissive' cells, which could lead to bias towards this susceptible cell type. It would be interesting to follow up on how smFISH at 48 hpi or beyond compares with the RT-qPCR measurements from the longitudinal studies. B.1.1.7 contains multiple mutations in the subgenomic portion, in particular, a new transcription regulatory core sequence (TRS) that gives rise to novel ORF9b sgRNA (Parker et al., 2021). Thorne et al. (2021) reported that elevated expression of the sgRNA encoding the innate immune antagonist ORF9b (also supported by Dr Aino Jarvelin's RNA-seq analysis of this study) gives B.1.1.7 a competitive advantage to evade host antiviral responses. Together with mutations in the Spike that are proposed to improve cell entry and the initial delay in accumulation of vRNA load, these features could provide B.1.1.7 with an advantage over the ancestral strains to enable its rapid dissemination across the human population.

6.2.1 Can smFISH be utilised in clinical settings?

The ease and sensitivity of smFISH can provide many opportunities for clinical applications, which can complement the widely used PCR and lateral flow tests. Particularly, recent technological developments that prioritise speed and convenience, enable a full round of smFISH to be completed in 13-20 minutes with minimal impact on sensitivity, but some impact on morphological preservation (Hepp et al., 2021; Shaffer et al., 2013). This will improve the capacity for high-throughput diagnosis or studies of many patient samples to characterise the presence of low-level

viral RNA below the detection threshold of other methods. The single-molecule sensitivity of the method will be particularly useful to elucidate whether the long-haul COVID (also known as post-COVD-19 syndrome) is due to lingering and stable viral RNA circulating in the body (Maccio et al., 2022; Mehandru & Merad, 2022). Moreover, the multiplexing ability of smFISH could be exploited to study whether the production of SARS-CoV-2 sgRNAs and negative-strands are biomarkers for the infection stage or disease severity (Alexandersen et al., 2020; Parker et al., 2021).

In fact, this methodology is not limited to tracking the infection dynamics of RNA viruses. In Chapter 5, I also described the application of smFISH to detecting the pgRNA precursor of a DNA virus, HBV. The single-molecule capacity offers new exciting avenues to investigate the complex dynamics of pgRNA production from cccDNA in acute infections or from the integrated virus associated with chronic infections (Yuen et al., 2018). Furthermore, recent developments of DNA-FISH are becoming increasingly compatible with RNA-FISH methods (Brown et al., 2022; Rouhanifard et al., 2018), which were often hampered by the harsh hybridisation condition requirements for the DNA-FISH. The combination of these methods holds great potential to provide an unprecedented insight into the vRNA and DNA replication processes of DNA-virus and retrovirus infections.

6.3 Conclusion

In conclusion, post-transcriptional regulation plays a widespread and significant role in *Drosophila* neurogenesis. However, the contribution of such mRNA metabolic control should be considered as a complement to transcriptional regulation, rather than an alternative. For example, Syp regulates the stability of $pros^{long}$ but it relies crucially on the 3'UTR isoform choice at the level of transcription. Similarly, in order for Syp to promote translation of *e93*, first transcriptional activation must occur and be integrated with post-transcriptional regulation of mRNA in the cytoplasm. The reciprocity of transcriptional and post-transcriptional regulation allows a finely tuned programme of NSC division and fate patterning that are immensely complex.

In response to the COVID-19 pandemic, I have successfully developed a quantitative viral smFISH assay and demonstrated its utility to detect SARS-CoV-2 RNA soon after host entry and follow its replication dynamics. Replication fitness of the virus is defined by the relationship of the virus with its host cell soon after the virus entry (Boersma et al., 2020; Singer et al., 2021). Therefore, studying the initial replication events and their impact on host antiviral responses is critical for our understanding of the success of virus variants. This study contributes to our understanding of SARS-CoV-2 RNA replication. The experimental methods and the key findings will be beneficial to future studies aiming to elucidate the molecular regulation of SARS-CoV-2 RNA metabolism.

Appendices

A

List of smFISH probe sequences

Table A.1: smFISH probe sequences

syp exon	imp exon	pros exon	pros intron	pros long UTR
cattaccttccgccattt	ttcagaccaacagcagct	tactactactgctattggcc	gtcttacgtctagatctaga	gttccggattttcattactc
tctggttaatgtcatcca	ctgccctcgaactcgaca	ggtgtgcttgaattgttatt	tgaatgctgctgctgaatgg	actcacctgactgacttatc
catcgccacggtcatcgg	ctggttcttgtccagctg	cgaatgggtggcaagagcag	ctcaagcgcgacgtaacgaa	aagcgaataccctctatctt
gcagcttgggataatcct	ctggttacgctggctacg	aaactggtgagttgctcgtg	cgcccacaaaacttaaaagt	taccgctcgaaaaatgctgc
ttcttgtccagaccgtat	gcatacccggatacggat	acaaggacgatgccgaactg	atttgtaacaaaagcccccg	gctttgagtttatgtcctta
gatctcatccagtttgcc	tgcgcagcggaaagtccg	accaaacaggttgccaaagg	cagtctcgaatttgggcaat	actagagacttttgctgctt
tcggcgtgagccaacttg	ccatttcgctctgcacaa	caaacagctcgttcagcatc	aagtcgagggtttcgatagg	cacttggcttatatggtgtg
aaactccttgagcgcgtc	gtcgaccaatgatggcgc	atggcgttgtccaaactctg	tcctgcgactacttcttgag	tacctatgctgacattcagt
attcaaggcaccatcgac	atcgtcctgatggtgctg	cgatcaacagctcggcactg	aatgacagccgaatggtgtt	gttccgtttttgtttttgg
ttccaggaactgtcccaa	cgagcacggctctgttgt	acttggacgtggaattgagg	tcacggactgacagttggag	tcctgctctgaaatctgtac
tgacacgtgctccaggtt	ctccttacgatggacgtc	ggagcaatgctgttgttatt	tttggcaacttcttgcgcaa	ctctacgtatttctaagccg
gcataggtaggcggactt	tttctccagcgatccaac	gcgagatcgatgcatttgtg	tagtcgttgtctttctgctc	tcttttgcaaaatgcagcgc
gtcggtacgtcttcatca	ccggattgccatagatcg	gatgcgagtgactggaactg	tatctggcagctgtcgtttt	gtgtgcaagagttctcattc
tgttggctggctcgactc	ttacaggcgttggtgcag	caaggagcgatcactgcagg	cgaatttacccacaccacgg	aaagtcttagccgagagtgt
gacctttgacagttgcgg	gctgcataacttccagga	tgttgcaactgagactggtg	cttccacggaataaccgtat	tgtgagcgaatgtgtgtcat
tcttgatcttgtcctcgt	tattcgtagagatggcct	atattgcgcagcatgtgatg	ctgttccattgagaacagcc	gtggccgaaactgcaattta
ctaatgtgtagccggtgc	gggcgagtattttgaggc	attgatactctccattgtgg	cctacctcacttcactttgg	gtttttgttataccgctaa
cgccgtatttacgctgac	cttgccaatgattcgtcc	ccatttccgaatcgtcgatc	caaatcacggcgactgcatg	gcagtccagttttcaatttt
cacatttccctcccaatg	ccgtttaatggtattgcc	gctcactcttgatcttgatg	tcggtatggtattcacttcg	actgatcttgtctgcttttc
aaacctcgcaaccgttgc	atcttcgtatcggtgtcc	gattctttttggtgggcag	cagctcggttctctgttttc	ggccgagaggtttttaattt
tgtccttgggtatcttgc	gtcgttgatagagctgac	acctcgtcaagcatatcgtc	aaatccaggtgagaccttgt	cctgaatgtttacaccgaga
agcggaatcagttcgtcc	gctccaaattgaagctgt	caggctatccatatcggatc	ccgacgatgaaagagggttg	tcaaagttttccgttttcca
attatgccgcagttctcg	atcccttgaccgtaatga	catcatgtccgagtgactgg	gacaaatcggagacgggcag	ttgttatcgctttgtttcgg
atcgggtccatcatgagt	cagcacgcgacatgttct	atgttttctgctccacgcaa	ttaagacgttgcgcatgtct	aaggttaactacccggagat
taaccacgatttgtgccc	cgcagcttggtgctaatt	gtttcttgagacatccgctg	agggtgtcttgcggacaatt	ttcgtattatcctttccttg
gcttcgcgatttgtgaat	gcagatcgttctcgtagc	cacgatgttctcgacgcgag	tgcgcaggtaggggaatttc	gttgtcgactagtcatgttc
tgtcgcactgcattgacg	aacatgaggctctgcggc	ctaatccggaggatggactg	gatttgttgtctttgtgct	tttttccaattagttcctgc
ttacctcgtataggccag	gccatcggatggagaccg	atttacctgcagttgtcctg	gaaattccgaacaaaccccg	catacactttccctatgtg
tatcatctggcgaactgt	ttgcccggtgtcgacatc	agagcttgcgcttcttacaa	acatcgcgcaagggtagcaa	tcaactcaagtagtgtcctc
gcaaaagccgcgattctt	acggcatgctcgtattga	cgaaattcagaccagctgcg	ccatgcaaagtagggttagg	atggggttaggattcatagc
gccttgtgtgactcgtac	ggcaaagctttgacacga	atcatgctctgcagattgag	aagacaagatgccaactccc	tcgatcgttggcaaatctca
cttcgtttggccaaagac	acactagccggagtcttg	agctcattggactcactgtc	aagaagtcgccgaagtcgtc	gtttcttggggaaggaactc
ttaattcttcctgtgcca	attggggaagaccggtgg	acacgcttctgttggatttg	atattcccacccatttttaac	ggtttattttcgaggcttga
ggcccagtcgactattat	gataggtggtttcctgta	tgggacttgagcgcattttt	cgagtaccagttaggaaatc	aaacggatctggatgtcctg
cattgtttgctcatccgg	cagcgttgttgggtatgt	acatatttctgctgcatctc	aaaccgttacatcccaacta	tgtttgtattgccttgagtg
cctgggtaagatttcgca	tggtgccgataatggcgc	cagctcgatatgatcactgc	cttagcaactcagcatgaga	atgttaccgaaatccttggt
tcagtttatcctctgaga	tgcttcggatatgcgagc	ttgtggttgggactaacgtc	ggaaaagcattttcccgaga	tatggagacgtggtcgcaaa
cgttccacttttccgtat	atgcgttcgagaagcgca	atggtgagggtgagctggaa	ctaggttaatttaggggggaa	cttttcacttgctcaactcg
gacggcgctatcacgatc	tcgaggggggctatttta	agcgatgtcttgggttttag		ctttagatcggacgttgtgg
ccattaaggccacgcata	cttgcgctccgtttgttg	atagcatgttagcgcctgag		cgcgttggttttagttacat
atattcgaggcgccgatc	cggtgtgccaacaatcgt	gacatcatcttgctcatcat		gtcatcgcttagtttttgtt
gggggtttggctagagag	atcatgtactgcgccttc	ctggagcagtggaggaaatc		agcagttatggtttttgctt
cctccttttttcttttgt	tcttcgcgcatcttctcg	aagaactgctggtacatggc		tttgcatttttagcccatat
gctcacgagcacgcagaa	tccgtgccacacatgaag	agcatttggaacttttccga		ggcttctaagagatttgcga
gacgcgcttgcatcattt	tcaaccgtgaggcgcaca	cgcatcattgagctattgtt		ttcatctcgtatcgttcttc
aggatgtgtcggcgacag	tgagagctggccacaagc	atctcggacttgagcacatc		atctacggtaggtgtatgtc
catgggcgtcaaggacat		gacgaagcgggtcacaatgg		ctaaaggttggccgaaacgg
tattcacggggtatcggc				aatgttaccggcactattgt

Table A.2: smFISH probe sequences, continued

e93 exon	e93 eUTR	GFP	YFP
atcctcttcactgtacgc	tgggccttgttttccgac	cggtgaacagctcctcgc	cggtgaacagctcctcgc
cgggcatctagcttgttg	agtttgctctccgacagc	cagctcgaccaggatggg	cagctcgaccaggatggg
gaatggagcgctgctcat	ttgcccgtttccttgacg	ctgaacttgtggccgttt	ctgaacttgtggccgttt
cttgaacagccggttgtc	tgggttggagccgaaaca	ttgccgtaggtggcatcg	ttgccgtaggtggcatcg
tcctggtcatggtgtttc	ggggttttgcctttgcag	ccggtggtgcagatgaac	gtggtgcagatcagcttc
ctcgtcgccatcatgatc	ttgttgcggtcgctcaag	agggtggtcacgagggtg	agggtggtcacgagggtg
catcatcgttggagtcct	ggcatcggcaaagaaagc	aagcactgcacgccgtag	aagcactgcaggccgtag
tgtcgacgcattgctgtc	tcgagtctaacaacgcgc	atgtggtcggggtagcgg	atgtggtcggggtagcgg
aactctgccggatacacc	ggttttcccttctcgcca	ttgaagaagtcgtgctgc	ttgaagaagtcgtgctgc
ctcagtttgcgcagttgt	tacctgagcgttttcgcc	acgtagccttcgggcatg	acgtagccttcgggcatg
cattgtgctcggacaggt	aagtagggccaaagtgcg	aagaagatggtgcgctcc	aagaagatggtgcgctcc
tttcggcgatccacgatc	ttttcggggccaactgca	tcttgtagttgccgtcgt	tcttgtagttgccgtcgt
gattggcactggcattgc	gggccattagcttgtgct	tcgaacttcacctcggcg	tcgaacttcacctcggcg
ttggcatccagcggaatg	tcgaaaggagggccaaag	atgcggttcaccagggtg	atgcggttcaccagggtg
agcatcagagtgtgcagc	gattgggggggggtttaac	gaagtcgatgcccttcag	gaagtcgatgcccttcag
cttcggcattgcaccaat	ttttggccctgtctgagc	caggatgttgccgtcctc	caggatgttgccgtcctc
ccttgataaagtcgccca	tacactccggggcacaaa	gttgtactccagcttgtg	gttgtactccagcttgtg
ttatgccaccactgttgg	gtcaacgccaaagcgaaa	atagacgttgtggctgtt	atagacgttgtggctgtt
agatttagcagtccctcg	tacgattcctttggctcg	cttctgcttgtcggccat	cgttcttctgcttgtcgg
aggagttcgtttccggag	gagcaaatattggccacg	gaagttcaccttgatgcc	tcttgaagttggccttga
atgggcgatgcggaatgc	acgatttcctgtttccgg	ctcgatgttgtggcggat	cgtcctcgatgttgtggc
ggtcagcaaactgttgcc	cttggctatgatcgatcc	tgttctgctggtagtggt	tgttctgctggtagtggt
tattgctggccatcatct	tggttgctgctgttgttg	ttgtcgggcagcagcacg	ttgtcgggcagcagcacg
ttcgttgatctggcgctg	ttgttatcacgtctgggc	gactgggtgctcaggtag	ggactggtagctcaggta
tttccgttcctgagactc	gcatatgcttagccaacg	ttggggtctttgctcagg	gttggggtctttgctcag
ttgctgctgcaatcgcta	ttttggtggttaggcgac	accatgtgatcgcgcttc	gaccatgtgatcgcgctt
ttcttgtatcccagcgag	tgctttgtctcgctttgc	ggcggtcacgaactccag	ggcggtcacgaactccag
caatgatcttggccaccg	tgagaatcccacttggca	cttgtacagctcgtccat	tccatgccgagagtgatc
acagcagattgggcgagg	cacatacgattttccggg		
agccgaatggtgctgttg	tcaaaacttttgcctggc		
ccaaactgtcgcggtcat	tttctgtgtgtgtctgtg		
accgacatttcacctctc	acatctcttatgtccacg		
gtacgccgtagtagctac	gatgtctcgaactgagca		
ttgaccttgtactccagc	taggcgtagtctaattgc		
cttcagtttgtccagctg	atacggtagacatatgcg		
tcttgagggcattgctca	tctccatgggcaatggaa		
tggggccagaacatgttc	atcgatcttgtcttctgc		
attgaagtccaggccgta	tttcagtgatgtgtgccg		
tgggcactcttcatcagg	actctcgcctattttccc		
cctgatgatgccatcgta	ttgccataccttgactg		
cagcagctgatcgagcag	gtctatgtcgaggacaac		
gcgtagtcattgttccga	atgtgcgaaaatcgcagc		
ccgcttgatgtctgcata	gtgtgtgaacggagagcc		
cttgttcttgttgtgcgc	tggactgaagcgtgcgca		
cgtttcggaatcatcccg	gattgcccgtattgcggg		
caggactgtgatgtccgg	cttggccaaatgcagcgc		
gtatggacgtggtgtggt	gttcgcagaactctccgc		
tcctataactgatccgcc	attcagcagttcagcaca		

Table A.3: smFISH probe sequences, continued

ORF1a_pos	ORF1b_neg	ORFN_pos	ORFS_pos
tagatcggcgccgtaactat	gacatgagtaaatttcccct	tgattttggggtccattatc	cacactgactagagactagt
gttatcgacatagcgagtgt	ttggcaaaccacgcgaacaa	acgtaatgcggggtgcattt	cagggggtaattgagttctg
agcatccgaacgtttgatga	gtgaatgcgtcatcatctga	agttgaatctgagggtccac	acaagtcctgagttgaatgt
ccagttgttcggacaaagtg	taagctcatgggacacttcg	cgttctccattctggttact	gacagggttatcaaacctct
caaagccacgtacgagcacg	tgcttgtcgattcagatctt	cgacgttgttttgatcgcgc	gtgcaattattcgcactaga
tttcagaacgttccgtgtac	accaggtacagctgttttaa	agacgcagtattattgggta	ctaccaatggttctaaagcc
taagctcacgcatgagttca	ggtgctggttctgataaagg	tgttgagtgagagcggtgaa	ataataagctgcagcaccag
ggtgacgcaactggatagac	tgttgctatgcctaatcttt	ggaatttaaggtcttccttg	tcctaggttgaagataaccc
aggttgttctaatggttgta	ggtgtaaagatggccatgta	aattggaacgccttgtcctc	cacagtctacagcatctgta
tgagttggacgtgtgttttc	acagatgcgcaaacaggttc	tcatctggactgctattggt	agattctgttggttggactc
acaacctatgttagcgctag	ttcctatggacagtacagtt	ctcttcggtagtagccaatt	ttcaccaaaagggcacaagt
tcctttattaccgttcttac	atgccttcgaacatatcgtt	caccaccacgaattcgtctg	tgcaaatctggtggcgttaa
ttatagcggccttctgtaaa	gttctatgactgacatagcc	cttggactgagatctttcat	ggagacactccataacactt
agaagaaccttgcggtaagc	aatttgggtgtggacattgc	tttgttagcaccatagggaa	ggtttgagattagacttcct
gagcaacataagcccgttaa	tctattactctgacagtcca	cagttgcaacccatatgatg	accattacaaggtgtgctac
ctttctgtacaatccctttg	atgtcgatagatatcctgct	gattgcagcattgttagcag	tactactctgtatggttggt
ctcgtcgcctaagtcaaatg	ggtgtatgcctatttttggaa	tagaagccttttggcaatgt	aacacctgtgcctgttaaac
ttaccagcacgtgctagaag	ttatgccacacattctgaca	gagaagaggcttgactgccg	gacagcatcagtagtgtcag
caaactggtgtaccaaccaa	gctattaagtgtgtacctca	gttgcgactacgtgatgagg	acagcaacctggttagaagt
tgtaccgagttcaactgtat	gttaagcgtgttgactggac	cctggagttgaatttcttga	tgtgcagttaacatcctgat
ttgacgtgcctctgataaga	atcatgactaggtgtctagc	caggagaagttcccctactg	ctgcatgaatagcaacaggg
tagtagttgtctgattgtcc	ctgtattgtcaagtccatgg	cagcaaagcaagagcagcat	gagttgttgacatgttcagc
aagtaacctttgttggtgca	gttggcatcattctattgga	agctggttcaatctgtcaag	caatgggtatgtcacactca
ccttggttgaatagtcttga	cacatggctttgagttgaca	ctttaccagacattttgctc	taactagcgcatatacctgc
ctcatattgagttgatggct	ctctctgacagagtcgtatt	tagtggcagtacgtttttgc	gccgaggagaattagtctga
gccaatttaaactcaccaga	atttttccagagttagtgct	agcttgtgttacattgtatg	gtgacatagtgtaggcaatg
tagagtcagcacacaaagcc	ctgtacctacaggttatgtt	tttgttctggaccacgtctg	gtgggtatggcaatagagtt
acttctgtgggaagtgtttc	cagctaggtttttctacagg	ctggtccccaaaatttcctt	gcacggtttaattgtgtaca
agtagtatgtagccatactc	gacatgtacgtgcatggatt	tcagttccttgtctgattag	acttcttgggtgtttttgtc
cacaggcgaactcatttact	ggcatacctaaggacatgac	caatttgcggccaatgtttg	acagtaaggccgttaaactt
atcatattaggtgcaagggc	cacctcagtgttgacactaa	tgaagcgctgggggcaaatt	tttggttgaccacatcttga
aaggctttaagtttagctcc	gtaatcactgggttacatcc	atgcgcgacattccgaagaa	agcgtgtttaaagcttgtgc
ctgaattgtgacatgctgga	cacgtaggaatgtggcaact	ccgaaggtgtgacttccatg	ttgcaccaaaattggagcta
cccaaccgtctctaagaaac	tcatcacagggctcagaata	atggcacctgtgtaggtcaa	tgaagtctgcctgtgatcaa
tgcgggagaaaattgatcgt	gctgtagcctcaaagatttt	atttggatctttgtcatcca	gttgagtcacatatgtctgc
cccctcttagtgtcaataaa	atctgcaattaacaggccac	ttgtatgcgtcaatatgctt	gtagcagcaagattagcaga
tctacccataaagccatcaa	ctgtgagtgctttggtttat	ggtaaggcttgagtttcatc	tccaagtacacactctgaca
agtagccaaatcagatgtga	ctataggtccagacatgttc	gaagagtcacagtttgctgt	ggacataagatgatagccct
cgcacagaattttgagcagt	ccacgcacattgctaactaa	atcatccaaatctgcagcag	aagactacaccatgaggtgc
tactgaatgccttcgagttc	cactatgtgtacattggcga	ggattgttgcaattgtttgg	gacaccttcacgaggaaagt
gtgttccagttttcttgaaa	gtgttgtcaatgccagatta		tgtgttgtctgtagtaatga
ggtaatcatcttcagtacca	agaattatacctgcacgtgc		tcaggttgcaaaggatcata
gcttttagaggcatgagtag	gtgctgacatcacatacagt		ctcctccttgaatgagtcta
catagggctgttcaagttga	ctatggtgatgctgttgttt		agcattaatgccagagatgt
ataggcacacttgttatggc	aacctagaccaccacttaac		ccaagttcttggagatcgat
gtaattcagatactggttgc	tacgtgaagtgctgtctgac		aaacctagccaaatgtacca
aagccaatcaaggacgggtt	gctcaaagctactgaggaga		tatggcaatcaagccagcta
aagtgtctcaccactacgac	cttgcatacgtagaccattc		ctcagagtcgtcttcatcaa

B

Syp targets in the brain

Due to space constraints only the top 100 targets are shown. The full list can be downloaded from this link: `https://drive.google.com/file/d/1-8GrL6UdarnC0AEIGxGXozA1a5HjyF5e/view?usp=sharing`

Table B.1: Syp iCLIP targets in the L3 brain

gene_id	gene_name	gene_biotype	n.binding.region	binding.region.mean	binding.region.sum	peak.score.per.region
FBgn0262972	lncRNA:cherub	ncRNA	396	13882.59848	5497509	13882.59848
FBgn0004595	pros	protein_coding	583	1456.88851	849366	1456.88851
FBgn0038826	Syp	protein_coding	422	662.15640	279430	662.15640
FBgn0026197	lncRNA:noe	ncRNA	66	3596.83333	237391	3596.83333
FBgn0027660	blot	protein_coding	66	3596.83333	237391	3596.83333
FBgn0027339	jim	protein_coding	287	602.79443	173002	602.79443
FBgn0010113	hdc	protein_coding	305	345.84262	105482	345.84262
FBgn0265297	pAbp	protein_coding	87	1058.25287	92068	1058.25287
FBgn0022893	Df31	protein_coding	60	1387.05000	83223	1387.05000
FBgn0266937	lncRNA:CR45388	ncRNA	250	280.91600	70229	280.91600
FBgn0065099	snRNA:7SK	snRNA	9	6361.33333	57252	6361.33333
FBgn0010300	brat	protein_coding	296	180.25000	53354	180.25000
FBgn0004838	Hrb27C	protein_coding	172	291.51163	50140	291.51163
FBgn0014163	fax	protein_coding	56	859.87500	48153	859.87500
FBgn0063127	lncRNA:CR33938	ncRNA	14	2548.64286	35681	2548.64286
FBgn0283451	br	protein_coding	175	195.25714	34170	195.25714
FBgn0003261	Rm62	protein_coding	52	573.07692	29800	573.07692
FBgn0031453	Bacc	protein_coding	62	448.56452	27811	448.56452
FBgn0020238	14-3-3epsilon	protein_coding	48	577.66667	27728	577.66667
FBgn0000308	chic	protein_coding	46	586.76087	26991	586.76087
FBgn0086675	fne	protein_coding	155	173.00645	26816	173.00645
FBgn0263396	sqd	protein_coding	149	174.02685	25930	174.02685
FBgn0283521	lola	protein_coding	276	91.81522	25341	91.81522
FBgn0263392	Tet	protein_coding	110	214.43636	23588	214.43636
FBgn0022959	yps	protein_coding	60	385.93333	23156	385.93333
FBgn0001105	Gbeta13F	protein_coding	87	240.71264	20942	240.71264
FBgn0286516	aqz	protein_coding	95	216.12632	20532	216.12632
FBgn0004828	His3.3B	protein_coding	34	592.41176	20142	592.41176
FBgn0086758	chinmo	protein_coding	143	140.36364	20072	140.36364
FBgn0003435	sm	protein_coding	181	106.35359	19250	106.35359
FBgn0000253	Cam	protein_coding	69	269.02899	18563	269.02899
FBgn0265276	l(3)neo38	protein_coding	67	268.92537	18018	268.92537
FBgn0260400	elav	protein_coding	102	175.18627	17869	175.18627

Table B.1: Syp iCLIP targets in the L3 brain *(continued)*

gene_id	gene_name	gene_biotype	n.binding.region	binding.region.mean	binding.region.sum	peak.score.per.region
FBgn0286870	lncRNA:enrr	ncRNA	102	175.18627	17869	175.18627
FBgn0066084	RpL41	protein_coding	8	1999.12500	15993	1999.12500
FBgn0019661	lncRNA:roX1	ncRNA	42	343.90476	14444	343.90476
FBgn0001234	lncRNA:Hsromega	ncRNA	146	98.74658	14417	98.74658
FBgn0266599	Hsc70-4	protein_coding	12	1191.50000	14298	1191.50000
FBgn0040007	RpL38	protein_coding	16	889.87500	14238	889.87500
FBgn0037874	Tctp	protein_coding	23	608.43478	13994	608.43478
FBgn0263352	Unr	protein_coding	115	121.53913	13977	121.53913
FBgn0036165	chrb	protein_coding	94	147.28723	13845	147.28723
FBgn0262743	Fs(2)Ket	protein_coding	126	108.92063	13724	108.92063
FBgn0029979	mahe	protein_coding	115	113.86087	13094	113.86087
FBgn0262739	AGO1	protein_coding	97	134.83505	13079	134.83505
FBgn0041094	scyl	protein_coding	62	193.91935	12023	193.91935
FBgn0028999	nerfin-1	protein_coding	38	311.31579	11830	311.31579
FBgn0053111	CG33111	protein_coding	94	125.24468	11773	125.24468
FBgn0260634	eIF4G2	protein_coding	94	125.24468	11773	125.24468
FBgn0003165	pum	protein_coding	103	112.94175	11633	112.94175
FBgn0264490	Eip93F	protein_coding	127	90.79528	11531	90.79528
FBgn0002645	Map205	protein_coding	58	197.27586	11442	197.27586
FBgn0004907	14-3-3zeta	protein_coding	39	282.61538	11022	282.61538
FBgn0065098	RNaseMRP:RNA	ncRNA	7	1567.57143	10973	1567.57143
FBgn0015778	rin	protein_coding	64	170.18750	10892	170.18750
FBgn0262734	eIF4H1	protein_coding	58	186.86207	10838	186.86207
FBgn0052423	shep	protein_coding	94	114.48936	10762	114.48936
FBgn0011766	E2f1	protein_coding	62	168.79032	10465	168.79032
FBgn0266410	CG45050	protein_coding	94	111.23404	10456	111.23404
FBgn0013981	His4r	protein_coding	14	744.57143	10424	744.57143
FBgn0000108	Appl	protein_coding	56	184.53571	10334	184.53571
FBgn0264922	smt3	protein_coding	16	621.37500	9942	621.37500
FBgn0262737	mub	protein_coding	90	108.57778	9772	108.57778
FBgn0039858	CycG	protein_coding	56	171.55357	9607	171.55357
FBgn0011935	tRNA:Asn-GTT-1-7	tRNA	4	2253.50000	9014	2253.50000
FBgn0262867	Ptr	protein_coding	4	2253.50000	9014	2253.50000
FBgn0027108	Inx2	protein_coding	66	136.28788	8995	136.28788
FBgn0037697	GstZ2	protein_coding	34	262.64706	8930	262.64706
FBgn0034802	CNBP	protein_coding	27	329.55556	8898	329.55556
FBgn0011817	nmo	protein_coding	89	99.56180	8861	99.56180
FBgn0284243	betaTub56D	protein_coding	14	632.92857	8861	632.92857
FBgn0039830	ATPsynC	protein_coding	24	358.62500	8607	358.62500
FBgn0000557	eEF1alpha2	protein_coding	23	371.47826	8544	371.47826
FBgn0004646	ogre	protein_coding	42	202.97619	8525	202.97619
FBgn0004432	Cyp1	protein_coding	23	368.08696	8466	368.08696
FBgn0001215	Hrb98DE	protein_coding	38	217.50000	8265	217.50000
FBgn0029687	Vap33	protein_coding	48	169.43750	8133	169.43750
FBgn0000042	Act5C	protein_coding	19	428.00000	8132	428.00000
FBgn0026250	eIF1A	protein_coding	20	386.60000	7732	386.60000
FBgn0020496	CtBP	protein_coding	75	102.45333	7684	102.45333
FBgn0010382	CycE	protein_coding	68	111.57353	7587	111.57353
FBgn0021895	ytr	protein_coding	24	316.00000	7584	316.00000
FBgn0010348	Arf79F	protein_coding	31	237.87097	7374	237.87097
FBgn0011666	msi	protein_coding	90	81.86667	7368	81.86667
FBgn0027066	Eb1	protein_coding	31	236.58065	7334	236.58065
FBgn0000116	Argk	protein_coding	38	191.23684	7267	191.23684
FBgn0011760	ctp	protein_coding	61	119.08197	7264	119.08197
FBgn0004108	Nrt	protein_coding	39	185.43590	7232	185.43590
FBgn0000546	EcR	protein_coding	38	189.92105	7217	189.92105
FBgn0001235	hth	protein_coding	94	76.52128	7193	76.52128
FBgn0267516	pre-rRNA:CR45856	rRNA	13	551.38462	7168	551.38462
FBgn0283477	SF2	protein_coding	36	197.05556	7094	197.05556
FBgn0015320	Ubc2	protein_coding	36	194.77778	7012	194.77778
FBgn0003930	snRNA:U4:39B	snRNA	6	1162.00000	6972	1162.00000
FBgn0032935	Atg18b	protein_coding	6	1162.00000	6972	1162.00000
FBgn0025741	PlexA	protein_coding	77	89.68831	6906	89.68831
FBgn0267522	28SrRNA-Psi:CR45862	pseudogene	11	626.81818	6895	626.81818
FBgn0263594	lost	protein_coding	23	290.78261	6688	290.78261
FBgn0035245	GC	protein_coding	37	179.75676	6651	179.75676
FBgn0035246	CG13928	protein_coding	37	179.75676	6651	179.75676

C

Syp targets in the whole larva

Due to space constraints only the top 100 targets are shown. The full list can be downloaded from this link: `https://drive.google.com/file/d/1--tibATx1Phuk3iN1D2REaHExiKc_k4a/view?usp=sharing`

Table C.1: Syp iCLIP targets in the whole L3 larva

gene_id	gene_name	gene_biotype	n.binding.region	binding.region.mean	binding.region.sum	peak.score.per.region
FBgn0262972	lncRNA:cherub	ncRNA	298	1301.1913	387755	1301.1913
FBgn0087002	apolpp	protein_coding	315	1017.9175	320644	1017.9175
FBgn0065099	snRNA:7SK	snRNA	13	23586.3077	306622	23586.3077
FBgn0038826	Syp	protein_coding	400	658.2250	263290	658.2250
FBgn0265297	pAbp	protein_coding	87	1977.3908	172033	1977.3908
FBgn0004595	pros	protein_coding	370	453.8270	167916	453.8270
FBgn0085195	CG34166	protein_coding	17	8918.4706	151614	8918.4706
FBgn0002565	Lsp2	protein_coding	16	8164.0000	130624	8164.0000
FBgn0283451	br	protein_coding	367	322.2316	118259	322.2316
FBgn0000308	chic	protein_coding	56	1159.1071	64910	1159.1071
FBgn0000416	Sap-r	protein_coding	18	3524.4444	63440	3524.4444
FBgn0263396	sqd	protein_coding	180	339.1167	61041	339.1167
FBgn0003377	Sgs7	protein_coding	9	5504.4444	49540	5504.4444
FBgn0002564	Lsp1gamma	protein_coding	37	1302.5405	48194	1302.5405
FBgn0038294	Mf	protein_coding	99	486.7071	48184	486.7071
FBgn0004838	Hrb27C	protein_coding	149	320.7248	47788	320.7248
FBgn0002562	Lsp1alpha	protein_coding	26	1834.8846	47707	1834.8846
FBgn0001234	lncRNA:Hsromega	ncRNA	203	231.6897	47033	231.6897
FBgn0035917	Zasp66	protein_coding	81	570.8025	46235	570.8025
FBgn0001224	Hsp23	protein_coding	9	5089.1111	45802	5089.1111
FBgn0266410	CG45050	protein_coding	150	281.9267	42289	281.9267
FBgn0010113	hdc	protein_coding	215	195.2558	41980	195.2558
FBgn0022893	Df31	protein_coding	49	848.4490	41574	848.4490
FBgn0001218	Hsc70-3	protein_coding	31	1277.5806	39605	1277.5806
FBgn0031228	ND-15	protein_coding	63	602.4762	37956	602.4762
FBgn0001142	Gs1	protein_coding	62	610.4355	37847	610.4355
FBgn0286516	aqz	protein_coding	130	290.5538	37772	290.5538
FBgn0005666	bt	protein_coding	280	132.9893	37237	132.9893
FBgn0027660	blot	protein_coding	50	741.6200	37081	741.6200
FBgn0027339	jim	protein_coding	205	180.6780	37039	180.6780
FBgn0026197	lncRNA:noe	ncRNA	48	769.0833	36916	769.0833
FBgn0014163	fax	protein_coding	59	599.4237	35366	599.4237
FBgn0000639	Fbp1	protein_coding	10	3330.0000	33300	3330.0000

Table C.1: Syp iCLIP targets in the whole L3 larva *(continued)*

gene_id	gene_name	gene_biotype	n.binding.region	binding.region.mean	binding.region.sum	peak.score.per.region
FBgn0004907	14-3-3zeta	protein_coding	44	750.1136	33005	750.1136
FBgn0000042	Act5C	protein_coding	57	568.0702	32380	568.0702
FBgn0033725	Cpr49Ac	protein_coding	49	660.0816	32344	660.0816
FBgn0010415	Sdc	protein_coding	84	373.0714	31338	373.0714
FBgn0037874	Tctp	protein_coding	24	1304.9583	31319	1304.9583
FBgn0058469	lncRNA:CR40469	ncRNA	5	6045.0000	30225	6045.0000
FBgn0039858	CycG	protein_coding	72	417.6528	30071	417.6528
FBgn0001149	GstD1	protein_coding	11	2717.5455	29893	2717.5455
FBgn0066084	RpL41	protein_coding	10	2875.8000	28758	2875.8000
FBgn0019661	lncRNA:roX1	ncRNA	48	579.3958	27811	579.3958
FBgn0262736	Vha16-1	protein_coding	47	589.9149	27726	589.9149
FBgn0014141	cher	protein_coding	36	766.3056	27587	766.3056
FBgn0261822	Bsg	protein_coding	109	252.8257	27558	252.8257
FBgn0284408	trol	protein_coding	88	308.4886	27147	308.4886
FBgn0003261	Rm62	protein_coding	56	469.0536	26267	469.0536
FBgn0265487	mbl	protein_coding	30	874.2667	26228	874.2667
FBgn0015222	Fer1HCH	protein_coding	24	1092.4583	26219	1092.4583
FBgn0000546	EcR	protein_coding	103	248.6408	25610	248.6408
FBgn0261238	Alh	protein_coding	52	483.2692	25130	483.2692
FBgn0014863	Mlp84B	protein_coding	49	508.3265	24908	508.3265
FBgn0266599	Hsc70-4	protein_coding	13	1797.0000	23361	1797.0000
FBgn0040007	RpL38	protein_coding	17	1350.5882	22960	1350.5882
FBgn0004687	Mlc-c	protein_coding	20	1138.8500	22777	1138.8500
FBgn0026084	cib	protein_coding	23	941.9130	21664	941.9130
FBgn0003721	Tm1	protein_coding	84	252.8810	21242	252.8810
FBgn0000261	Cat	protein_coding	17	1242.2941	21119	1242.2941
FBgn0034398	CG15098	protein_coding	15	1327.8667	19918	1327.8667
FBgn0052672	Atg8a	protein_coding	28	710.7143	19900	710.7143
FBgn0004828	His3.3B	protein_coding	33	601.6061	19853	601.6061
FBgn0266446	CG45076	protein_coding	74	262.9459	19458	262.9459
FBgn0052062	Rbfox1	protein_coding	63	308.2540	19420	308.2540
FBgn0032949	Lamp1	protein_coding	20	967.5000	19350	967.5000
FBgn0002772	Mlc1	protein_coding	18	1059.1111	19064	1059.1111
FBgn0085446	CG34417	protein_coding	57	330.9474	18864	330.9474
FBgn0003162	Pu	protein_coding	24	784.5417	18829	784.5417
FBgn0027066	Eb1	protein_coding	29	645.8276	18729	645.8276
FBgn0001105	Gbeta13F	protein_coding	85	216.5647	18408	216.5647
FBgn0002535	Lcp4	protein_coding	8	2284.3750	18275	2284.3750
FBgn0026250	eIF1A	protein_coding	18	1010.8333	18195	1010.8333
FBgn0000064	Ald1	protein_coding	52	349.6538	18182	349.6538
FBgn0004169	up	protein_coding	18	1010.0556	18181	1010.0556
FBgn0286818	Pdi	protein_coding	15	1198.0667	17971	1198.0667
FBgn0005671	Vha55	protein_coding	35	511.6000	17906	511.6000
FBgn0036030	CG6767	protein_coding	47	379.2979	17827	379.2979
FBgn0283535	Vha26	protein_coding	51	344.2353	17556	344.2353
FBgn0011224	heph	protein_coding	136	128.5515	17483	128.5515
FBgn0052423	shep	protein_coding	115	150.4174	17298	150.4174
FBgn0262737	mub	protein_coding	101	164.4851	16613	164.4851
FBgn0052064	S-Lap4	protein_coding	28	593.0000	16604	593.0000
FBgn0027835	Dp1	protein_coding	55	300.7273	16540	300.7273
FBgn0034903	sona	protein_coding	40	410.7750	16431	410.7750
FBgn0002645	Map205	protein_coding	73	224.3562	16378	224.3562
FBgn0045823	vsg	protein_coding	47	344.7021	16201	344.7021
FBgn0000116	Argk	protein_coding	15	1075.2000	16128	1075.2000
FBgn0041094	scyl	protein_coding	53	304.1132	16118	304.1132
FBgn0261836	Msp300	protein_coding	112	143.4911	16071	143.4911
FBgn0285952	eEF5	protein_coding	22	728.0909	16018	728.0909
FBgn0000043	Act42A	protein_coding	12	1327.2500	15927	1327.2500
FBgn0086758	chinmo	protein_coding	78	203.6154	15882	203.6154
FBgn0000253	Cam	protein_coding	65	244.0462	15863	244.0462
FBgn0000370	crc	protein_coding	61	259.3770	15822	259.3770
FBgn0284245	eEF1alpha1	protein_coding	11	1432.0909	15753	1432.0909
FBgn0053113	Rtnl1	protein_coding	36	436.7222	15722	436.7222
FBgn0036921	RhoGDI	protein_coding	31	503.8387	15619	503.8387
FBgn0015766	Msr-110	protein_coding	25	609.3200	15233	609.3200
FBgn0014184	Oda	protein_coding	35	432.8571	15150	432.8571
FBgn0040732	CG16926	protein_coding	15	978.1333	14672	978.1333

D

Summary of smFISH protein and RNA screen results (Syp targets)

Result of protein and RNA distribution survey in Syp targets.

Summary of mRNA and protein distribution survey in Syp target protein-trap lines. Presence of mRNA or Protein expression are depicted as 1-present or 0-absent. Gene expression was scored in multiple nervous system cell types and synaptic compartments.

D. Summary of smFISH protein and RNA screen results (Syp targets)

			Mushroom body						Optic lobe				Ventral nerve cord				Neuroblast				Progeny/Neuron					
			Soma		Calyx		Lobe		Soma		Neuropil		Soma		Neuropil		VNC		Central brain		VNC		Central brain		Discordance	
Gene ID	Gene name	Protein-trapID	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	RNA	Protein	NB lineage	Any
FBgn0003435	**sm**	CPTI002828	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	0	1	0	1	1	1	1	TRUE	TRUE
FBgn0010113	**hdc**	MiMIC00406	1	0	0	0	0	0	1	1	0	0	1	1	0	0	1	0	1	0	1	1	1	1	TRUE	TRUE
FBgn0085387	**shakB**	MiMIC02168	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	FALSE	FALSE
FBgn0264001	**bru-3**	MiMIC02379	1	1	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	FALSE	TRUE
FBgn0264490	**eip93f**	MiMIC05200	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0	1	0	TRUE	TRUE
FBgn0011666	**msi**	MiMIC00977	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	0	0	0	1	0	1	0	TRUE	TRUE
FBgn0267336	**glut4ef**	MiMIC07274	1	1	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	FALSE	TRUE
FBgn0261822	**bsg**	CPTI100062	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	FALSE	FALSE
FBgn0261552	**ps**	CPTI001261	1	0	1	0	1	0	1	0	1	0	1	1	1	0	1	0	1	0	1	0	1	0	TRUE	TRUE
FBgn0004838	**hrb27c**	Davislab	0	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	FALSE	TRUE
FBgn0264006	**dysc**	MiMIC02729	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	TRUE	TRUE
FBgn0000253	**cam**	CPTI001284	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	FALSE	TRUE
FBgn0040765	**luna**	MiMIC05859	1	1	0	1	0	0	1	1	0	1	1	1	0	1	0	0	1	1	1	1	1	1	TRUE	TRUE
FBgn0011206	**bol**	CPTI004519	1	0	1	1	1	1	1	0	1	0	1	0	1	1	1	0	1	0	1	0	1	0	TRUE	TRUE
FBgn0264270	**sxl**	BDSC51532	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	FALSE	FALSE
FBgn0263396	**sqd**	2017	1	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	TRUE	TRUE
FBgn0285926	**Imp**	BDSC60237	1	1	1	0	1	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	FALSE	TRUE
FBgn0038826	**Syp**	Titlowetal. 2020	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	TRUE	TRUE
FBgn0262656	**myc**	BDSC81274	0	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	1	1	0	0	0	0	FALSE	FALSE
FBgn0010313	**corto**	fTRG318889	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	1	1	1	0	1	0	TRUE	TRUE
FBgn0010313	**jim**	BDSC42275	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	0	1	0	1	1	1	1	TRUE	TRUE
FBgn0004880	**scrt**	fTRG318073	1	1	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	FALSE	FALSE
FBgn0265487	**mbl**	BDSC59758	1	1	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0	TRUE	TRUE
																									TOTAL 13	TOTAL 18

E

Summary of smFISH protein and RNA screen results (200 YFP-traps)

Summary annotation of 200 YFP-trap screen result.

Summary heatmap of mRNA and protein distribution screen in 200 YFP-trap fly lines. Discordant mRNA and protein expression pattern is represented in dark purple, while concordance is shown in light pink. Each row represent a gene and each column represents different nervous system cell types or synaptic compartments. The presented visual was created by Maria Kiourlappou.

E. Summary of smFISH protein and RNA screen results (200 YFP-traps)



References

Acheampong, K. K., Schaff, D. L., Emert, B. L., Lake, J., Reffsin, S., Shea, E. K., Comar, C. E., Litzky, L. A., Khurram, N. A., Linn, R. L., Feldman, M., Weiss, S. R., Montone, K. T., Cherry, S., & Shaffer, S. M. (2021). *Multiplexed detection of SARS-CoV-2 genomic and subgenomic RNA using in situ hybridization* (p. 2021.08.11.455959). bioRxiv. https://doi.org/10.1101/2021.08.11.455959

Aldridge, S., & Teichmann, S. A. (2020). Single cell transcriptomics comes of age. *Nature Communications*, *11*(1), 4307. https://doi.org/10.1038/s41467-020-18158-5

Alexandersen, S., Chamings, A., & Bhatta, T. R. (2020). SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. *Nat Commun*, *11*(1), 6059. https://doi.org/10.1038/s41467-020-19883-7

Allan, C., Burel, J. M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., Macdonald, D., Moore, W. J., Neves, C., Patterson, A., Porter, M., Tarkowska, A., Loranger, B., Avondo, J., Lagerstedt, I., Lianas, L., Leo, S., Hands, K., Hay, R. T., … Swedlow, J. R. (2012). OMERO: Flexible, model-driven data management for experimental biology. *Nat Methods*, *9*(3), 245–253. https://doi.org/10.1038/nmeth.1896

Angelini, M. M., Akhlaghpour, M., Neuman, B. W., & Buchmeier, M. J. (2013). Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio*, *4*(4), e00524–13. https://doi.org/10.1128/mBio.00524-13

Arefin, B., Parvin, F., Bahrampour, S., Stadler, C. B., & Thor, S. (2019). Drosophila Neuroblast Selection Is Gated by Notch, Snail, SoxB, and EMT Gene Interplay. *Cell Reports*, *29*(11), 3636–3651.e3. https://doi.org/10.1016/j.celrep.2019.11.038

Aughey, G. N., Estacio Gomez, A., Thomson, J., Yin, H., & Southall, T. D. (2018). CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. *eLife*, *7*, e32341. https://doi.org/10.7554/eLife.32341

Bae, B., & Miura, P. (2020). Emerging Roles for 3′ UTRs in Neurons. *International Journal of Molecular Sciences*, *21*(10), 3413. https://doi.org/10.3390/ijms21103413

Bahar Halpern, K., & Itzkovitz, S. (2016). Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues. *Methods*, *98*, 134–142. https://doi.org/10.1016/j.ymeth.2015.11.015

Bannai, H., Fukatsu, K., Mizutani, A., Natsume, T., Iemura, S.-I., Ikegami, T., Inoue, T., &

Mikoshiba, K. (2004). An RNA-interacting protein, SYNCRIP (heterogeneous nuclear ribonuclear protein Q1/NSAP1) is a component of mRNA granule transported with inositol 1,4,5-trisphosphate receptor type 1 mRNA in neuronal dendrites. *The Journal of Biological Chemistry*, *279*(51), 53427–53434. https://doi.org/10.1074/jbc.M409732200

Bayraktar, O. A., Boone, J. Q., Drummond, M. L., & Doe, C. Q. (2010). Drosophila type II neuroblast lineages keep Prospero levels low to generate large clones that contribute to the adult brain central complex. *Neural Dev*, *5*, 26. https://doi.org/10.1186/1749-8104-5-26

Bayraktar, O. A., & Doe, C. Q. (2013). Combinatorial temporal patterning in progenitors expands neural diversity. *Nature*, *498*(7455), 449–455. https://doi.org/10.1038/nature12266

Belkowski, L. S., & Sen, G. C. (1987). Inhibition of vesicular stomatitis viral mRNA synthesis by interferons. *Journal of Virology*. https://doi.org/10.1128/jvi.61.3.653-660.1987

Bello, B., Reichert, H., & Hirth, F. (2006). The brain tumor gene negatively regulates neural progenitor cell proliferation in the larval central brain of Drosophila. *Development*, *133*(14), 2639–2648. https://doi.org/10.1242/dev.02429

Berdnik, D., Török, T., González-Gaitán, M., & Knoblich, J. A. (2002). The endocytic protein alpha-Adaptin is required for numb-mediated asymmetric cell division in Drosophila. *Developmental Cell*, *3*(2), 221–231. https://doi.org/10.1016/s1534-5807(02)00215-0

Berger, C., Harzer, H., Burkard, T. R., Steinmann, J., van der Horst, S., Laurenson, A. S., Novatchkova, M., Reichert, H., & Knoblich, J. A. (2012). FACS purification and transcriptome analysis of drosophila neural stem cells reveals a role for Klumpfuss in self-renewal. *Cell Rep*, *2*(2), 407–418. https://doi.org/10.1016/j.celrep.2012.07.008

Best Rocha, A., Stroberg, E., Barton, L. M., Duval, E. J., Mukhopadhyay, S., Yarid, N., Caza, T., Wilson, J. D., Kenan, D. J., Kuperman, M., Sharma, S. G., & Larsen, C. P. (2020). Detection of SARS-CoV-2 in formalin-fixed paraffin-embedded tissue sections using commercially available reagents. *Lab Invest*, *100*(11), 1485–1489. https://doi.org/10.1038/s41374-020-0464-x

Betschinger, J., Mechtler, K., & Knoblich, J. A. (2006). Asymmetric segregation of the tumor suppressor brat regulates self-renewal in Drosophila neural stem cells. *Cell*, *124*(6), 1241–1253. https://doi.org/10.1016/j.cell.2006.01.038

Billman, M. R., Rueda, D., & Bangham, C. R. M. (2017). *Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1* (No. 2:87). Wellcome Open Research. https://doi.org/10.12688/wellcomeopenres.12469.2

Boersma, S., Rabouw, H. H., Bruurs, L. J. M., Pavlovic, T., van Vliet, A. L. W., Beumer, J., Clevers, H., van Kuppeveld, F. J. M., & Tanenbaum, M. E. (2020). Translation and Replication Dynamics of Single RNA Viruses. *Cell*, *183*(7), 1930–1945 e23. https://doi.org/10.1016/j.cell.2020.10.019

Boone, J. Q., & Doe, C. Q. (2008). Identification of Drosophila type II neuroblast lineages

containing transit amplifying ganglion mother cells. *Developmental Neurobiology*, *68*(9), 1185–1195. https://doi.org/10.1002/dneu.20648

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Briese, M., Saal-Bauernschubert, L., Ji, C., Moradi, M., Ghanawi, H., Uhl, M., Appenzeller, S., Backofen, R., & Sendtner, M. (2018). hnRNP R and its main interactor, the noncoding RNA 7SK, coregulate the axonal transcriptome of motoneurons. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(12), E2859–E2868. https://doi.org/10.1073/pnas.1721670115

Broadus, J., Fuerstenberg, S., & Doe, C. Q. (1998). Staufen-dependent localization of prospero mRNA contributes to neuroblast daughter-cell fate. *Nature*, *391*(6669), 792–795. https://doi.org/10.1038/35861

Brown, J. M., De Ornellas, S., Parisi, E., Schermelleh, L., & Buckle, V. J. (2022). RASER-FISH: Non-denaturing fluorescence in situ hybridization for preservation of three-dimensional interphase chromatin structure. *Nature Protocols*, 1–26. https://doi.org/10.1038/s41596-022-00685-8

Buccitelli, C., & Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics*, *21*(10), 630–644. https://doi.org/10.1038/s41576-020-0258-4

Buchbender, A., Mutter, H., Sutandy, F. X. R., Kortel, N., Hanel, H., Busch, A., Ebersberger, S., & Konig, J. (2019). Improved library preparation with the new iCLIP2 protocol. *Methods*. https://doi.org/10.1016/j.ymeth.2019.10.003

Burgess, H. M., & Mohr, I. (2015). Cellular 5'-3' mRNA exonuclease Xrn1 controls double-stranded RNA accumulation and anti-viral responses. *Cell Host Microbe*, *17*(3), 332–344. https://doi.org/10.1016/j.chom.2015.02.003

Burke, J. M., St Clair, L. A., Perera, R., & Parker, R. (2021). Rapid decay of host basal mRNAs during SARS-CoV-2 infection perturbs host antiviral mRNA biogenesis and export. *bioRxiv*, 2021.04.19.440452. https://doi.org/10.1101/2021.04.19.440452

Busch, A., Bruggemann, M., Ebersberger, S., & Zarnack, K. (2019). iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. *Methods*. https://doi.org/10.1016/j.ymeth.2019.11.008

Calviello, L., & Ohler, U. (2017). Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome. *Trends in Genetics*, *33*(10), 728–744. https://doi.org/10.1016/j.tig.2017.08.003

Caly, L., Druce, J., Roberts, J., Bond, K., Tran, T., Kostecki, R., Yoga, Y., Naughton, W., Taiaroa, G., Seemann, T., Schultz, M. B., Howden, B. P., Korman, T. M., Lewin, S. R., Williamson,

D. A., & Catton, M. G. (2020). Isolation and rapid sharing of the 2019 novel coronavirus (SARS-CoV-2) from the first patient diagnosed with COVID-19 in Australia. *Med J Aust*, *212*(10), 459–462. https://doi.org/10.5694/mja2.50569

Cao, Y., Xu, X., Kitanovski, S., Song, L., Wang, J., Hao, P., & Hoffmann, D. (2021). Comprehensive Comparison of RNA-Seq Data of SARS-CoV-2, SARS-CoV and MERS-CoV Infections: Alternative Entry Routes and Innate Immune Responses. *Frontiers in Immunology*, *12*.

Carlson, C. R., Asfaha, J. B., Ghent, C. M., Howard, C. J., Hartooni, N., Safari, M., Frankel, A. D., & Morgan, D. O. (2020). Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Mol Cell*, *80*(6), 1092–1103 e4. https://doi.org/10.1016/j.molcel.2020.11.025

Carossino, M., Ip, H. S., Richt, J. A., Shultz, K., Harper, K., Loynachan, A. T., Del Piero, F., & Balasuriya, U. B. R. (2020). Detection of SARS-CoV-2 by RNAscope((R)) in situ hybridization and immunohistochemistry techniques. *Arch Virol*, *165*(10), 2373–2377. https://doi.org/10.1007/s00705-020-04737-w

Carrasco, J., Rauer, M., Hummel, B., Grzejda, D., Alfonso-Gonzalez, C., Lee, Y., Wang, Q., Puchalska, M., Mittler, G., & Hilgers, V. (2020). ELAV and FNE Determine Neuronal Transcript Signatures through EXon-Activated Rescue. *Mol Cell*, *80*(1), 156–163 e6. https://doi.org/10.1016/j.molcel.2020.09.011

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., & Hentze, M. W. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, *149*(6), 1393–1406. https://doi.org/10.1016/j.cell.2012.04.031

Caviness, V. S., Nowakowski, R. S., & Bhide, P. G. (2009). Neocortical neurogenesis: Morphogenetic gradients and beyond. *Trends in Neurosciences*, *32*(8), 443–450. https://doi.org/10.1016/j.tins.2009.05.003

Ceron, J., Tejedor, F. J., & Moya, F. (2006). A primary cell culture of Drosophila postembryonic larval neuroblasts to study cell cycle and asymmetric division. *European Journal of Cell Biology*, *85*(6), 567–575. https://doi.org/10.1016/j.ejcb.2006.02.006

Chell, J. M., & Brand, A. H. (2010). Nutrition-responsive glia control exit of neural stem cells from quiescence. *Cell*, *143*(7), 1161–1173. https://doi.org/10.1016/j.cell.2010.12.007

Chia, W., Somers, W. G., & Wang, H. (2008). Drosophila neuroblast asymmetric divisions: Cell cycle regulators, asymmetric protein localization, and tumorigenesis. *The Journal of Cell Biology*, *180*(2), 267–272. https://doi.org/10.1083/jcb.200708159

Choksi, S. P., Southall, T. D., Bossing, T., Edoff, K., de Wit, E., Fischer, B. E., van Steensel, B., Micklem, G., & Brand, A. H. (2006). Prospero acts as a binary switch between self-renewal and differentiation in Drosophila neural stem cells. *Developmental Cell*, *11*(6), 775–789.

https://doi.org/10.1016/j.devcel.2006.09.015

Chou, Y., Heaton, N. S., Gao, Q., Palese, P., Singer, R. H., Singer, R., & Lionnet, T. (2013). Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathogens*, *9*(5), e1003358. https://doi.org/10.1371/journal.ppat.1003358

Chou, Y., & Lionnet, T. (2018). Single-Molecule Sensitivity RNA FISH Analysis of Influenza Virus Genome Trafficking. In Y. Yamauchi (Ed.), *Influenza Virus: Methods and Protocols* (pp. 195–211). Springer. https://doi.org/10.1007/978-1-4939-8678-1_10

Chu, H., Chan, J. F., Yuen, T. T., Shuai, H., Yuan, S., Wang, Y., Hu, B., Yip, C. C., Tsang, J. O., Huang, X., Chai, Y., Yang, D., Hou, Y., Chik, K. K., Zhang, X., Fung, A. Y., Tsoi, H. W., Cai, J. P., Chan, W. M., … Yuen, K. Y. (2020). Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: An observational study. *Lancet Microbe*, *1*(1), e14–e23. https://doi.org/10.1016/S2666-5247(20)30004-5

Cioni, J.-M., Koppers, M., & Holt, C. E. (2018). Molecular control of local translation in axon development and maintenance. *Current Opinion in Neurobiology*, *51*, 86–94. https://doi.org/10.1016/j.conb.2018.02.025

Collier, D. A., De Marco, A., Ferreira, I. A. T. M., Meng, B., Datir, R. P., Walls, A. C., Kemp, S. A., Bassi, J., Pinto, D., Silacci-Fregni, C., Bianchi, S., Tortorici, M. A., Bowen, J., Culap, K., Jaconi, S., Cameroni, E., Snell, G., Pizzuto, M. S., Pellanda, A. F., … Gupta, R. K. (2021). Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature*, *593*(7857), 136–141. https://doi.org/10.1038/s41586-021-03412-7

Colonques, J., Ceron, J., Reichert, H., & Tejedor, F. J. (2011). A transient expression of Prospero promotes cell cycle exit of Drosophila postembryonic neurons through the regulation of Dacapo. *PloS One*, *6*(4), e19342. https://doi.org/10.1371/journal.pone.0019342

Conesa, A., Nueda, M. J., Ferrer, A., & Talón, M. (2006). maSigPro: A method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics*, *22*(9), 1096–1102. https://doi.org/10.1093/bioinformatics/btl056

Contreras, E. G., Egger, B., Gold, K. S., & Brand, A. H. (2018). Dynamic Notch signalling regulates neural stem cell state progression in the Drosophila optic lobe. *Neural Development*, *13*(1), 25. https://doi.org/10.1186/s13064-018-0123-8

Corman, V. M., Muth, D., Niemeyer, D., & Drosten, C. (2018). Hosts and Sources of Endemic Human Coronaviruses. *Advances in Virus Research*, *100*, 163–188. https://doi.org/10.1016/bs.aivir.2018.01.001

Cortese, M., Lee, J. Y., Cerikan, B., Neufeldt, C. J., Oorschot, V. M. J., Kohrer, S., Hennies, J., Schieber, N. L., Ronchi, P., Mizzon, G., Romero-Brey, I., Santarella-Mellwig, R., Schorb, M., Boermel, M., Mocaer, K., Beckwith, M. S., Templin, R. M., Gross, V., Pape, C., …

Bartenschlager, R. (2020). Integrative Imaging Reveals SARS-CoV-2-Induced Reshaping of Subcellular Morphologies. *Cell Host Microbe*, *28*(6), 853–866 e5. https://doi.org/10.1016/j.chom.2020.11.003

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, *12*, 138–163.

Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, L., Aibar, S., Makhzami, S., Christiaens, V., Bravo Gonzalez-Blas, C., Poovathingal, S., Hulselmans, G., Spanier, K. I., Moerman, T., Vanspauwen, B., Geurs, S., Voet, T., Lammertyn, J., Thienpont, B., … Aerts, S. (2018). A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell*, *174*(4), 982–998 e20. https://doi.org/10.1016/j.cell.2018.05.057

Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., van Zandvoort, K., Silverman, J. D., CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz, K., … Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, *372*(6538), eabg3055. https://doi.org/10.1126/science.abg3055

de Pretis, S., Kress, T., Morelli, M. J., Melloni, G. E. M., Riva, L., Amati, B., & Pelizzola, M. (2015). INSPEcT: A computational tool to infer mRNA synthesis, processing and degradation dynamics from RNA- and 4sU-seq time course experiments. *Bioinformatics*, *31*(17), 2829–2835. https://doi.org/10.1093/bioinformatics/btv288

Desmyter, J., Melnick, J. L., & Rawls, W. E. (1968). Defectiveness of Interferon Production and of Rubella Virus Interference in a Line of African Green Monkey Kidney Cells (Vero). *Journal of Virology*, *2*(10), 955–961. https://doi.org/10.1128/jvi.2.10.955-961.1968

Dhir, A., Dhir, S., Borowski, L. S., Jimenez, L., Teitell, M., Rotig, A., Crow, Y. J., Rice, G. I., Duffy, D., Tamby, C., Nojima, T., Munnich, A., Schiff, M., de Almeida, C. R., Rehwinkel, J., Dziembowski, A., Szczesny, R. J., & Proudfoot, N. J. (2018). Mitochondrial double-stranded RNA triggers antiviral signalling in humans. *Nature*, *560*(7717), 238–242. https://doi.org/10.1038/s41586-018-0363-0

Dicken, S. J., Murray, M. J., Thorne, L. G., Reuschl, A.-K., Forrest, C., Ganeshalingham, M., Muir, L., Kalemera, M. D., Palor, M., McCoy, L. E., Jolly, C., Towers, G. J., Reeves, M. B., & Grove, J. (2021). *Characterisation of B.1.1.7 and Pangolin coronavirus spike provides insights on the evolutionary trajectory of SARS-CoV-2* (p. 2021.03.22.436468). bioRxiv. https://doi.org/10.1101/2021.03.22.436468

Dieck, S. tom, Kochen, L., Hanus, C., Bartnik, I., Nassim-Assir, B., Merk, K., Mosler, T., Garg, S., Bunse, S., Tirrell, D. A., & Schuman, E. M. (2015). Direct visualization of

identified and newly synthesized proteins in situ. *Nature Methods*, *12*(5), 411–414. https://doi.org/10.1038/nmeth.3319

Dillard, C., Narbonne-Reveau, K., Foppolo, S., Lanet, E., & Maurange, C. (2018). Two distinct mechanisms silence chinmo in Drosophila neuroblasts and neuroepithelial cells to limit their self-renewal. *Development*, *145*(2). https://doi.org/10.1242/dev.154534

Dinesh, D. C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., & Boura, E. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog*, *16*(12), e1009100. https://doi.org/10.1371/journal.ppat.1009100

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Doe, C. Q. (2017a). Temporal Patterning in the Drosophila CNS. *Annu Rev Cell Dev Biol*, *33*, 219–240. https://doi.org/10.1146/annurev-cellbio-111315-125210

Doe, C. Q. (2017b). Temporal Patterning in the Drosophila CNS. *Annual Review of Cell and Developmental Biology*, *33*(1), 219–240. https://doi.org/10.1146/annurev-cellbio-111315-125210

Doe, C. Q., & Goodman, C. S. (1985). Early events in insect neurogenesis. II. The role of cell interactions and cell lineage in the determination of neuronal precursor cells. *Developmental Biology*, *111*(1), 206–219. https://doi.org/10.1016/0012-1606(85)90446-4

Doganay, S., Lee, M. Y., Baum, A., Peh, J., Hwang, S. Y., Yoo, J. Y., Hergenrother, P. J., Garcia-Sastre, A., Myong, S., & Ha, T. (2017). Single-cell analysis of early antiviral gene expression reveals a determinant of stochastic IFNB1 expression. *Integr Biol (Camb)*, *9*(11), 857–867. https://doi.org/10.1039/c7ib00146k

Dowall, S., Salguero, F. J., Wiblin, N., Fotheringham, S., Hatch, G., Parks, S., Gowan, K., Harris, D., Carnell, O., Fell, R., Watson, R., Graham, V., Gooch, K., Hall, Y., Mizen, S., & Hewson, R. (2021). Development of a Hamster Natural Transmission Model of SARS-CoV-2 Infection. *Viruses*, *13*(11). https://doi.org/10.3390/v13112251

Duning, K., Buck, F., Barnekow, A., & Kremerskothen, J. (2008). SYNCRIP, a component of dendritically localized mRNPs, binds to the translation regulator BC200 RNA. *Journal of Neurochemistry*, *105*(2), 351–359. https://doi.org/10.1111/j.1471-4159.2007.05138.x

Enquist, L. W. (2009). Virology in the 21st Century. *Journal of Virology*, *83*(11), 5296–5308. https://doi.org/10.1128/JVI.00151-09

Eymieux, S., Rouille, Y., Terrier, O., Seron, K., Blanchard, E., Rosa-Calatrava, M., Dubuisson, J., Belouzard, S., & Roingeard, P. (2021). Ultrastructural modifications induced by SARS-CoV-2 in Vero cells: A kinetic analysis of viral factory formation, viral particle morphogenesis and virion release. *Cell Mol Life Sci*. https://doi.org/10.1007/s00018-020-03745-y

Femino, A. M., Fay, F. S., Fogarty, K., & Singer, R. H. (1998). Visualization of single RNA transcripts in situ. *Science*, *280*(5363), 585–590. https://doi.org/10.1126/

science.280.5363.585

Fiege, J. K., Thiede, J. M., Nanda, H. A., Matchett, W. E., Moore, P. J., Montanari, N. R., Thielen, B. K., Daniel, J., Stanley, E., Hunter, R. C., Menachery, V. D., Shen, S. S., Bold, T. D., & Langlois, R. A. (2021). Single cell resolution of SARS-CoV-2 tropism, antiviral responses, and susceptibility to therapies in primary human airway epithelium. *PLOS Pathogens*, *17*(1), e1009292. https://doi.org/10.1371/journal.ppat.1009292

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., & Stern-Ginossar, N. (2021). The coding capacity of SARS-CoV-2. *Nature*, *589*(7840), 125–130. https://doi.org/10.1038/s41586-020-2739-1

Fitzgerald, M., & Shenk, T. (1981). The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell*, *24*(1), 251–260. https://doi.org/10.1016/0092-8674(81)90521-3

Franks, A., Airoldi, E., & Slavov, N. (2017). Post-transcriptional regulation across human tissues. *PLoS Comput Biol*, *13*(5), e1005535. https://doi.org/10.1371/journal.pcbi.1005535

Fraser, D., & Kærn, M. (2009). A chance at survival: Gene expression noise and phenotypic diversification strategies. *Molecular Microbiology*, *71*(6), 1333–1340. https://doi.org/10.1111/j.1365-2958.2009.06605.x

Frediansyah, A., Tiwari, R., Sharun, K., Dhama, K., & Harapan, H. (2021). Antivirals for COVID-19: A critical review. *Clinical Epidemiology and Global Health*, *9*, 90–98. https://doi.org/10.1016/j.cegh.2020.07.006

Fung, M., Otani, I., Pham, M., & Babik, J. (2021). Zoonotic coronavirus epidemics: Severe acute respiratory syndrome, Middle East respiratory syndrome, and coronavirus disease 2019. *Annals of Allergy, Asthma & Immunology*, *126*(4), 321–337. https://doi.org/10.1016/j.anai.2020.11.021

Galloway, S. E., Paul, P., MacCannell, D. R., Johansson, M. A., Brooks, J. T., MacNeil, A., Slayton, R. B., Tong, S., Silk, B. J., Armstrong, G. L., Biggerstaff, M., & Dugan, V. G. (2021). Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. *Morbidity and Mortality Weekly Report*, *70*(3), 95–99. https://doi.org/10.15585/mmwr.mm7003e2

Garcia-Moreno, M., Noerenberg, M., Ni, S., Jarvelin, A. I., Gonzalez-Almela, E., Lenz, C. E., Bach-Pages, M., Cox, V., Avolio, R., Davis, T., Hester, S., Sohier, T. J. M., Li, B., Heikel, G., Michlewski, G., Sanz, M. A., Carrasco, L., Ricci, E. P., Pelechano, V., … Castello, A. (2019). System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Mol Cell*, *74*(1), 196–211 e11. https://doi.org/10.1016/j.molcel.2019.01.017

Gaspar, I., Wippich, F., & Ephrussi, A. (2017). Enzymatic production of single-molecule FISH and

RNA capture probes. *RNA*, *23*(10), 1582–1591. https://doi.org/10.1261/rna.061184.117

Geuens, T., Bouhy, D., & Timmerman, V. (2016). The hnRNP family: Insights into their role in health and disease. *Hum Genet*, *135*(8), 851–867. https://doi.org/10.1007/s00439-016-1683-5

Glock, C., Biever, A., Tushev, G., Nassim-Assir, B., Kao, A., Bartnik, I., Tom Dieck, S., & Schuman, E. M. (2021). The translatome of neuronal cell bodies, dendrites, and axons. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(43), e2113929118. https://doi.org/10.1073/pnas.2113929118

Goldberg, I. G., Allan, C., Burel, J.-M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P. K., & Swedlow, J. R. (2005). The Open Microscopy Environment (OME) Data Model and XML file: Open tools for informatics and quantitative analysis in biological imaging. *Genome Biology*, *6*(5), R47. https://doi.org/10.1186/gb-2005-6-5-r47

Grosskortenhaus, R., Robinson, K. J., & Doe, C. Q. (2006). Pdm and Castor specify late-born motor neuron identity in the NB7-1 lineage. *Genes Dev*, *20*(18), 2618–2627. https://doi.org/10.1101/gad.1445306

Guan, W., Bellemin, S., Bouchet, M., Venkatasubramanian, L., Guillermin, C., Laurençon, A., Chérif, K., Darmas, A., Godin, C., Urdy, S., Mann, R. S., & Enriquez, J. (2021). *The post-transcriptional regulation of TFs in immature motoneurons shapes the axon-muscle connectome* (p. 2021.10.22.465474). bioRxiv. https://doi.org/10.1101/2021.10.22.465474

Guan, W., Bouchet, M., Darmas, A., & Enriquez, J. (2021). *Shaping the size of a neuronal lineage: The role of Imp and Syp RBPs in the precise elimination of neurons by apoptosis* (p. 2021.11.06.467542). bioRxiv. https://doi.org/10.1101/2021.11.06.467542

Guerini-Rocco, E., Taormina, S. V., Vacirca, D., Ranghiero, A., Rappa, A., Fumagalli, C., Maffini, F., Rampinelli, C., Galetta, D., Tagliabue, M., Ansarin, M., & Barberis, M. (2020). SARS-CoV-2 detection in formalin-fixed paraffin-embedded tissue specimens from surgical resection of tongue squamous cell carcinoma. *J Clin Pathol*, *73*(11), 754–757. https://doi.org/10.1136/jclinpath-2020-206635

Guo, K., Wustoni, S., Koklu, A., Díaz-Galicia, E., Moser, M., Hama, A., Alqahtani, A. A., Ahmad, A. N., Alhamlan, F. S., Shuaib, M., Pain, A., McCulloch, I., Arold, S. T., Grünberg, R., & Inal, S. (2021). Rapid single-molecule detection of COVID-19 and MERS antigens via nanobody-functionalized organic electrochemical transistors. *Nature Biomedical Engineering*, *5*(7), 666–677. https://doi.org/10.1038/s41551-021-00734-9

Hackstadt, T., Chiramel, A. I., Hoyt, F. H., Williamson, B. N., Dooley, C. A., Beare, P. A., de Wit, E., Best, S. M., & Fischer, E. R. (2021). Disruption of the Golgi Apparatus and Contribution of the Endoplasmic Reticulum to the SARS-CoV-2 Replication Complex. *Viruses*, *13*(9).

https://doi.org/10.3390/v13091798

Hailstone, M., Waithe, D., Samuels, T. J., Yang, L., Costello, I., Arava, Y., Robertson, E., Parton, R. M., & Davis, I. (2020). CytoCensus, mapping cell identity and division in tissues and organs using machine learning. *eLife*, *9*, e51085. https://doi.org/10.7554/eLife.51085

Halstead, J. M., Lin, Y. Q., Durraine, L., Hamilton, R. S., Ball, G., Neely, G. G., Bellen, H. J., & Davis, I. (2014). Syncrip/hnRNP Q influences synaptic transmission and regulates BMP signaling at the Drosophila neuromuscular synapse. *Biol Open*, *3*(9), 839–849. https://doi.org/10.1242/bio.20149027

Hansen, D. V., Lui, J. H., Parker, P. R. L., & Kriegstein, A. R. (2010). Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature*, *464*(7288), 554–561. https://doi.org/10.1038/nature08845

Hansen, H. T., Rasmussen, S. H., Adolph, S. K., Plass, M., Krogh, A., Sanford, J., Nielsen, F. C., & Christiansen, J. (2015). Drosophila Imp iCLIP identifies an RNA assemblage coordinating F-actin formation. *Genome Biol*, *16*, 123. https://doi.org/10.1186/s13059-015-0687-0

Hansen, M. M. K., Desai, R. V., Simpson, M. L., & Weinberger, L. S. (2018). Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. *Cell Systems*, *7*(4), 384–397.e6. https://doi.org/10.1016/j.cels.2018.08.002

Healy, C. P., Adler, F. R., & Deans, T. L. (2020). Rolling Signal-based Ripley's K: A new algorithm to identify spatial patterns in histological specimens. *bioRxiv*, 2020.05.21.109314. https://doi.org/10.1101/2020.05.21.109314

Hepp, C., Shiaelis, N., Robb, N. C., Vaughan, A., Matthews, P. C., Stoesser, N., Crook, D., & Kapanidis, A. N. (2021). Viral detection and identification in 20 min by rapid single-particle fluorescence in-situ hybridization of viral RNA. *Scientific Reports*, *11*(1), 19579. https://doi.org/10.1038/s41598-021-98972-z

Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., von Haeseler, A., Zuber, J., & Ameres, S. L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods*, *14*(12), 1198–1204. https://doi.org/10.1038/nmeth.4435

Hilgers, V., Lemke, S. B., & Levine, M. (2012). ELAV mediates 3' UTR extension in the Drosophila nervous system. *Genes Dev*, *26*(20), 2259–2264. https://doi.org/10.1101/gad.199653.112

Hilgers, V., Perry, M. W., Hendrix, D., Stark, A., Levine, M., & Haley, B. (2011). Neural-specific elongation of 3' UTRs during Drosophila development. *Proc Natl Acad Sci U S A*, *108*(38), 15864–15869. https://doi.org/10.1073/pnas.1112672108

Hirata, J., Nakagoshi, H., Nabeshima, Y., & Matsuzaki, F. (1995). Asymmetric segregation of the homeodomain protein Prospero duringDrosophila development. *Nature*, *377*(6550),

627–630. https://doi.org/10.1038/377627a0

Hobert, O. (2010). Neurogenesis in the nematode Caenorhabditis elegans. *WormBook: The Online Review of C. Elegans Biology*, 1–24. https://doi.org/10.1895/wormbook.1.12.2

Hobor, F., Dallmann, A., Ball, N. J., Cicchini, C., Battistelli, C., Ogrodowicz, R. W., Christodoulou, E., Martin, S. R., Castello, A., Tripodi, M., Taylor, I. A., & Ramos, A. (2018). A cryptic RNA-binding domain mediates Syncrip recognition and exosomal partitioning of miRNA targets. *Nature Communications*, *9*(1), 831. https://doi.org/10.1038/s41467-018-03182-3

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., & Pohlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, *181*(2), 271–280 e8. https://doi.org/10.1016/j.cell.2020.02.052

Holguera, I., & Desplan, C. (2018). Neuronal specification in space and time. *Science (New York, N.Y.)*, *362*(6411), 176–180. https://doi.org/10.1126/science.aas9435

Holt, C. E., Martin, K. C., & Schuman, E. M. (2019). Local translation in neurons: Visualization and function. *Nature Structural & Molecular Biology*, *26*(7), 557–566. https://doi.org/10.1038/s41594-019-0263-5

Holt, C. E., & Schuman, E. M. (2013). The Central Dogma Decentralized: New Perspectives on RNA Function and Local Translation in Neurons. *Neuron*, *80*(3), 648–657. https://doi.org/10.1016/j.neuron.2013.10.036

Homem, C. C., & Knoblich, J. A. (2012). Drosophila neuroblasts: A model for stem cell biology. *Development*, *139*(23), 4297–4310. https://doi.org/10.1242/dev.080515

Hu, Y., Comjean, A., Perkins, L. A., Perrimon, N., & Mohr, S. E. (2015). GLAD: An Online Database of G ene L ist A nnotation for *DRosophila* . *Journal of Genomics*, *3*, 75–81. https://doi.org/10.7150/jgen.12863

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., Tajnik, M., Konig, J., & Ule, J. (2014). iCLIP: Protein-RNA interactions at nucleotide resolution. *Methods*, *65*(3), 274–287. https://doi.org/10.1016/j.ymeth.2013.10.011

Huston, N. C., Wan, H., Strine, M. S., de Cesaris Araujo Tavares, R., Wilen, C. B., & Pyle, A. M. (2021). Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol Cell*, *81*(3), 584–598 e5. https://doi.org/10.1016/j.molcel.2020.12.041

Ignatiadis, N., Klaus, B., Zaugg, J. B., & Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods*, *13*(7), 577–580. https://doi.org/10.1038/nmeth.3885

Ikeshima-Kataoka, H., Skeath, J. B., Nabeshima, Y., Doe, C. Q., & Matsuzaki, F. (1997). Miranda directs Prospero to a daughter cell during Drosophila asymmetric divisions. *Nature*, *390*(6660),

625–629. https://doi.org/10.1038/37641

Imbert, A., Ouyang, W., Safieddine, A., Coleno, E., Zimmer, C., Bertrand, E., Walter, T., & Mueller, F. (2021). *FISH-quant v2: A scalable and modular analysis tool for smFISH image analysis* (p. 2021.07.20.453024). bioRxiv. https://doi.org/10.1101/2021.07.20.453024

Iselin, L., Palmalux, N., Kamel, W., Simmonds, P., Mohammed, S., & Castello, A. (2022). Uncovering viral RNA-host cell interactions on a proteome-wide scale. *Trends in Biochemical Sciences*, *47*(1), 23–38. https://doi.org/10.1016/j.tibs.2021.08.002

Iserman, C., Roden, C. A., Boerneke, M. A., Sealfon, R. S. G., McLaughlin, G. A., Jungreis, I., Fritch, E. J., Hou, Y. J., Ekena, J., Weidmann, C. A., Theesfeld, C. L., Kellis, M., Troyanskaya, O. G., Baric, R. S., Sheahan, T. P., Weeks, K. M., & Gladfelter, A. S. (2020). Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol Cell*, *80*(6), 1078–1091 e6. https://doi.org/10.1016/j.molcel.2020.11.041

Iserman, C., Roden, C., Boerneke, M., Sealfon, R., McLaughlin, G., Jungreis, I., Park, C., Boppana, A., Fritch, E., Hou, Y. J., Theesfeld, C., Troyanskaya, O. G., Baric, R. S., Sheahan, T. P., Weeks, K., & Gladfelter, A. S. (2020). Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate. *bioRxiv*. https://doi.org/10.1101/2020.06.11.147199

Isshiki, T., Pearson, B., Holbrook, S., & Doe, C. Q. (2001). Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell*, *106*(4), 511–521. https://doi.org/10.1016/s0092-8674(01)00465-2

Janssens, D. H., Komori, H., Grbac, D., Chen, K., Koe, C. T., Wang, H., & Lee, C.-Y. (2014). Earmuff restricts progenitor cell potential by attenuating the competence to respond to self-renewal factors. *Development (Cambridge, England)*, *141*(5), 1036–1046. https://doi.org/10.1242/dev.106534

Jessell, T. M. (2000). Neuronal specification in the spinal cord: Inductive signals and transcriptional codes. *Nature Reviews. Genetics*, *1*(1), 20–29. https://doi.org/10.1038/35049541

Jiang, H. W., Zhang, H. N., Meng, Q. F., Xie, J., Li, Y., Chen, H., Zheng, Y. X., Wang, X. N., Qi, H., Zhang, J., Wang, P. H., Han, Z. G., & Tao, S. C. (2020). SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell Mol Immunol*, *17*(9), 998–1000. https://doi.org/10.1038/s41423-020-0514-8

Jiao, J., Duan, C., Xue, L., Liu, Y., Sun, W., & Xiang, Y. (2020). DNA nanoscaffold-based SARS-CoV-2 detection for COVID-19 diagnosis. *Biosens Bioelectron*, *167*, 112479. https://doi.org/10.1016/j.bios.2020.112479

Jürges, C., Dölken, L., & Erhard, F. (2018). Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics*, *34*(13), i218–i226. https://doi.org/10.1093/bioinformatics/bty256

Kaltschmidt, J. A., & Brand, A. H. (2002). Asymmetric cell division: Microtubule dynamics and spindle asymmetry. *Journal of Cell Science*, *115*(Pt 11), 2257–2264. https://doi.org/10.1242/jcs.115.11.2257

Kamel, W., Noerenberg, M., Cerikan, B., Chen, H., Jarvelin, A. I., Kammoun, M., Lee, J. Y., Shuai, N., Garcia-Moreno, M., Andrejeva, A., Deery, M. J., Johnson, N., Neufeldt, C. J., Cortese, M., Knight, M. L., Lilley, K. S., Martinez, J., Davis, I., Bartenschlager, R., … Castello, A. (2021). Global analysis of protein-RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection. *Mol Cell*. https://doi.org/10.1016/j.molcel.2021.05.023

Kanca, O., Bellen, H. J., & Schnorrer, F. (2017). Gene Tagging Strategies To Assess Protein Expression, Localization, and Function in Drosophila. *Genetics*, *207*(2), 389–412. https://doi.org/10.1534/genetics.117.199968

Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., Nakane, T., Zivanov, J., Neufeldt, C. J., Cerikan, B., Lu, J. M., Peukes, J., Xiong, X., Kräusslich, H.-G., Scheres, S. H. W., Bartenschlager, R., & Briggs, J. A. G. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*, *588*(7838), 498–502. https://doi.org/10.1038/s41586-020-2665-2

Kepecs, A., & Fishell, G. (2014). Interneuron cell types are fit to function. *Nature*, *505*(7483), 318–326. https://doi.org/10.1038/nature12983

Khudayberdiev, S., Soutschek, M., Ammann, I., Heinze, A., Rust, M. B., Baumeister, S., & Schratt, G. (2021). The cytoplasmic SYNCRIP mRNA interactome of mammalian neurons. *RNA Biology*, *18*(9), 1252–1264. https://doi.org/10.1080/15476286.2020.1830553

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, *181*(4), 914–921 e10. https://doi.org/10.1016/j.cell.2020.04.011

Kimura, S., Matsumiya, T., Shiba, Y., Nakanishi, M., Hayakari, R., Kawaguchi, S., Yoshida, H., & Imaizumi, T. (2018). The Essential Role of Double-Stranded RNA-Dependent Antiviral Signaling in the Degradation of Nonself Single-Stranded RNA in Nonimmune Cells. *J Immunol*, *201*(3), 1044–1052. https://doi.org/10.4049/jimmunol.1800456

Kiourlappou, M., Taylor, S., & Davis, I. (2022). *Zegami user manual for data exploration: "Systematic analysis of YFP gene traps reveals common discordance between mRNA and protein across the nervous system"*. Zenodo. https://doi.org/10.5281/zenodo.6374012

Kissler, S. M., Fauver, J. R., Mack, C., Tai, C. G., Breban, M. I., Watkins, A. E., Samant, R. M., Anderson, D. J., Metti, J., Khullar, G., Baits, R., MacKay, M., Salgado, D., Baker, T., Dudley, J. T., Mason, C. E., Ho, D. D., Grubaugh, N. D., & Grad, Y. H. (2021). Viral Dynamics of SARS-CoV-2 Variants in Vaccinated and Unvaccinated Persons. *The New England Journal of Medicine*, *385*(26), 2489–2491. https://doi.org/10.1056/NEJMc2102507

Kitamura, T., & Imafuku, M. (2015). Behavioural mimicry in flight path of Batesian intraspecific

polymorphic butterfly Papilio polytes. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1809), 20150483. https://doi.org/10.1098/rspb.2015.0483

Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., Stanifer, M. L., Boulant, S., Bartenschlager, R., & Chlanda, P. (2020). SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun*, *11*(1), 5885. https://doi.org/10.1038/s41467-020-19619-7

Knoblich, J. A. (2008). Mechanisms of asymmetric stem cell division. *Cell*, *132*(4), 583–597. https://doi.org/10.1016/j.cell.2008.02.007

Knoops, K., Kikkert, M., Worm, S. H. E. van den, Zevenhoven-Dobbe, J. C., Meer, Y. van der, Koster, A. J., Mommaas, A. M., & Snijder, E. J. (2008). SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PLOS Biology*, *6*(9), e226. https://doi.org/10.1371/journal.pbio.0060226

Konstantinides, N., Holguera, I., Rossi, A. M., Escobar, A., Dudragne, L., Chen, Y.-C., Tran, T. N., Martínez Jaimes, A. M., Özel, M. N., Simon, F., Shao, Z., Tsankova, N. M., Fullard, J. F., Walldorf, U., Roussos, P., & Desplan, C. (2022). A complete temporal transcription factor series in the fly visual system. *Nature*, *604*(7905), 316–322. https://doi.org/10.1038/s41586-022-04564-w

Kopecky-Bromberg, S. A., Martínez-Sobrido, L., Frieman, M., Baric, R. A., & Palese, P. (2007). Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *Journal of Virology*, *81*(2), 548–557. https://doi.org/10.1128/JVI.01782-06

Korenbaum, E., & Rivero, F. (2002). Calponin homology domains at a glance. *Journal of Cell Science*, *115*(18), 3543–3545. https://doi.org/10.1242/jcs.00003

Kusmartseva, I., Wu, W., Syed, F., Van Der Heide, V., Jorgensen, M., Joseph, P., Tang, X., Candelario-Jalil, E., Yang, C., Nick, H., Harbert, J. L., Posgai, A. L., Paulsen, J. D., Lloyd, R., Cechin, S., Pugliese, A., Campbell-Thompson, M., Vander Heide, R. S., Evans-Molina, C., … Atkinson, M. A. (2020). Expression of SARS-CoV-2 Entry Factors in the Pancreas of Normal Organ Donors and Individuals with COVID-19. *Cell Metab*, *32*(6), 1041–1051 e6. https://doi.org/10.1016/j.cmet.2020.11.005

Lam, L. T., Pickeral, O. K., Peng, A. C., Rosenwald, A., Hurt, E. M., Giltnane, J. M., Averett, L. M., Zhao, H., Davis, R. E., Sathyamoorthy, M., Wahl, L. M., Harris, E. D., Mikovits, J. A., Monks, A. P., Hollingshead, M. G., Sausville, E. A., & Staudt, L. M. (2001). Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biology*, *2*(10), research0041.1. https://doi.org/10.1186/gb-2001-2-10-research0041

Landskron, L., Steinmann, V., Bonnay, F., Burkard, T. R., Steinmann, J., Reichardt, I., Harzer, H., Laurenson, A. S., Reichert, H., & Knoblich, J. A. (2018). The asymmetrically segregating

lncRNA cherub is required for transforming stem cells into malignant cells. *Elife*, *7*. https://doi.org/10.7554/eLife.31347

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laue, M., Kauter, A., Hoffmann, T., Möller, L., Michel, J., & Nitsche, A. (2021). Morphometry of SARS-CoV and SARS-CoV-2 particles in ultrathin plastic sections of infected Vero cell cultures. *Scientific Reports*, *11*(1), 3515. https://doi.org/10.1038/s41598-021-82852-7

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput Biol*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Lean, F. Z. X., Lamers, M. M., Smith, S. P., Shipley, R., Schipper, D., Temperton, N., Haagmans, B. L., Banyard, A. C., Bewley, K. R., Carroll, M. W., Brookes, S. M., Brown, I., & Nunez, A. (2020). Development of immunohistochemistry and in situ hybridisation for the detection of SARS-CoV and SARS-CoV-2 in formalin-fixed paraffin-embedded specimens. *Sci Rep*, *10*(1), 21894. https://doi.org/10.1038/s41598-020-78949-0

Lee, C.-Y., Wilkinson, B. D., Siegrist, S. E., Wharton, R. P., & Doe, C. Q. (2006). Brat is a Miranda cargo protein that promotes neuronal differentiation and inhibits neuroblast self-renewal. *Developmental Cell*, *10*(4), 441–449. https://doi.org/10.1016/j.devcel.2006.01.017

Lee, F. C. Y., & Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell*, *69*(3), 354–369. https://doi.org/10.1016/j.molcel.2018.01.005

Lee, J. Y., Wing, P. A., Gala, D. S., Noerenberg, M., Järvelin, A. I., Titlow, J., Zhuang, X., Palmalux, N., Iselin, L., Thompson, M. K., Parton, R. M., Prange-Barczynska, M., Wainman, A., Salguero, F. J., Bishop, T., Agranoff, D., James, W., Castello, A., McKeating, J. A., & Davis, I. (2022). Absolute quantitation of individual SARS-CoV-2 RNA molecules provides a new paradigm for infection dynamics and variant differences. *eLife*, *11*, e74153. https://doi.org/10.7554/eLife.74153

Lee, P. T., Zirin, J., Kanca, O., Lin, W. W., Schulze, K. L., Li-Kroeger, D., Tao, R., Devereaux, C., Hu, Y., Chung, V., Fang, Y., He, Y., Pan, H., Ge, M., Zuo, Z., Housden, B. E., Mohr, S. E., Yamamoto, S., Levis, R. W., … Bellen, H. J. (2018a). A gene-specific T2A-GAL4 library for Drosophila. *Elife*, *7*. https://doi.org/10.7554/eLife.35574

Lee, P.-T., Zirin, J., Kanca, O., Lin, W.-W., Schulze, K. L., Li-Kroeger, D., Tao, R., Devereaux, C., Hu, Y., Chung, V., Fang, Y., He, Y., Pan, H., Ge, M., Zuo, Z., Housden, B. E., Mohr, S. E., Yamamoto, S., Levis, R. W., … Bellen, H. J. (2018b). A gene-specific T2A-GAL4 library for Drosophila. *eLife*, *7*, e35574. https://doi.org/10.7554/eLife.35574

Lee, S., Lee, Y.-S., Choi, Y., Son, A., Park, Y., Lee, K.-M., Kim, J., Kim, J.-S., & Kim, V. N. (2021). The SARS-CoV-2 RNA interactome. *Molecular Cell*, *81*(13), 2838–2850.e6. https://doi.org/10.1016/j.molcel.2021.04.022

Lee, S., Wei, L., Zhang, B., Goering, R., Majumdar, S., Wen, J., Taliaferro, J. M., & Lai, E. C. (2021). ELAV/Hu RNA binding proteins determine multiple programs of neural alternative splicing. *PLOS Genetics*, *17*(4), e1009439. https://doi.org/10.1371/journal.pgen.1009439

Legland, D., Arganda-Carreras, I., & Andrey, P. (2016). MorphoLibJ: Integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics*, *32*(22), 3532–3534. https://doi.org/10.1093/bioinformatics/btw413

Lelieveld, S. H., Reijnders, M. R. F., Pfundt, R., Yntema, H. G., Kamsteeg, E.-J., de Vries, P., de Vries, B. B. A., Willemsen, M. H., Kleefstra, T., Löhner, K., Vreeburg, M., Stevens, S. J. C., van der Burgt, I., Bongers, E. M. H. F., Stegmann, A. P. A., Rump, P., Rinne, T., Nelen, M. R., Veltman, J. A., … Gilissen, C. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nature Neuroscience*, *19*(9), 1194–1196. https://doi.org/10.1038/nn.4352

Li, P., Yang, X., Wasser, M., Cai, Y., & Chia, W. (1997). Inscuteable and Staufen mediate asymmetric localization and segregation of prospero RNA during Drosophila neuroblast cell divisions. *Cell*, *90*(3), 437–447. https://doi.org/10.1016/s0092-8674(00)80504-8

Li, Y., Renner, D. M., Comar, C. E., Whelan, J. N., Reyes, H. M., Cardenas-Diaz, F. L., Truitt, R., Tan, L. H., Dong, B., Alysandratos, K. D., Huang, J., Palmer, J. N., Adappa, N. D., Kohanski, M. A., Kotton, D. N., Silverman, R. H., Yang, W., Morrisey, E. E., Cohen, N. A., & Weiss, S. R. (2021). SARS-CoV-2 induces double-stranded RNA-mediated innate immune responses in respiratory epithelial-derived cells and cardiomyocytes. *Proc Natl Acad Sci U S A*, *118*(16). https://doi.org/10.1073/pnas.2022643118

Liu, J., Babka, A. M., Kearney, B. J., Radoshitzky, S. R., Kuhn, J. H., & Zeng, X. (2020). Molecular detection of SARS-CoV-2 in formalin-fixed, paraffin-embedded specimens. *JCI Insight*, *5*(12). https://doi.org/10.1172/jci.insight.139042

Liu, L.-Y., Long, X., Yang, C.-P., Miyares, R. L., Sugino, K., Singer, R. H., & Lee, T. (2020). Mamo decodes hierarchical temporal gradients into terminal neuronal fate. *eLife*, *8*, e48056. https://doi.org/10.7554/eLife.48056

Liu, S.-L., Wang, Z.-G., Xie, H.-Y., Liu, A.-A., Lamb, D. C., & Pang, D.-W. (2020). Single-Virus Tracking: From Imaging Methodologies to Virological Applications. *Chemical Reviews*, *120*(3), 1936–1979. https://doi.org/10.1021/acs.chemrev.9b00692

Liu, X., Shen, J., Xie, L., Wei, Z., Wong, C., Li, Y., Zheng, X., Li, P., & Song, Y. (2020). Mitotic Implantation of the Transcription Factor Prospero via Phase Separation Drives Terminal Neuronal Differentiation. *Developmental Cell*, *52*(3), 277–293.e8. https://doi.org/10.1016/j.devcel.2019.11.019

Liu, Y., Beyer, A., & Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, *165*(3), 535–550. https://doi.org/10.1016/j.cell.2016.

03.014

Liu, Z., Yang, C. P., Sugino, K., Fu, C. C., Liu, L. Y., Yao, X., Lee, L. P., & Lee, T. (2015). Opposing intrinsic temporal gradients guide neural stem cell production of varied neuronal fates. *Science*, *350*(6258), 317–320. https://doi.org/10.1126/science.aad1886

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lowe, N., Rees, J. S., Roote, J., Ryder, E., Armean, I. M., Johnson, G., Drummond, E., Spriggs, H., Drummond, J., Magbanua, J. P., Naylor, H., Sanson, B., Bastock, R., Huelsmann, S., Trovisco, V., Landgraf, M., Knowles-Barley, S., Armstrong, J. D., White-Cooper, H., … St Johnston, D. (2014). Analysis of the expression patterns, subcellular localisations and interaction partners of Drosophila proteins using a pigP protein trap library. *Development*, *141*(20), 3994–4005. https://doi.org/10.1242/dev.111054

Maccio, U., Zinkernagel, A. S., Schuepbach, R., Probst-Mueller, E., Frontzek, K., Brugger, S. D., Hofmaenner, D. A., Moch, H., & Varga, Z. (2022). Long-Term Persisting SARS-CoV-2 RNA and Pathological Findings: Lessons Learnt From a Series of 35 COVID-19 Autopsies. *Frontiers in Medicine*, *9*.

Malka, Y., Steiman-Shimony, A., Rosenthal, E., Argaman, L., Cohen-Daniel, L., Arbib, E., Margalit, H., Kaplan, T., & Berger, M. (2017). Post-transcriptional 3 -UTR cleavage of mRNA transcripts generates thousands of stable uncapped autonomous RNA fragments. *Nat Commun*, *8*(1), 2029. https://doi.org/10.1038/s41467-017-02099-7

Maneewongvatana, S., & Mount, D. (1999). Analysis of approximate nearest neighbor searching with clustered point sets. *Data Structures, Near Neighbor Searches, and Methodology*.

Marshall, O. J., & Brand, A. H. (2017). Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. *Nature Communications*, *8*(1), 2271. https://doi.org/10.1038/s41467-017-02385-4

Marx, V. (2021). Method of the Year: Spatially resolved transcriptomics. *Nature Methods*, *18*(1), 9–14. https://doi.org/10.1038/s41592-020-01033-y

Maurange, C., Cheng, L., & Gould, A. P. (2008). Temporal transcription factors and their targets schedule the end of neural proliferation in Drosophila. *Cell*, *133*(5), 891–902. https://doi.org/10.1016/j.cell.2008.03.034

Mayr, C. (2016). Evolution and Biological Roles of Alternative 3′UTRs. *Trends in Cell Biology*, *26*(3), 227–237. https://doi.org/10.1016/j.tcb.2015.10.012

Mayr, C. (2017). Regulation by 3'-Untranslated Regions. *Annu Rev Genet*, *51*, 171–194. https://doi.org/10.1146/annurev-genet-120116-024704

McDermott, S. M., Meignin, C., Rappsilber, J., & Davis, I. (2012). Drosophila Syncrip binds the gurken mRNA localisation signal and regulates localised transcripts during axis specification.

Biol Open, *1*(5), 488–497. https://doi.org/10.1242/bio.2012885

McDermott, S. M., Yang, L., Halstead, J. M., Hamilton, R. S., Meignin, C., & Davis, I. (2014). Drosophila Syncrip modulates the expression of mRNAs encoding key synaptic proteins required for morphology at the neuromuscular junction. *RNA*, *20*(10), 1593–1606. https://doi.org/10.1261/rna.045849.114

McLean, D. J., & Skowron Volponi, M. A. (2018). Trajr: An R package for characterisation of animal trajectories. *Ethology*, *124*(6), 440–448. https://doi.org/10.1111/eth.12739

Mehandru, S., & Merad, M. (2022). Pathological sequelae of long-haul COVID. *Nature Immunology*, *23*(2), 194–202. https://doi.org/10.1038/s41590-021-01104-y

Mendonca, L., Howe, A., Gilchrist, J. B., Sheng, Y., Sun, D., Knight, M. L., Zanetti-Domingues, L. C., Bateman, B., Krebs, A. S., Chen, L., Radecke, J., Li, V. D., Ni, T., Kounatidis, I., Koronfel, M. A., Szynkiewicz, M., Harkiolaki, M., Martin-Fernandez, M. L., James, W., & Zhang, P. (2021). Correlative multi-scale cryo-imaging unveils SARS-CoV-2 assembly and egress. *Nat Commun*, *12*(1), 4629. https://doi.org/10.1038/s41467-021-24887-y

Meyer, K., Köster, T., Nolte, C., Weinholdt, C., Lewinski, M., Grosse, I., & Staiger, D. (2017). Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7. *Genome Biology*, *18*, 204. https://doi.org/10.1186/s13059-017-1332-x

Mili, S., Moissoglu, K., & Macara, I. G. (2008). Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature*, *453*(7191), 115–119. https://doi.org/10.1038/nature06888

Miller, S., & Krijnse-Locker, J. (2008). Modification of intracellular membrane structures for virus replication. *Nature Reviews. Microbiology*, *6*(5), 363–374. https://doi.org/10.1038/nrmicro1890

Mizutani, A., Fukuda, M., Ibata, K., Shiraishi, Y., & Mikoshiba, K. (2000). SYNCRIP, a Cytoplasmic Counterpart of Heterogeneous Nuclear Ribonucleoprotein R, Interacts with Ubiquitous Synaptotagmin Isoforms *. *Journal of Biological Chemistry*, *275*(13), 9823–9831. https://doi.org/10.1074/jbc.275.13.9823

Morin, X., Daneman, R., Zavortink, M., & Chia, W. (2001). A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. *Proc Natl Acad Sci U S A*, *98*(26), 15050–15055. https://doi.org/10.1073/pnas.261408198

Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., & Zimmer, C. (2013). FISH-quant: Automatic counting of transcripts in 3D FISH images. *Nat Methods*, *10*(4), 277–278. https://doi.org/10.1038/nmeth.2406

Nadadhur, A. G., Leferink, P. S., Holmes, D., Hinz, L., Cornelissen-Steijger, P., Gasparotto, L., & Heine, V. M. (2018). Patterning factors during neural progenitor induction determine regional identity and differentiation potential in vitro. *Stem Cell Research*, *32*, 25–34. https:

//doi.org/10.1016/j.scr.2018.08.017

Nagarkar-Jaiswal, S., Lee, P. T., Campbell, M. E., Chen, K., Anguiano-Zarate, S., Gutierrez, M. C., Busby, T., Lin, W. W., He, Y., Schulze, K. L., Booth, B. W., Evans-Holm, M., Venken, K. J., Levis, R. W., Spradling, A. C., Hoskins, R. A., & Bellen, H. J. (2015). A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. *Elife*, *4*. https://doi.org/10.7554/eLife.05338

Nassal, M. (2015). HBV cccDNA: Viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut*, *64*(12), 1972–1984. https://doi.org/10.1136/gutjnl-2015-309809

Neumuller, R. A., Richter, C., Fischer, A., Novatchkova, M., Neumuller, K. G., & Knoblich, J. A. (2011). Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. *Cell Stem Cell*, *8*(5), 580–593. https://doi.org/10.1016/j.stem.2011.02.022

Nguyen, D. T. T., Lu, Y., Chu, K. L., Yang, X., Park, S.-M., Choo, Z.-N., Chin, C. R., Prieto, C., Schurer, A., Barin, E., Savino, A. M., Gourkanti, S., Patel, P., Vu, L. P., Leslie, C. S., & Kharas, M. G. (2020). HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. *Nature Communications*, *11*(1), 2026. https://doi.org/10.1038/s41467-020-15814-8

Nystrom, S. L., Niederhuber, M. J., & McKay, D. J. (2020). Expression of E93 provides an instructive cue to control dynamic enhancer activity and chromatin accessibility during development. *Development (Cambridge, England)*, *147*(6), dev181909. https://doi.org/10.1242/dev.181909

Padrón, A., Iwasaki, S., & Ingolia, N. T. (2019). Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules. *Molecular Cell*, *75*(4), 875–887.e5. https://doi.org/10.1016/j.molcel.2019.07.030

Pahl, M. C., Doyle, S. E., & Siegrist, S. E. (2019). E93 Integrates Neuroblast Intrinsic State with Developmental Time to Terminate MB Neurogenesis via Autophagy. *Curr Biol*, *29*(5), 750–762 e3. https://doi.org/10.1016/j.cub.2019.01.039

Parker, M. D., Lindsey, B. B., Shah, D. R., Hsu, S., Keeley, A. J., Partridge, D. G., Leary, S., Cope, A., State, A., Johnson, K., Ali, N., Raghei, R., Heffer, J., Smith, N., Zhang, P., Gallis, M., Louka, S. F., Whiteley, M., Foulkes, B. H., … de Silva, T. I. (2021). Altered Subgenomic RNA Expression in SARS-CoV-2 B.1.1.7 Infections. *bioRxiv*, 2021.03.02.433156. https://doi.org/10.1101/2021.03.02.433156

Pascale, A., Amadio, M., & Quattrone, A. (2008). Defining a neuron: Neuronal ELAV proteins. *Cellular and Molecular Life Sciences: CMLS*, *65*(1), 128–140. https://doi.org/10.1007/s00018-007-7017-y

Patil, S., Fribourg, M., Ge, Y., Batish, M., Tyagi, S., Hayot, F., & Sealfon, S. C. (2015). Single-cell analysis shows that paracrine signaling by first responder cells shapes the interferon-beta response to viral infection. *Sci Signal*, *8*(363), ra16. https://doi.org/10.1126/

scisignal.2005728

Perlman, S., & Netland, J. (2009). Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews Microbiology*, *7*(6), 439–450. https://doi.org/10.1038/nrmicro2147

Pilaz, L. J., & Silver, D. L. (2015). Post-transcriptional regulation in corticogenesis: How RNA-binding proteins help build the brain. *Wiley Interdiscip Rev RNA*, *6*(5), 501–515. https://doi.org/10.1002/wrna.1289

Pinder, D. A., & Witherick, M. E. (1972). The Principles, Practice and Pitfalls of Nearest-neighbour Analysis. *Geography*, *57*(4), 277–288.

Planas, D., Bruel, T., Grzelak, L., Guivel-Benhassine, F., Staropoli, I., Porrot, F., Planchais, C., Buchrieser, J., Rajah, M. M., Bishop, E., Albert, M., Donati, F., Prot, M., Behillil, S., Enouf, V., Maquart, M., Smati-Lafarge, M., Varon, E., Schortgen, F., … Schwartz, O. (2021). Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nature Medicine*, *27*(5), 917–924. https://doi.org/10.1038/s41591-021-01318-5

Pollen, A. A., Nowakowski, T. J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C. R., Shuga, J., Liu, S. J., Oldham, M. C., Diaz, A., Lim, D. A., Leyrat, A. A., West, J. A., & Kriegstein, A. R. (2015). Molecular Identity of Human Outer Radial Glia During Cortical Development. *Cell*, *163*(1), 55–67. https://doi.org/10.1016/j.cell.2015.09.004

Poon, C. L., Mitchell, K. A., Kondo, S., Cheng, L. Y., & Harvey, K. F. (2016). The Hippo Pathway Regulates Neuroblasts and Brain Size in Drosophila melanogaster. *Curr Biol*, *26*(8), 1034–1042. https://doi.org/10.1016/j.cub.2016.02.009

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., & Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods*, *5*(10), 877–879. https://doi.org/10.1038/nmeth.1253

Raj, A., & van Oudenaarden, A. (2009). Single-Molecule Approaches to Stochastic Gene Expression. *Annual Review of Biophysics*, *38*(1), 255–270. https://doi.org/10.1146/annurev.biophys.37.032807.125928

Ramanan, V., Trehan, K., Ong, M. L., Luna, J. M., Hoffmann, H. H., Espiritu, C., Sheahan, T. P., Chandrasekar, H., Schwartz, R. E., Christine, K. S., Rice, C. M., van Oudenaarden, A., & Bhatia, S. N. (2016). Viral genome imaging of hepatitis C virus to probe heterogeneous viral infection and responses to antiviral therapies. *Virology*, *494*, 236–247. https://doi.org/10.1016/j.virol.2016.04.020

Ravindra, N. G., Alfajaro, M. M., Gasque, V., Huston, N. C., Wan, H., Szigeti-Buck, K., Yasumoto, Y., Greaney, A. M., Habet, V., Chow, R. D., Chen, J. S., Wei, J., Filler, R. B., Wang, B., Wang, G., Niklason, L. E., Montgomery, R. R., Eisenbarth, S. C., Chen, S., … Wilen, C. B. (2021). Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes. *PLOS Biology*,

19(3), e3001143. https://doi.org/10.1371/journal.pbio.3001143

Rebollo, E., Roldán, M., & Gonzalez, C. (2009). Spindle alignment is achieved without rotation after the first cell cycle in Drosophila embryonic neuroblasts. *Development (Cambridge, England)*, *136*(20), 3393–3397. https://doi.org/10.1242/dev.041822

Ren, Q., Yang, C. P., Liu, Z., Sugino, K., Mok, K., He, Y., Ito, M., Nern, A., Otsuna, H., & Lee, T. (2017). Stem Cell-Intrinsic, Seven-up-Triggered Temporal Factor Gradients Diversify Intermediate Neural Progenitors. *Curr Biol*, *27*(9), 1303–1313. https://doi.org/10.1016/j.cub.2017.03.047

Rensen, E., Pietropaoli, S., Mueller, F., Weber, C., Souquere, S., Sommer, S., Isnard, P., Rabant, M., Gibier, J.-B., Terzi, F., Simon-Loriere, E., Rameix-Welti, M.-A., Pierron, G., Barba-Spaeth, G., & Zimmer, C. (2022). Sensitive visualization of SARS-CoV-2 RNA with CoronaFISH. *Life Science Alliance*, *5*(4). https://doi.org/10.26508/lsa.202101124

Reynoso, M. A., Juntawong, P., Lancia, M., Blanco, F. A., Bailey-Serres, J., & Zanetti, M. E. (2015). Translating Ribosome Affinity Purification (TRAP) followed by RNA sequencing technology (TRAP-SEQ) for quantitative assessment of plant translatomes. *Methods in Molecular Biology (Clifton, N.J.)*, *1284*, 185–207. https://doi.org/10.1007/978-1-4939-2444-8_9

Ripley, B. D. (1979). Tests of "Randomness" for Spatial Point Patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*(3), 368–374.

Rossi, A. M., & Desplan, C. (2020). Extrinsic activin signaling cooperates with an intrinsic temporal program to increase mushroom body neuronal diversity. *eLife*, *9*, e58880. https://doi.org/10.7554/eLife.58880

Rouhanifard, S. H., Mellis, I. A., Dunagin, M., Bayatpour, S., Jiang, C. L., Dardani, I., Symmons, O., Emert, B., Torre, E., Cote, A., Sullivan, A., Stamatoyannopoulos, J. A., & Raj, A. (2018). ClampFISH detects individual nucleic acid molecules using click chemistry-based amplification. *Nature Biotechnology*. https://doi.org/10.1038/nbt.4286

Russo, J., & Wilusz, J. (2017). Trick or TREAT: A Scary-Good New Approach for Single-Molecule mRNA Decay Analysis. *Molecular Cell*, *68*(3), 476–477. https://doi.org/10.1016/j.molcel.2017.10.027

Salguero, F. J., White, A. D., Slack, G. S., Fotheringham, S. A., Bewley, K. R., Gooch, K. E., Longet, S., Humphries, H. E., Watson, R. J., Hunter, L., Ryan, K. A., Hall, Y., Sibley, L., Sarfas, C., Allen, L., Aram, M., Brunt, E., Brown, P., Buttigieg, K. R., … Carroll, M. W. (2021). Comparison of rhesus and cynomolgus macaques as an infection model for COVID-19. *Nature Communications*, *12*(1), 1260. https://doi.org/10.1038/s41467-021-21389-9

Samson, J., Cronin, S., & Dean, K. (2018). BC200 (BCYRN1) – The shortest, long, non-coding RNA associated with cancer. *Non-Coding RNA Research*, *3*(3), 131–143. https://doi.org/10.1016/j.ncrna.2018.05.003

Samuels, T. J., Arava, Y., Järvelin, A. I., Robertson, F., Lee, J. Y., Yang, L., Yang, C.-P., Lee, T., Ish-Horowicz, D., & Davis, I. (2020). Neuronal upregulation of Prospero protein is driven by alternative mRNA polyadenylation and Syncrip-mediated mRNA stabilisation. *Biology Open*, *9*(5), bio049684. https://doi.org/10.1242/bio.049684

Samuels, T. J., Järvelin, A. I., Ish-Horowicz, D., & Davis, I. (2020). Imp/IGF2BP levels modulate individual neural stem cell growth and division through myc mRNA stability. *eLife*, *9*, e51529. https://doi.org/10.7554/eLife.51529

Sarov, M., Barz, C., Jambor, H., Hein, M. Y., Schmied, C., Suchold, D., Stender, B., Janosch, S., KJ, V. V., Krishnan, R., Krishnamoorthy, A., Ferreira, I. R., Ejsmont, R. K., Finkl, K., Hasse, S., Kämpfer, P., Plewka, N., Vinis, E., Schloissnig, S., … Schnorrer, F. (2016). A genome-wide resource for the analysis of protein localisation in Drosophila. *eLife*, *5*, e12068. https://doi.org/10.7554/eLife.12068

Schoggins, J. W., & Rice, C. M. (2011). Interferon-stimulated genes and their antiviral effector functions. *Current Opinion in Virology*, *1*(6), 519–525. https://doi.org/10.1016/j.coviro.2011.10.008

Schoof, E. M., Furtwängler, B., Üresin, N., Rapin, N., Savickas, S., Gentil, C., Lechman, E., Keller, U. auf dem, Dick, J. E., & Porse, B. T. (2021). Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nature Communications*, *12*(1), 3341. https://doi.org/10.1038/s41467-021-23667-y

Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.-A., Leibundgut, M., Thiel, V., Mühlemann, O., & Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nature Structural & Molecular Biology*, *27*(10), 959–966. https://doi.org/10.1038/s41594-020-0511-8

Senussi, Y., Swank, Z. N., & Walt, D. R. (2021). The optimization of an ultra-sensitive single molecule assay for the multiplexed detection of SARS-CoV-2 antigens. *American Journal of Clinical Pathology*, *156*(Supplement_1), S145. https://doi.org/10.1093/ajcp/aqab191.309

Sethna, P. B., Hofmann, M. A., & Brian, D. A. (1991). Minus-strand copies of replicating coronavirus mRNAs contain antileaders. *J Virol*, *65*(1), 320–325. https://doi.org/10.1128/JVI.65.1.320-325.1991

Shaffer, S. M., Wu, M.-T., Levesque, M. J., & Raj, A. (2013). Turbo FISH: A Method for Rapid Single Molecule RNA FISH. *PLOS ONE*, *8*(9), e75120. https://doi.org/10.1371/journal.pone.0075120

Shang, C., Liu, Z., Zhu, Y., Lu, J., Ge, C., Zhang, C., Li, N., Jin, N., Li, Y., Tian, M., & Li, X. (2022). SARS-CoV-2 Causes Mitochondrial Dysfunction and Mitophagy Impairment. *Frontiers in Microbiology*, *12*.

Shigeoka, T., Jung, H., Jung, J., Turner-Bridger, B., Ohk, J., Lin, J. Q., Amieux, P. S., & Holt,

C. E. (2016). Dynamic Axonal Translation in Developing and Mature Visual Circuits. *Cell*, *166*(1), 181–192. https://doi.org/10.1016/j.cell.2016.05.029

Shulla, A., & Randall, G. (2015). Spatiotemporal analysis of hepatitis C virus infection. *PLoS Pathog*, *11*(3), e1004758. https://doi.org/10.1371/journal.ppat.1004758

Simmonds, P., Williams, S., & Harvala, H. (2021). Understanding the outcomes of COVID-19 - does the current model of an acute respiratory infection really fit? *The Journal of General Virology*, *102*(3). https://doi.org/10.1099/jgv.0.001545

Singer, Z. S., Ambrose, P. M., Danino, T., & Rice, C. M. (2021). Quantitative measurements of early alphaviral replication dynamics in single cells reveals the basis for superinfection exclusion. *Cell Syst*, *12*(3), 210–219 e3. https://doi.org/10.1016/j.cels.2020.12.005

Skeath, J. B., & Thor, S. (2003). Genetic control of Drosophila nerve cord development. *Current Opinion in Neurobiology*, *13*(1), 8–15. https://doi.org/10.1016/s0959-4388(03)00007-2

Smibert, P., Miura, P., Westholm, J. O., Shenker, S., May, G., Duff, M. O., Zhang, D., Eads, B. D., Carlson, J., Brown, J. B., Eisman, R. C., Andrews, J., Kaufman, T., Cherbas, P., Celniker, S. E., Graveley, B. R., & Lai, E. C. (2012). Global patterns of tissue-specific alternative polyadenylation in Drosophila. *Cell Rep*, *1*(3), 277–289. https://doi.org/10.1016/j.celrep.2012.01.001

Smith, T. S., Heger, A., & Sudbery, I. (2017). UMI-tools: Modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Research*, gr.209601.116. https://doi.org/10.1101/gr.209601.116

Snijder, E. J., Decroly, E., & Ziebuhr, J. (2016). The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Advances in Virus Research*, *96*, 59–126. https://doi.org/10.1016/bs.aivir.2016.08.008

Sola, I., Almazán, F., Zúñiga, S., & Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annual Review of Virology*, *2*(1), 265–288. https://doi.org/10.1146/annurev-virology-100114-055218

Spana, E. P., & Doe, C. Q. (1995). The prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in Drosophila. *Development (Cambridge, England)*, *121*(10), 3187–3195. https://doi.org/10.1242/dev.121.10.3187

Stoiber, M. H., Olson, S., May, G. E., Duff, M. O., Manent, J., Obar, R., Guruharsha, K. G., Bickel, P. J., Artavanis-Tsakonas, S., Brown, J. B., Graveley, B. R., & Celniker, S. E. (2015). Extensive cross-regulation of post-transcriptional regulatory networks in Drosophila. *Genome Res*, *25*(11), 1692–1702. https://doi.org/10.1101/gr.182675.114

Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: A generalist algorithm for cellular segmentation. *Nat Methods*, *18*(1), 100–106. https://doi.org/10.1038/s41592-020-01018-x

Stueland, M., Wang, T., Park, H. Y., & Mili, S. (2019). RDI Calculator: An Analysis Tool to Assess RNA Distributions in Cells. *Sci Rep*, *9*(1), 8267. https://doi.org/10.1038/s41598-019-44783-2

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N. M., & Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature*, *519*(7544), 491–494. https://doi.org/10.1038/nature14280

Sun, J., Bonaguidi, M. A., Jun, H., Guo, J. U., Sun, G. J., Will, B., Yang, Z., Jang, M.-H., Song, H., Ming, G., & Christian, K. M. (2015). A septo-temporal molecular gradient of Sfrp3 in the dentate gyrus differentially regulates quiescent adult hippocampal neural stem cell activation. *Molecular Brain*, *8*(1), 52. https://doi.org/10.1186/s13041-015-0143-9

Suzuki, A., & Ohno, S. (2006). The PAR-aPKC system: Lessons in polarity. *Journal of Cell Science*, *119*(Pt 6), 979–987. https://doi.org/10.1242/jcs.02898

Svitkin, Y. V., Yanagiya, A., Karetnikov, A. E., Alain, T., Fabian, M. R., Khoutorsky, A., Perreault, S., Topisirovic, I., & Sonenberg, N. (2013). Control of translation and miRNA-dependent repression by a novel poly(A) binding protein, hnRNP-Q. *PLoS Biology*, *11*(5), e1001564. https://doi.org/10.1371/journal.pbio.1001564

Syed, M. H., Mark, B., & Doe, C. Q. (2017). Steroid hormone induction of temporal gene expression in Drosophila brain neuroblasts generates neuronal and glial diversity. *Elife*, *6*. https://doi.org/10.7554/eLife.26287

Sysoev, V. O., Fischer, B., Frese, C. K., Gupta, I., Krijgsveld, J., Hentze, M. W., Castello, A., & Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nature Communications*, *7*, 12128. https://doi.org/10.1038/ncomms12128

Targett-Adams, P., Boulant, S., & McLauchlan, J. (2008). Visualization of double-stranded RNA in cells supporting hepatitis C virus RNA replication. *J Virol*, *82*(5), 2182–2195. https://doi.org/10.1128/JVI.01565-07

Taylor, M. W. (2014). *Viruses and Man: A History of Interactions*. Springer International Publishing. https://doi.org/10.1007/978-3-319-07758-1

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E. J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A. J., Engelbrecht, S., Van Zyl, G., … de Oliveira, T. (2020). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*, 2020.12.21.20248640. https://doi.org/10.1101/2020.12.21.20248640

Thorne, L. G., Bouhaddou, M., Reuschl, A.-K., Zuliani-Alvarez, L., Polacco, B., Pelin, A., Batra, J., Whelan, M. V. X., Ummadi, M., Rojc, A., Turner, J., Obernier, K., Braberg, H., Soucheray, M., Richards, A., Chen, K.-H., Harjai, B., Memon, D., Hosmillo, M., … Krogan,

N. J. (2021). Evolution of enhanced innate immune evasion by the SARS-CoV-2 B.1.1.7 UK variant. *bioRxiv*, 2021.06.06.446826. https://doi.org/10.1101/2021.06.06.446826

Titlow, J. S., Kiourlappou, M., Palanca, A., Lee, J. Y., Gala, D. S., Ennis, D., Yu, J. J. S., Young, F. L., Pinto, D. M. S., Garforth, S., Francis, H. S., Strivens, F., Mulvey, H., Dallman-Porter, A., Thornton, S., Arman, D., Järvelin, A. I., Thompson, M. K., Kounatidis, I., … Davis, I. (2022). *Systematic analysis of YFP gene traps reveals common discordance between mRNA and protein across the nervous system* (p. 2022.03.21.485142). bioRxiv. https://doi.org/10.1101/2022.03.21.485142

Titlow, J. S., Yang, L., Parton, R. M., Palanca, A., & Davis, I. (2018). Super-Resolution Single Molecule FISH at the Drosophila Neuromuscular Junction. *Methods Mol Biol*, *1649*, 163–175. https://doi.org/10.1007/978-1-4939-7213-5_10

Titlow, J., Robertson, F., Järvelin, A., Ish-Horowicz, D., Smith, C., Gratton, E., & Davis, I. (2020). Syncrip/hnRNP Q is required for activity-induced Msp300/Nesprin-1 expression and new synapse formation. *The Journal of Cell Biology*, *219*(3), e201903135. https://doi.org/10.1083/jcb.201903135

Torres-Méndez, A., Pop, S., Bonnal, S., Almudi, I., Avola, A., Roberts, R. J. V., Paolantoni, C., Alcaina-Caro, A., Martín-Anduaga, A., Haussmann, I. U., Morin, V., Casares, F., Soller, M., Kadener, S., Roignant, J.-Y., Prieto-Godino, L., & Irimia, M. (2022). Parallel evolution of a splicing program controlling neuronal excitability in flies and mammals. *Science Advances*, *8*(4), eabk0445. https://doi.org/10.1126/sciadv.abk0445

Tushev, G., Glock, C., Heumuller, M., Biever, A., Jovanovic, M., & Schuman, E. M. (2018). Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron*, *98*(3), 495–511 e6. https://doi.org/10.1016/j.neuron.2018.03.030

Tutucci, E., Livingston, N. M., Singer, R. H., & Wu, B. (2018). Imaging mRNA In Vivo, from Birth to Death. *Annu Rev Biophys*, *47*, 85–106. https://doi.org/10.1146/annurev-biophys-070317-033037

Ule, J., Hwang, H. W., & Darnell, R. B. (2018). The Future of Cross-Linking and Immunoprecipitation (CLIP). *Cold Spring Harb Perspect Biol*, *10*(8). https://doi.org/10.1101/cshperspect.a032243

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., & Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science (New York, N.Y.)*, *302*(5648), 1212–1215. https://doi.org/10.1126/science.1090095

V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat Rev Microbiol*, *19*(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C.,

Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., ... Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature*, *583*(7818), 711–719. https://doi.org/10.1038/s41586-020-2077-3

Van Nostrand, E. L., Gelboin-Burkhart, C., Wang, R., Pratt, G. A., Blue, S. M., & Yeo, G. W. (2017). CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. *Methods (San Diego, Calif.)*, *118–119*, 50–59. https://doi.org/10.1016/j.ymeth.2016.12.007

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., & Yeo, G. W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*, *13*(6), 508–514. https://doi.org/10.1038/nmeth.3810

Van Nostrand, E. L., Pratt, G. A., Yee, B. A., Wheeler, E. C., Blue, S. M., Mueller, J., Park, S. S., Garcia, K. E., Gelboin-Burkhart, C., Nguyen, T. B., Rabano, I., Stanton, R., Sundararaman, B., Wang, R., Fu, X.-D., Graveley, B. R., & Yeo, G. W. (2020). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biology*, *21*(1), 90. https://doi.org/10.1186/s13059-020-01982-9

Vegivinti, C. T. R., Evanson, K. W., Lyons, H., Akosman, I., Barrett, A., Hardy, N., Kane, B., Keesari, P. R., Pulakurthi, Y. S., Sheffels, E., Balasubramanian, P., Chibbar, R., Chittajallu, S., Cowie, K., Karon, J., Siegel, L., Tarchand, R., Zinn, C., Gupta, N., ... Touchette, J. (2022). Efficacy of antiviral therapies for COVID-19: A systematic review of randomized controlled trials. *BMC Infectious Diseases*, *22*(1), 107. https://doi.org/10.1186/s12879-022-07068-0

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., ... Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, *593*(7858), 266–269. https://doi.org/10.1038/s41586-021-03470-x

von Kügelgen, N., & Chekulaeva, M. (2020). Conservation of a core neurite transcriptome across neuronal types and species. *WIREs RNA*, *11*(4), e1590. https://doi.org/10.1002/wrna.1590

Wang, H., Ouyang, Y., Somers, W. G., Chia, W., & Lu, B. (2007). Polo inhibits progenitor self-renewal and regulates Numb asymmetry by phosphorylating Pon. *Nature*, *449*(7158), 96–100. https://doi.org/10.1038/nature06056

Wang, S. (2019). Single Molecule RNA FISH (smFISH) in Whole-Mount Mouse Embryonic Organs. *Current Protocols in Cell Biology*, *83*(1), e79. https://doi.org/10.1002/cpcb.79

Weber, F., Wagner, V., Rasmussen, S. B., Hartmann, R., & Paludan, S. R. (2006). Double-

stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. *J Virol*, *80*(10), 5059–5064. https://doi.org/10.1128/JVI.80.10.5059-5064.2006

Wei, L., Lee, S., Majumdar, S., Zhang, B., Sanfilippo, P., Joseph, B., Miura, P., Soller, M., & Lai, E. C. (2020). Overlapping Activities of ELAV/Hu Family RNA Binding Proteins Specify the Extended Neuronal 3' UTR Landscape in Drosophila. *Mol Cell*, *80*(1), 140–155 e6. https://doi.org/10.1016/j.molcel.2020.09.007

Weigmann, K., & Lehner, C. F. (1995). Cell fate specification by even-skipped expression in the Drosophila nervous system is coupled to cell cycle progression. *Development (Cambridge, England)*, *121*(11), 3713–3721. https://doi.org/10.1242/dev.121.11.3713

Wheeler, E. C., Van Nostrand, E. L., & Yeo, G. W. (2018). Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *WIREs RNA*, *9*(1), e1436. https://doi.org/10.1002/wrna.1436

Wilcox, A. H., Delwart, E., & Diaz-Munoz, S. L. (2019). Next-generation sequencing of dsRNA is greatly improved by treatment with the inexpensive denaturing reagent DMSO. *Microb Genom*, *5*(11). https://doi.org/10.1099/mgen.0.000315

Wing, P. A. C., Keeley, T. P., Zhuang, X., Lee, J. Y., Prange-Barczynska, M., Tsukuda, S., Morgan, S. B., Harding, A. C., Argles, I. L. A., Kurlekar, S., Noerenberg, M., Thompson, C. P., Huang, K. A., Balfe, P., Watashi, K., Castello, A., Hinks, T. S. C., James, W., Ratcliffe, P. J., … McKeating, J. A. (2021). Hypoxic and pharmacological activation of HIF inhibits SARS-CoV-2 infection of lung epithelial cells. *Cell Rep*, *35*(3), 109020. https://doi.org/10.1016/j.celrep.2021.109020

Wolff, G., Melia, C. E., Snijder, E. J., & Bárcena, M. (2020). Double-Membrane Vesicles as Platforms for Viral Replication. *Trends in Microbiology*, *28*(12), 1022–1033. https://doi.org/10.1016/j.tim.2020.05.009

Wu, Y. C., Chen, C. H., Mercer, A., & Sokol, N. S. (2012). Let-7-complex microRNAs regulate the temporal identity of Drosophila mushroom body neurons via chinmo. *Dev Cell*, *23*(1), 202–209. https://doi.org/10.1016/j.devcel.2012.05.013

Yang, C. P., Samuels, T. J., Huang, Y., Yang, L., Ish-Horowicz, D., Davis, I., & Lee, T. (2017). Imp and Syp RNA-binding proteins govern decommissioning of Drosophila neural stem cells. *Development*, *144*(19), 3454–3464. https://doi.org/10.1242/dev.149500

Yang, L., Titlow, J., Ennis, D., Smith, C., Mitchell, J., Young, F. L., Waddell, S., Ish-Horowicz, D., & Davis, I. (2017). Single molecule fluorescence in situ hybridisation for quantitating post-transcriptional regulation in Drosophila brains. *Methods*, *126*, 166–176. https://doi.org/10.1016/j.ymeth.2017.06.025

Yano, M., Hayakawa-Yano, Y., & Okano, H. (2016). RNA regulation went wrong in neurodevelopmental disorders: The example of Msi/Elavl RNA binding proteins. *Int J Dev Neurosci*,

55, 124–130. https://doi.org/10.1016/j.ijdevneu.2016.01.002

Yasugi, T., & Nishimura, T. (2016). Temporal regulation of the generation of neuronal diversity in Drosophila. *Dev Growth Differ*, *58*(1), 73–87. https://doi.org/10.1111/dgd.12245

Yuen, M.-F., Chen, D.-S., Dusheiko, G. M., Janssen, H. L. A., Lau, D. T. Y., Locarnini, S. A., Peters, M. G., & Lai, C.-L. (2018). Hepatitis B virus infection. *Nature Reviews. Disease Primers*, *4*, 18035. https://doi.org/10.1038/nrdp.2018.35

Zacharioudaki, E., Housden, B. E., Garinis, G., Stojnic, R., Delidakis, C., & Bray, S. J. (2016). Genes implicated in stem cell identity and temporal programme are directly targeted by Notch in neuroblast tumours. *Development (Cambridge, England)*, *143*(2), 219–231. https://doi.org/10.1242/dev.126326

Zarnegar, B. J., Flynn, R. A., Shen, Y., Do, B. T., Chang, H. Y., & Khavari, P. A. (2016). irCLIP platform for efficient characterization of protein-RNA interactions. *Nat Methods*, *13*(6), 489–492. https://doi.org/10.1038/nmeth.3840

Zhang, J., Zhang, Y., Kang, J.-Y., Chen, S., He, Y., Han, B., Liu, M.-F., Lu, L., Li, L., Yi, Z., & Chen, L. (2021). Potential transmission chains of variant B.1.1.7 and co-mutations of SARS-CoV-2. *Cell Discovery*, *7*(1), 1–10. https://doi.org/10.1038/s41421-021-00282-1

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., Kingston, R. E., Borowsky, M., & Lee, J. T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Molecular Cell*, *40*(6), 939–953. https://doi.org/10.1016/j.molcel.2010.12.011

Zhao, M., Yu, Y., Sun, L. M., Xing, J. Q., Li, T., Zhu, Y., Wang, M., Yu, Y., Xue, W., Xia, T., Cai, H., Han, Q. Y., Yin, X., Li, W. H., Li, A. L., Cui, J., Yuan, Z., Zhang, R., Zhou, T., … Li, T. (2021). GCG inhibits SARS-CoV-2 replication by disrupting the liquid phase condensation of its nucleocapsid protein. *Nat Commun*, *12*(1), 2114. https://doi.org/10.1038/s41467-021-22297-8

Zhu, S., Lin, S., Kao, C.-F., Awasaki, T., Chiang, A.-S., & Lee, T. (2006). Gradients of the Drosophila Chinmo BTB-zinc finger protein govern neuronal temporal identity. *Cell*, *127*(2), 409–422. https://doi.org/10.1016/j.cell.2006.08.045