UNIVERSITY OF CALIFORNIA, SAN DIEGO

Integrative analysis of the IGF2BP/IMP family of RNA binding proteins in human embryonic stem cells

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Anne Conway

Committee in charge:

Professor Gene Yeo, Chair
Professor Leanne Jones, Co-Chair
Professor Lawrence Goldstein
Professor Bing Ren
Professor Jean Wang
Professor Karl Willert

2014

The Dissertation of Anne Conway is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Leanne Jones for giving me the opportunity to perform my thesis work in the lab and whose support and guidance have been invaluable during my journey. I would also like to acknowledge professors Gene Yeo and Karl Willert for their mentoring as well as sharing their space and resources with me during the final year of my PhD.

I would like to acknowledge the past and present members of the Jones, Yeo, and Willert labs for your continued encouragement and support. Our thoughtful conversations and scientific discussions have greatly aided my graduate school experience and career development.

I would also like to acknowledge my parents, friends and family. Without their patience, love and support throughout the entirety of my academic career I would not have been able to accomplish all that I have.

Chapter 2, in full, is an adaptation of material being submitted for publication as Conway, A.E., Wilbert. M.L.W., Landais, S., Sundaraman, B., Liang, T.Y., Pratt G., Essex, A., Hoon, S., Jones, D.L., Yeo, G.W., "IGF2BP/IMP family of RNA-binding proteins control an RNA network in pluripotent stem cells to maintain cell survival". The dissertation author is a primary investigator and author for this paper.

VITA

EDUCATION

University of California, San Diego	Ph.D.	Biomedical Sciences	2014
University of California, Berkeley	B.A.	Molecular and Cellular Biology	2006

RESEARCH EXPERIENCE

2008-2014 Graduate Student, Biomedical Sciences Program, University of California, San Diego, La Jolla, CA.
Determined the function of a family of RNA binding proteins in human pluripotent stem cells.

2006-2008 Research Associate II, Department of Molecular Cellular and Developmental Biology, University of California, Los Angeles, Westwood, CA.
Investigated the role of Retinoic Acid induced differentiation in germ cell tumor formation and established an in vivo germ cell tumor model in the lab.

2005-2006 Research Assistant I, Department of Plant and Micbrobial Biology, University of California, Berekely. Berekley, CA.
Performed PCR screens to test for insertion of a transgene in a genetically modified strain of sorghum.

HONORS AND AWARDS

2013 Best Poster Presentation Award UCLA Department of Molecular, Cellular, and Developmental Biology Annual Meeting

2010-2013 Genetics Training Grant Fellowship (NIH T32 GM00866), University of California, San Diego

2005-2006 Biology Scholars Program Fellowship, University of California, Berkeley

2002-2006 Alumni Leadership Scholarship, University of California, Berkeley

PUBLICATIONS

Kim Y, Deshpande A, Dai Y, Kim JJ, Lindgren A, **Conway A**, Clark AT, Wong DT. Cyclin-dependent kinase 2-associating protein 1 Commits Murine Embryonic Stem Cell Differentiation through Retinoblastoma Protein Regulation. J Biol Chem. June 2009. Epub ahead of print.

Karumbayaram S, Novitch BG, Patterson M, Umbach JA, Richter L, Lindgren A, **Conway AE**, Clark AT, Goldman SA, Plath K, Wiedau-Pazos M, Kornblum HI,

Lowry WE. Directed differentiation of human-induced pluripotent stem cells generates active motor neurons. *Stem Cells.* April 2009.

Park TS, Galic Z, **Conway AE**, Lindgren A, van Handel BJ, Magnusson M, Richter L, Teitell MA, Mikkola HK, Lowry WE, Plath K, Clark AT. Derivation of primordial germ cells from human embryonic and induced pluripotent stem cells is significantly improved by coculture with human fetal gonadal cells. *Stem Cells.* April 2009.

Xie W, Song C, Young NL, Sperling AS, Xu F, Sridharan R, **Conway AE**, Garcia BA, Plath K, Clark AT, Grunstein M. Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Molecular Cell.* February 2009.

Conway, AE., Lindgren, A., Galic, Z., Pyle, D., Wu, H., Zack, J., Pelligrini, M., Teitell, MA., Clark, AT. A self-renewal program controls the expansion of genetically unstable cancer stem cells in pluripotent stem cell-derived tumors. *Stem Cells.* January 2009.

TEACHING EXPERIENCE

2013	Volunteer Lecturer, Stem Cell Awareness Day, Oceanside High School, Oceanside, CA.
2012	Volunteer Lecturer, Stem Cell Awareness Day, La Jolla High School, La Jolla, CA.
2012	Invited Lecturer for BIOM 201, Seminars in Biomedical Research University of California, San Diego, La Jolla CA.
2010	Teaching Assistant, BICD 130, Embryos, Genes, and Development, University of California, San Diego, La Jolla CA.

ABSTRACT OF THE DISSERTATION

Integrative analysis of the IGF2BP/IMP family of RNA binding proteins
in human embryonic stem cells

by

Anne Conway

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2014

Professor Gene Yeo, Chair
Professor Leanne Jones, Co-Chair

Stem cell research is revolutionizing the way scientists think about mammalian development. It is a field full of promise to treat many debilitating and degenerative diseases. Regulation of stem cells by transcription factors that bind to specific DNA loci to control gene regulatory networks have been well studied in various stem cell

populations. However, the mechanisms by which the regulation of post-transcriptional RNA processing influences the timely, highly specialized differentiation of stem cells are not well understood. Only recently have researchers started focusing attention on the importance of RNA regulation in the control of gene expression and, thus, to impact human health.

RNA binding proteins (RBPs) bind to and regulate RNA metabolism at multiple levels from processing within the nucleus, nuclear export, and transport to organelles within the cytoplasm. In addition, RBPs can regulate RNA stability and protein translation. Techniques to study and identify the RNA substrates regulated by RNA binding proteins have only recently been developed within the last decade, with the availability of high-throughput sequencing revolutionizing established techniques, enabling transcriptome-wide views of RNA biology.

My dissertation research focuses on the function of the IGF2BP/IMP family of RNA binding proteins. In particular, I have investigated the role of IGF2BP1/IMP1, in early human development using human embryonic stem cells (hESCs) as a model system. I have used CLIP-seq to identify the endogenous targets of IMP1 in hESCs and have determined the mechanism of regulation for the novel target *ITGB5*. Additionally, I have uncovered a role for IMP1 in early neural progenitor cell populations and in localization of target RNAs. Ongoing studies comparing and contrasting the various targets and functions of IMP1 and IMP2 in hESCs at the genome-wide level will pave the way for future investigators to dissect the individual functions for each of these RBPs. This work will provide a resource for those aiming

to understand the mechanisms of how these proteins control their target RNAs during development and tumorigenesis.

CHAPTER 1 - INTRODUCTION

RBPs in development and disease

Every cell in an organism contains the same genetic information, however, it is how that genetic information is interpreted that leads to distinct cell types and functions. One could argue the first layer of distinction is at the level of gene transcription when cell type specific transcription factors transcribe the genetic information into RNA. For it is this RNA message that then gets translated into protein, the action-packed machines of the cell, that allows for true cellular distinction. Alternatively, RNA binding proteins (RBPs) are the master regulators of RNA metabolism, and thus may be the true interpreters of the genetic code. RBPs control whether the RNA message is degraded, processed correctly into mRNA, localized to the correct cellular organelle, and translated (or not) and thus it is transcriptome regulation by RBPs that ultimately determines cell fate.

A tangible example of the consequences of deficient RBP function is the onset of several neurodegenerative diseases. Spinal muscular atrophy (SMA) is characterized by degeneration of lower motor neurons and severe muscle atrophy (Lefebvre et al., 1995). A primary cause for this disease is homozygous deletion in the Survival of Motor Neuron 1 (SMN1) gene. SMN1 is a RBP essential for the assembly of small nuclear ribonucleoproteins (snRNPs) that interact to form the spliceosome (Pellizzoni et al., 1998). Loss of SMN1 leads to extensive splicing defects in which leads to

the vulnerability and eventual demise of affected motor neurons. In addition to SMA, a subset of Autism Spectrum Disorders (ASDs) can be caused by reduction of the RBPs Fragile X Mental Retardation protein (FMRP) and RNA Binding Protein FOX 1 (RBFOX1) (Yu et al., 1991; Oberle et al., 1991; Kremer et al., 1991; Verkerk et al., 1991; Weyn-Vanhentenryck, 2014). Fragile X syndrome (FXS) is a neurodevelopmental disorder with autistic phenotypes caused by depletion of FMRP. FMRP is necessary for the formation of neural synapses and its transcription is reduced via histone hypoacetylation and hypermethylation, as well as DNA hypermethylation (Coffee et al., 1999; Kumari et al., 2012). Due to reduced FMRP levels, dendrites of neurons from FXS patients have immature, abnormally long and dense spines, supporting a role for FMRP in synaptic pruning and maturation. In addition to FMRP, reduced expression of RBFOX1 has recently been associated with autism. RBFOX1 is necessary for appropriate neuronal excitation through control of the alternative splicing of several genes, some of which are known as autism susceptibility genes (Weyn-Vanhentenryck et al., 2014). Additional neurodegenerative diseases associated with mis-regulation of RBPs include Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD), of which the RBPs TDP-43 and FUS/TAF15/EWS are implicated, as well as Myotonic Dystrophy, (DM), which is caused by sequestration of MBNL1 (Neumann et al., 2006; Mankodi et al., 2002; Wang et al., 2012). While RBPs in neurodegenerative diseases are certainly a stimulating case study, what about a role for RBPs in very early development of an

organism? There is not as much known about this subject, which is one of the reasons I chose to pursue it further.

Human embryonic stem cells (hESCs) are pluripotent cells that propagate readily in culture as undifferentiated cells and can be induced to differentiate into all of the various cell types that make up the mammalian embryo. Their differentiation program *in vitro* very closely approximates the developmental pathways and timing inside a human embryo, which makes them an excellent model for studying normal human development. Additionally, their fast proliferation rate is ideal for performing large-scale genome wide studies and also provides a substantial source material to make more specialized cell types such as neurons and cardiomyocytes. These properly specialized cell types can be used as a model for drug screening and toxicity studies, which is advantageous to the typical transformed tumor cell line since they actually represent the tissues the drug is designed to cure. For these reasons I have focused on using human embryonic stem cells as a model system to investigate the roles of RBPs in human development and disease.

RBPs in pluripotent stem cells: RBFOX2, MBNL, LIN28

Due to decreases in the cost of sequencing and improvements in technology efficiency transcriptome-wide studies of RBPs and their target RNAs have recently become achievable. Using **c**ross-**l**inking followed by **i**mmuno**p**recipitation of protein-RNA complexes and high-throughput **seq**uencing of isolated transcripts (CLIP-seq), we and others have identified the genome-wide binding preferences of many different RBPs. These studies described, primarily a network of RNAs that are targeted by

RBPs in hESCs, and provided insight into how these RBPs regulate their target RNA networks. The first of such studies describes the RNA network regulated by the splicing factor RBFOX2. Interestingly, RBFOX2 is expressed in hESCs while RBFOX1 and 3 are not, indicating a role in pluripotent stem cells that is specific to RBFOX2 (Yeo et al., 2009). The significant RBFOX2 binding sites clustered around alternatively spliced (AS) exons that were exclusive to stem cells. Gene ontology analysis found that RBFOX2 binding sites are enriched in other splicing factors suggesting that RBFOX2 controls an entire stem cell specific splicing program. The functional consequence of RBFOX2 binding was assayed by RT-PCR and it became apparent that there was a position-dependent affect; RBFOX2 binding upstream represses AS exon usage while binding downstream leads to inclusion. The resulting survival phenotype from loss of RBFOX2 in hESCs suggested RBFOX2 is required for cell survival specifically in stem cells, but not in other differentiated cell types, further highlighting the importance of a stem cell specific program.

As mentioned previously, sequestration of MBNL1 in muscle cells can lead to Myotonic Dystrophy. Additionally, there is a function for MBNL1 (and MBNL2) in regulating pluripotency. Intriguingly, although MBNL1 and 2 are expressed at low levels in hESCs, MBNL overexpression in hESCs led to premature differentiation and knockdown in differentiated cells resulted in reversion towards an embryonic stem cell-like state (Han et al., 2013). This latter affect is also correlated with an increase in reprogramming efficiency, thus MBNL1 and 2 have a negative affect on pluripotency (Han et al., 2013). The affects of MBNL1 and 2 in stem cells were mediated through

regulation of AS patterns specific to hESCs (Dai et al., 2013). A primary example of this is MBNL regulation of the hESC specific form of the forkhead family transcription factor FOXP1 (Gabut et al., 2011). The hESC specific FOXP1 has an altered DNA binding capacity that stimulates the expression of the pluripotency factors OCT4, NANOG, NR5A2, and GDF3, while concomitantly repressing genes required for ESC differentiation (Gabut et al., 2011). Conserved MBNL CLIP-seq binding sites were present adjacent to the ESC-specific FOXP1 exon 18b in hESCs and loss of MBNL in differentiated cells led to inclusion of this exon, thus reverting the expression patterns to a more ESC-like state (Han et al., 2013).

Finally, LIN28 is the first RBP that comes to mind for many people when discussing RBPs and pluripotency because LIN28 is the only non-transcritption factor to date that has been directly able to reprogram differentiated cells to the pluripotent state (Yu et al., 2007). Additionally, at the single cell level during reprogramming stochastic activation of LIN28 is the only RBP and one of the top four most associated factors with successful reprogramming to the pluripotent state, which was surprising since it correlated much more efficiently than other well known pluripotency transcription factors such as FBXO15, FGF4, and OCT4 (Buganim et al., 2012). LIN28 primarily acts through regulation of its target RNAs in hESC to promote proliferation. Whether or not loss of LIN28 leads to differentiation is contested in the literature (though the studies were done in different cell lines) and may be different in "naïve" mESC and "primed" hESC, likely due to the different pluripotent state of the two cell types (Nichols and Smith, 2009). Genome-wide analysis of LIN28 binding

targets in hESC was first performed by Peng et al., 2011 using RIP-seq and has since been replicated by Wilbert et al., 2012 using CLIP-seq. The primary conclusions from these reports are that LIN28 binds both miRNAs and mRNAs and regulates translation of splicing factor abundance resulting in thousands of splicing changes downstream. Additionally, LIN28 binds a largely overlapping set of targets in hESC and other cell types where CLIP-seq has been performed such as 293 cells (Wilbert et al., 2012, Hafner et al., 2013, Graf et al., 2013) suggesting that LIN28 binding of target RNAs may not be cell-type specific. Genome-wide studies for LIN28 and the other previously mentioned RBPs reveal avenues by which these proteins directly impact gene regulatory networks through regulation of their mRNA targets in hESCs. These studies provide a valuable framework for future characterization of the molecular roles of other RNA binding proteins in human pluripotent stem cells.

IGF2BPs are members of an evolutionarily conserved family of RNA binding proteins

Discovery

The IGF2BP family of RNA binding proteins was discovered simultaneously in different model organisms and systems. Jeff Ross' lab was interested in determining how cytoplasmic mRNA stability was regulated and how elements specifically in the coding region of the *MYC* oncogene controlled its stability. It had been previously discovered that a sequence within the coding amino acids determined the rapid turnover of *MYC* RNA (Wisdom and Lee, 1991). To test whether this region was bound by a trans-acting factor, a competition experiment was performed. Polysomes

containing *MYC* mRNA were isolated from a human erythroleukemia cell line (k562) and incubated in a cell-free RNA decay system (Bernstein et al., 1992). These reactions were supplemented with individual exogenous competitor RNA fragments corresponding to different portions of *MYC* mRNA. The idea was if a trans-acting factor bound to a specific region of the *MYC* mRNA, the factor might be titrated away by a competitor sequence homologous to that region, thereby altering the *MYC* half-life. Indeed, they discovered a 182bp sense strand competitor fragment from the *MYC* C-terminus, that, when added to the cell-free system, induced destabilization of *MYC* mRNA (Bernstein et al., 1992). To determine whether this was indeed due to a RBP binding the *MYC* coding region stability determinant (CRD), a gel-shift assay was performed with the 182bp CRD radiolabeled and incubated with polysomes from k562 cells. Indeed, there was a slower migrating complex at around ~75bp that was sensitive to the addition of proteases and the protein was then named CRD-BP.

Around that same time, Robert Singer's lab was interested in determining how mRNAs were localized to different areas of the cell, contributing to direct cellular migration and polarization. They were especially interested in how a 54 nucleotide region of the *ACTB* 3′ UTR termed the "zipcode" allowed for asymmetrical localization of the β-actin transcript to the leading edge of the lamellepodium in chicken embryo fibroblasts (CEFs) (Kislauskis et al., 1994). To test whether it was a trans-acting RBP, zipcode-containing probes were labeled with biotin, immobilized onto a streptavidin membrane and incubated overnight with CEF cell extracts. Bound proteins were then eluted from the column and a ~68kD zipcode-binding protein

(ZBP1) was identified (Ross et al., 1997). Further characterization of the ZBP1-zipcode interaction was performed and determined to be consistent with previous findings that *ACTB* localization was dependent on ATP (energy) and the actin cytoskeleton, but not protein translation and microtubules (Sundell et al., 1990; Sundell et al., 1991; Latham et al., 1994; Ross et al., 1997).

In addition to the discovery of ZBP1 and CRD-BP, other IMP family members have been discovered and studied in *Xenopus* and *Drosophila*. It is thought the highly conserved family was created by 2 gene duplications between the separation of vertebrates and invertebrates (Runge et al., 2000)(Figure 1A). The *Xenopus* othologue is called Vg1RBP due to binding and polarization of the *Vg1* RNA to the vegetal pole of developing *Xenopus* oocytes (Schwartz et al., 1992). Vg1RBP binds the *Vg1* 3′ UTR and allows for transport to the vegetal pole via microtubules (Deshler et al., 1998; Havin et al., 1998). Upon similarity in structure to the *Xenopus* Vg1 Rbp, the zebrafish family member was also called Vg1 RBP (Figure 1A)(Zhang et al., 1999). *Drosophila* IMP (*dIMP*) was discovered to also be very highly conserved with the other family members and was also expressed in early development (Runge et al., 2000). Specifically in *Drosophila* and not necessarily the other family members there was a striking expression in the developing brain and neural tissues. Further evaluation of dIMP mutants during development demonstrated a requirement for dIMP protein in synaptogenesis (Boylan et al., 2008). Mutant animals are unable to stand up after following over and are not able to crawl up the side of the vial. Motility screens also identify IMP acting in oogenesis, although this function is not essential. IMP localized

transcripts in oogenesis are still able to be transported to the correct cellular compartment in its absence (Boylan et al., 2008). In *Drosophila*, loss-of-function Imp mutations are zygotic lethal, and mutants die later as pharate adults reiterating a requirement for IMP proteins in normal embryonic development.

Finally last, but not least was the characterization of the IGF2BP family of RBPs with their namesake target *IGF2* in early murine development (Nielsen J., 1999). The observation was made that specific isoforms of *IGF2* are spatially and temporally regulated at the level of translation during development and it was hypothesized that a trans-acting factor may be the root cause. To test if trans-acting factors were indeed regulating the *IGF2* RNA, differentially expressed 5′ UTRs were incubated with cytoplasmic cell lysate from RD rhabdomysoarcoma cells and UV crosslinked and then RNAse digested and run on a gel (Nielsen J., 1999). A strong band at 69kD was purified and surprisingly was found to be very highly conserved with the already discovered ZBP1 and CRD-BP RBPs. Furthermore, based on the protein sequences there were 2 additional independent mammalian family members that shared high homology with one another and shared the 6 RNA binding domains, thus these proteins became called IGF2BP1-3. The high evolutionary conservation of the IMP RBPs across phyla underscores the functional importance of these RBPs in normal development and disease.

IMP family RBPs share a similar modular structure

The mammalian IMP RNA binding proteins contain six different RNA binding domains, 2 proximal RRM domains at the N terminus and 4 KH domains in the C-terminus of the protein (Figure 1B)(Nielsen et al., 1999). The 4 KH domains are very highly conserved across species and have individually been shown to facilitate binding target RNAs (Nielsen et al., 2002; Wachter et al., 2013). Additionally, IMP proteins have 2 nuclear export signals (NES) and are able to shuttle in and out of the nucleus (Nielsen, 2003). However, they have no known nuclear localization signals, which suggests that their transport into the nucleus is either RNA dependent and/or facilitated through interactions with a co-factor. Localization of IMPs is dependent on the RNA binding domains as disruption of RNA binding or exogenous overexpression allows for IMP mis-localization within the nucleus (Nielsen et al., 2003). Mis-localization upon over-expression is important to consider when identifying endogenous interactions with target RNAs, as some studies that have been performed utilized over-expressed, tagged versions of the IMP proteins.

Post-transcriptional modifications and phosphorylation status have been investigated for the mammalian family members, IMP1-3. It was first discovered that ZBP1 (IMP1) phosphorylation by the Src family kinases at residue Tyr396 caused the release and subsequent translation of *ACTB* RNA at the leading edge of neuroblastoma cells (Huttlemaier et al., 2005). Interestingly, though IMP1 was shown to associate with *ACTB* in the nucleus mutations at this site had no effect on the localization of the *ACTB* mRNA in the cytoplasm; however, the binding efficiency with ZBP1 (IMP1)

was significantly reduced. These findings were significant, as it was the first study to investigate a mechanism for how IMP RBPs regulate target RNAs temporally and spatially *in vivo*. Further studies discussed later went on to discover specific mechanisms for this spatial and temporal regulation of *ACTB* in developing neurons.

Translational regulation of target RNAs by IMP family member phosphorylation is not only relevant for *ACTB* regulation, in fact, this seems to be a common mechanism for regulation of *IGF2* translation. Two recent studies by the Avruch lab investigating the role of first IMP2 and then also IMP1 and IMP3 phosphorylation in promoting *IGF2* translation were recently published (Dai et al., 2013; Dai et al., 2011). IMP2 is dually phosphorylated at Ser162 and Ser164 by mTOR in a rapamycin-inhibitable manner both *in vitro* as well as in RD Rhabdomyosarcoma cells (Dai et al., 2011). Phosphorylation at both sites promotes IMP2 binding to the IGF2 leader 3 mRNA 5′ UTR, and translational initiation of this mRNA through eIF-4E- and 5′ cap-independent internal ribosomal entry. The interaction of IMP2 and the *IGF2* mRNA is inhibited by mutations in the Ser162 and Ser164 residues as well as the addition of Rapamycin. Interestingly, IMP1 and IMP3 are also phosphorylated to promote translation of *IGF2*, but it appears to be through an independent mechanism. IMP1 and IMP3 are both phosphorylated by mTORC2 at residues Ser181 and Ser183, respectively; however, phosphorylation remains at these sites irrespective of rapamycin addition (Dai et al., 2013). Additionally, the phosphorylation status seems to be primarily dependent on RNA binding as addition of RNAse to RD extracts quickly diminished the phosphorylation signal without

interrupting IMP total protein levels (Dai et al., 2013). Furthermore, RNase-induced dephosphorylation can be fully rescued by phosphatase inhibitors, thus suggesting that slow turnover of some IMP1–RNA complexes limits susceptibility to dephosphorylation (Dai et al., 2013). Finally it was shown that IMP1 Ser181-phosphorylation was required for *IGF2* translation by cap-independent internal ribosomal entry in a similar fashion to IMP2. In summary, although *ACTB* and *IGF2* are only two RNA targets, this conserved method of translational regulation by phosphorylation status could likely be relevant for other IMP targets.

Evolutionarily conserved binding motifs- a summary of previously defined target sequences

Understanding the sequences and motifs bound by IMP RBPs remains widely under investigation. Early studies in various model organisms identified binding motifs based on repeated sequences enriched within target RNAs. For example, the zipcode motif that ZBP-1 binds includes a 6-nucleotide tandem repeat, ACACCC, which, when mutated, abrogates ZBP-1 binding (Kislauskis et al., 1993; Kislauskis et al., 1994; Ross et al., 1997). Vg1 RBP-binding sites contain a different hexanucleotide repeat, UUUCUA, and are generally AU rich (Deshler et al., 1998; Havin et al., 1998). Furthermore, an "IMP binding element" or IBE, UUUAY (Y being C or U), was determined by systematic evolution of ligands by exponential enrichment (SELEX) *in vitro* and is found 13 times within the *Drosophila Oscar (osk)* 3′ UTR (Munro et al.,

2006). This third motif seems to be a hybrid of the two described in chick (*ACTB*) and *Xenopus* (*Vg1*).

Upon crystallization of the KH3 and KH4 RBDs with the *ACTB* 3′ UTR, it became apparent that target RNA binding by the IMPs may be more dependent on a structural relationship than originally thought. ZBP1 (IMP1) KH34 was shown to assemble as a single structure by adopting an anti-parallel pseudodimer arrangement, which places the RNA-binding surfaces on opposing ends of the structure (Chao et al., 2010). This spatial organization of the latter two KH domains allows the protein to recognize its targets through sequence-specific contacts distributed over multiple stretches of RNA. Chao and colleagues further determined that ZBP1 (IMP1) binds specific nucleotides in a bi-partite motif in the *ACTB* 3′ UTR that consists of a GGACU "anchor" flanked by 2 AC rich regions on either side located within 7-30 bases (Patel et al., 2012). The fact that this structure is made of only 2 of the 6 RNA binding domains provides support for the provocative theory that all 6 of the IMP RBDs can associate with RNA and thus provide many different sequences and surfaces by which the protein can interact with other proteins and target RNAs.

The first and only genome-wide study of IMP RNA targets to date was performed as part of a larger resource study where a new technique was developed and several RBPs were analyzed. PAR-CLIP (**P**hoto**a**ctivatable-**R**ibonucleoside-Enhanced **C**ross**l**inking and **I**mmuno**p**recipitation) was designed to overcome the shortfall of other CLIP techniques where the location of the RBP-bound crosslinked nucleotide is not easily identifiable. PAR-CLIP accomplishes this by incorporating 4-thiouridine

(4SU) into transcripts of cultured cells and then identifying the precise RBP binding sites by scoring for thymidine (T) to cytidine (C) transitions following cDNA sequencing (Hafner et al., 2010). PAR-CLIP was performed in HEK293 cells by over-expressing FLAG/HA-tagged IMPs 1-3 and immunoprecipitating the bound RNA targets followed by cDNA library prep and sequencing. An analysis of the IMP bound sequences identified the tetramer CAUH (H = A, U, or C) as the most highly enriched motif and this sequence was found in the majority (75%) of all bound sequences (Hafner et al., 2010). Interestingly, a second CAUH motif was found in close proximity to the crosslinked site 30% of the time which supports the previously discussed ideas of repeats and multiple binding sites being necessary for IMP binding. While PAR-CLIP does appear to provide an accurate assessment of the cross-linked nucleotide, this specific study did not go into much detail about alternative IMP binding motifs, such as those longer than 4 base pairs or specific bipartite motifs. An additional caveat to this approach is that it was performed over-expressing the 3 IMP proteins. Over-expression of the IMPs can lead to protein mis-localization within the cell (Nielsen et al., 2003) and/or mis-sedimentation of target RNAs in a polysome gradient (Bell et al., 2013) so this is not ideal for genome-wide studies. However, this study was a much-needed first step to identifying the IMP-bound transcriptome and further independent validation and analysis of the target RNAs identified using PAR-CLIP will surely add to the details provided by Hafner and colleagues. Taken together, the currently available data suggest a significant structural complexity of IGF2BP-RNA binding combined with few enriched sequences containing AU-rich repeats and

multiple interaction sites within a single target RNA.

PPI-Partners in different RNPs

In addition to the sequences bound by the IMPs, a general knowledge of their protein-protein interaction (PPI) partners can also provide insight into their function. It has been well established that IMPs interact with binding partners in ribonucleoprotein complexes (RNPs) and that interactions can be both RNA dependent and independent. IMP proteins are named, in part, for their regulation of *IGF2* mRNA, which was also the first mRNA found to be regulated by LIN28 at the level of translation (Polesskaya et al., 2007). In particular, all three IMP RBPs co-immunoprecipitate with LIN28 in an RNA dependent manner in proliferating, but not differentiating, C2C12 muscle cells (Polesskaya et al., 2007). This is due, in part, to the fact that levels of IMP proteins decrease during muscle differentiation, but is also likely suggestive of a more general function for IMP proteins in highly proliferative cells.

A second paper analyzing IMP1 RNPs during an *in vitro* model of differentiation isolated IMP1 RNPs in differentiating P19 embryonal carcinoma cell neurons. HuD had previously been shown to bind both microtubules and the Tau RNA in differentiated P19 neurons, but the PPI partners were not identified. To test this RNPs containing recombinant HuD-GST RNPs were isolated and bound proteins and RNAs were determined by a GST-pull down (Atlas et al., 2004). IMP-1 was found to associate with HuD and G3BP-1 proteins in an RNA-dependent manner and also binds directly to tau mRNA in a location distinct from HuD and G3BP-1 (Atlas et al., 2004). These results were especially interesting considering the differences in temporal

expression of these interacting proteins during the neural differentiation process. IMP1 expression peaks in early differentiation and declined around the same time that HuD and Tau protein levels start increasing, which suggest there are additional factors controlling the temporal association and functions of these RBPs. Further studies will elucidate how levels of the proteins in this RNP are temporally controlled.

Immunoprecipitation followed by mass spectrometry identified binding partners of IMP1 in HEK293 cells (Jonson et al., 2007). In contrast to previous reports in the P19 neurons HEK293 RNP granules were described to be unique, distinct entities that are different from Staufen and FMRP RNP granules, P-bodies, and stress granules (SGs) both in size (300nM) and by segregation of IMP1 from markers such as G3BP1 following cellular fractionation (Atlas et al., 2004; Jonson et al., 2007; Nielsen et al., 2002). Consistently however protein components of HEK293 IMP1 RNPs include mostly other RNA binding proteins, including many members of the hnRNP family, YB1/major core protein, ILF2, ILF3, PABP1, PABP2, PABP4 as well as nucleolin, RNA helicase A, 40 S ribosomal proteins, and the nuclear cap-binding protein CBP80 (Table 1)(Jonson et al., 2007). The lack of translation initiation factors and 60 S ribosomal subunits indicates that bound mRNAs in these RNPs are likely untranslated (Jonson et al., 2007). RNA targets isolated from these RNPs included many components of the ubiquitin-dependent protein degradation pathway and other proteostatic processes providing more evidence that IMP1 RNPs may have a role in promoting localized translation/degradation of the associated RNAs at discrete loci within a cell.

In a more recent study in U2OS osteosarcoma cancer cells a similar approach was used to identify proteins in Flag-tagged IMP1 RNPs that mediate *MYC* RNA stability (Weidensdorfer et al., 2009). Flag-tagged IMP was immunoprecipitated, and enriched bands were analyzed by mass spectrometry. Similarly to the IMP1 RNP described in HEK293 cells, Weidensdorfer and colleagues pulled down IMP3 and YB1 as well, which was also confirmed by Butter et al., 2009. However, in contrast to the study in HEK293 cells, U2OS IMP1 RNPs *did* associate with STAUFEN1 and 2. Additionally, this report specifically performed the IP +/- RNAse and was able to show that with the exception of STAUFEN1 and DDX3X all of the IMP1 PPIs are RNA dependent. Interestingly, these studies highlight the complexity of unique IMP RNPs and that PPIs may be temporally and spatially specific in different cell types.

Last but not least is the discovery that IMP1 can both homo-dimerize and hetero-dimerize with IMP2 and IMP3 (Jonson et al., 2007; Nielsen et al., 2004; Nielsen et al., 2002; Git et al., 2002; Butter et al., 2009). This finding is stimulating because it provides a mechanism as to how IMP family members bind the same targets (Hafner et al., 2010) and may lead to redundancy in their overall functions within a cell. Interestingly, visualization of motile IMP1 RNP granule trafficking by tagged-IMP fluorescent proteins alone suggests that there are consistently multiple members to a single RNP otherwise the signal wouldn't be visible (Nielsen et al., 2002). Moreover, in an *in vitro* reaction this IMP1-IMP1 interaction is described to be both cooperative and sequential. Specifically, binding of the first IMP family member only provides an unstable interaction, where binding of a second IMP1 molecule to the

same RNA stabilizes the interaction within the RNP into a "locked" state that is more stable with a longer half life (Nielsen et al., 2004). This could be important for stability of the complex as it travels long distances in the cell or for selectivity. IMP family dimerization is also significant because this self-interaction is evolutionarily conserved; the IMP3 *Xenopus* orthologue Vg1RBP homodimerizes through its KH3 and 4 domains, and its association with the *Vg1* RNA and provides stability to the RNP (Git et al., 2002). A summary of IMP1 binding partners is provided in Table 1.

IMPs display an oncofetal expression pattern

During mammalian development

Analysis of IMP1-3 protein expression in early mammalian development was first analyzed by immunofluorescence imaging on sections from E12.5-E15.5 mouse embryos. Positive immunostaining was found at the basal layer of the developing epidermis of the skin, where IMP-1 was located at the basal plasma cell membrane, but not in the dermis suggesting an enrichment in the more primitive tissues (Nielsen et al., 1999). The IMPs were are also expressed in the developing epithelia of the lung and the intestine (Nielsen et al., 1999). IMPs were also present in the developing muscle tissue (Nielsen et al., 1999; Polesskaya et al., 2007). In addition to the mouse embryonic tissue, dot blots containing RNA from human fetal tissues were tested for IMP expression. These experiments revealed IMP-1, IMP-2, and IMP-3 mRNAs are expressed in human embryonic liver, lung, kidney, thymus, and placenta. Furthermore, IMP1-3 protein was expressed even earlier during development in E3.5 blastocysts

and in both male and female gonads (Hammer et al., 2005), cementing a role for these proteins in the earliest stages of mammalian development.

Functional evidence from an IMP1 deficient mouse strain suggested that IMP1 is required for normal embryonic development. IMP1 knockout mice are significantly smaller than their wild-type littermates, due to a general decrease in cellular proliferation (Hansen et al., 2004). IMP1 mutants also exhibit reduced survival, which was attributed to intestinal hypoplasia beginning in embryonic stages and becoming more apparent after birth. The most severe phenotype of IMP1 deficient mice was necrotic patches present in the intestines, which likely attributed to intestinal dysfunction (Hansen et al., 2004). Further investigation, focused on other endodermal tissues, included the intestines, liver and kidneys. Interestingly, there were few gene expression changes at the RNA level at E12.5 by microarray analysis when IMP RNA expression is reported to have peaked (Hansen et al., 2004; Nielsen et al., 1999); however, there were more changes (albeit few in general) in postnatal intestine, liver and kidney tissue samples that generally correlated with a mis-regulation of extracellular matrix (ECM) proteins (Hansen et al., 2004). Mis-regulation of the ECM and thus cell-cell interactions may have played a hand in the improper location of proliferating intestinal crypt cells further up the crypt rather than in the basal layer (Hansen et al., 2004).

Additionally, there was an overall deficiency in organ size, the IMP1 -/- mice organs were approximately14% smaller on average starting at day E17.5 and those that survived were approximately 45% smaller 1 week after birth which is likely due

to a decrease in proliferation and IGF2 translation (Hansen et al., 2004). Notably, there seemed to be neurological damage in the few surviving IMP1 deficient mice demonstrated by aggressive behavior, restlessness, and circular movements (Hansen et al., 2004). IMP family member expression in the early developing nervous systems of zebrafish, *Xenopus*, and *Drosophila* is also evolutionarily conserved (Mueller-Pillasch et al., 1999; Nielsen et al., 1999; Zhang et al., 1999; Nishino et al., 2013). A recent paper by Nishino et al. followed up on these findings and investigated the role of IMP1 specifically in mouse embryonic neural development.

A conserved role for IMP1 in neural development

Gary Bassell's lab and collaborators have put together a nice repertoire of papers specifically investigating the role of mRNA transport and local protein synthesis by ZBP1 (IMP1) in neurons of the CNS and PNS. In a model system of rat cultured hippocampal neurons, Tiruchinapalli et al., showed that ZBP1 was distributed in dendrites in the form of granules that spatially associated with *ACTB* and that movement of these granules responded quickly to increased synaptic activity with addition of KCl (Tiruchinapalli et al., 2003). Apparently, ZBP1 is actually required for *ACTB* localization to dendrites because loss of ZBP1 lead to mis-localization of *ACTB* (Eom et al., 2003). Furthermore, stimulation of ACTB dependent dendritic density by BNDF was impaired with anti-ZBP1 morpholinos (Eom et al., 2003). BDNF also plays a role in ZBP1 localized *ACTB* expression in axons. Huttelmaier et al., showed phosphorylation of ZBP1 by SRC family kinases allowed for *ACTB* release and

ultimately translation (Huttelmaier et al., 2005), however, it wasn't clear how spatial regulation of this phosphorylation site was physiologically relevant in a cellular system. Turns out, ZBP1 phosphorylation is involved in local translation of *ACTB* in growth cones and growth cone turning in response to BDNF and Netrin-1 signaling, but not developmental axonal outgrowth (Sasaki et al., 2010; Welshans et al., 2011). This function for ZBP1 to promote localized Actin synthesis during growth cone guidance is evolutionarily conserved all the way to *Xenopus* (Yao et al., 2006). The above studies showed that ZBP1-dependent localized actin translation is crucial for axon guidance, but the mechanism of how ZBP1-ACTIN RNPs are actually transported along the axon to growth cone was still unknown.

Axonal transport of diverse cargos is required for normal neuronal function, thus impeding full function of molecular motors often leads to neurodegenerative diseases such as ALS, Alzheimer's, and Parkinson's. Myosin Va (MyoVa) is an abundant processive myosin that transports varied cargoes and is important for neural development, thus it was a good candidate to test for regulation of ZBP1 RNP transport (Lewis et al., 2009; Tamada et al., 2010). Nalavadi et al., showed that MyoVa associates with ZBP1 and plays an inhibitory role causing accumulation of ZBP1 in axons of cultured hippocampal neurons (Nalavadi et al., 2012). MyoVa inhibition resulted in increased transport dynamics and reversal of orientation of ZBP1 particles in neurons (Nalavadi et al., 2012). Additionally, IMP1 was shown to colocalize with SMN1 in E13.5 primary motor neurons and in SMN1 deficient cells localization of IMP1 was specifically depleted in the axon and not in the cell body

(Fallini et al., 2013). This suggests that axonal localization of IMP1 may also be dependent on interacting partners within RNP granules. These studies provided important new insight into how IMP family mRNA-binding proteins are transported along axons to neuronal growth cones and provided a mechanistic function for IMP1 during neural development.

To add to the previously discussed studies that found significant roles for IMP family members in localization and trafficking, a recent publication suggests a function for IMP1 in specifying neural cell fate. Nishino et al. found that in IMP1 deficient mice there is pre-mature differentiation in the dorsal telencephalon (as labeled by loss of Pax6+ cells and gain of Tuj1+ cells), which is likely due to premature cell cycle exit from loss of CyclinD (Nishino et al., 2013). Moreover, this phenotypic effect takes place specifically at E14.5-E18.5, while not at E12.5 which emphasizes the spatial and temporal function of IMP1 during embryonic development. It has been well characterized that WNT signaling plays a role in the maintenance of Pax6+ progenitor cell populations and indeed, exogenous WNT3a added to explant medium increased the expression of *Imp1 in vitro* (Nishino et al., 2013). To test whether this same effect was present *in vivo* APC deficient mice and CTNNB1 mutant mice were examined and both induced and repressed (respectively) IMP1 expression *in vivo* as one would expect if this was the case (Nishino et al., 2013). Finally they showed that IMP1 indeed was a target of the *let-7* family of miRNAs *in vivo,* and that the effect of *let-7* was rescued by over-expression of Lin28a. All of these phenotypes

are relevant to the role of IMP1 during embryonic development and/or tumorigenesis and will be discussed later in this report.

Cancer

IMP biology has a fundamental impact on human health as these RBPs are upregulated in many different types of common cancers and their expression is often correlated with poor patient prognosis (Ross et al., 2001);(Gu et al., 2004; Hammer et al., 2005); (Dimitriadis et al., 2007; Wachter et al., 2012). It remains a mystery as to how IMP RBPs are upregulated during tumorigenesis, however one potential mechanism may be the general downregulation of microRNAs (miRNAs). MiRNAs are a class of ~21-23 nucleotide non-coding RNAs that act as post-transcriptional regulators of 3′ UTRs. Targeting of miRNAs to 3′ UTRs results in decreased gene expression likely by an increase in RNA degradation, but also perhaps a block in translation (Guo et al., 2010). These ncRNAs have critical roles in diverse biological processes that encompass development, proliferation, apoptosis, stress response, and fat metabolism.

The IMP family mRNAs has been shown to be targeted for degradation by the *let-7* family of miRNAs (Boyerinas et al., 2008). The *let-7* family is highly evolutionarily conserved and all members appear to be up-regulated towards the end of embryonic development (Abbott et al., 2005; Schulman et al., 2005). *Let-7* expression is maintained in the adult, but can be down-regulated during the early stages of tumor development, suggesting that *let-7* regulated oncofetal genes (LOGs)

may become re-expressed in cancer cells (Boyerinas et al., 2008; Yu et al., 2007). This is exactly the case for regulation of *IMP1*. Human *IMP1* was originally identified to be an *in vitro* target of *let-7* based on six *let-7* seed sequences present in the 3'UTR, five of which are conserved in all mammalian species. Furthermore, addition of exogenous *let-7* to two cancer cell lines, k562 and HEPG2, resulted in the downregulation of IMP1 while down-regulation of *let-7* due to addition of antisense oligonucleotides, resulted in upregulation of IMP1 (Boyerinas et al., 2008). These results signify the importance of post-transcriptional regulation of IMP1 by *let-7* during development and provide a potential mechanism for how IMP RBPs become upregulated in human cancers.

An alternative mechanism by which IMP could be upregulated during tumorigenesis could be by a direct increase in *IMP1* transcription. There is some evidence in 293T cells that β-catenin directly upregulates IMP1 transcription by LEF/TCF4 binding (Noubissi et al., 2005) and that MYC can upregulate CRD-BP expression as well in a positive feedback loop (Noubissi et al., 2010). Remarkably, both of these genes are IMP1 binding targets and therefore complete a feed-forward loop of IMP1/target expression. Consequences of this upregulation are evident in a mouse mammary tumor model where IMP1 is over-expressed in mammary epithelial cells of pregnant and lactating females under the whey-acidic-protein (WAP) promoter (Tessier et al., 2004). IMP1 OE mice form tumors capable of metastasis at 95% efficiency, whereas control mice do not ever generate tumors, providing evidence that IMP1 may act as an oncogene (Tessier et al., 2004). Interestingly, β-catenin has also

been shown to activate IMP1 transcription in human breast cancer cell lines, and IMP1 promotes the stability of the β-catenin RNA (Gu et al., 2008). These data, along with the previous report suggesting WNT signaling promotes IMP1 expression during murine neural development (Nishino et al., 2013), suggest that the IMP-β-catenin interaction network may be conserved across tissue types and in various human cancers. Further investigation of these feed-forward loops will solidify how IMP1 is upregulated and promotes tumorigenesis.

While the cause of IMP upregulation during tumorigenesis is only beginning to be investigated, the end result has been extensively studied. Consistently across multiple tumor types, IMP family member expression is correlated with poor patient prognosis across multiple tumor types (See Bell, 2013 for review). An enrichment of IMP1 expression in breast, lung, brain, ovarian, testicular and colon cancers was associated with late stage tumors and lower patient survival (Ioannidis et al., 2003; Ioannidis et al., 2004; Gu et al., 2004; Hammer et al., 2005; Dimitriadis et al., 2007). Interestingly, the colon cancer study found that IMP1 expression was enriched in samples that also expressed another oncofetal RBP, MUSHASHI1 (Dimitriadis et al., 2007). This suggests that coordinated networks of RBPs may be hijacked during cellular transformation. IMP3 expression has also been extensively correlated with many different types of cancers and has been clinically associated with poor prognosis in hepatocellular, gastrointestinal, and prostate cancers (Hu et al., 2014; Kim et al., 2014; Szarvas et al., 2014). While IMP2 has not been as extensively correlated with

tumors as the other 2 mammalian family members, it also seems to be associated with poor prognosis (Alajez et al., 2012).

Prominent histological expression of these RBPs in so many different tumor types led to the idea that they could be targeted for destruction as a cancer therapy. Disruption of the *MYC*-IMP1 interaction in k562 cells using antisense oligonucleotides (ASOs) targeted against the interaction site between IMP1 and the *MYC* CRD resulted in a 70% decrease in k562 cell proliferation in a dose dependent manner (Coulis et al., 2000). A caveat to this study, however, was that a second control ASO that targeted a different, unique region of the *MYC* RNA was also able to have a 50% reduction in k562 cell proliferation, suggesting that there may be other regions of target RNAs that lead to instability and not just an interaction with a RBP.

Recently King et al. used ASOs to target the IMP1-*CD44* interaction *in vitro* and in Hela cells (King et al., 2014). These ASOs targeted the *CD44* 3′ UTR at a site known to interact with IMP1. Unfortunately, however, these ASOs worked well in *in vitro* assays, but didn't have any functional affect once tested in Hela cells. ASO therapy is an area of high interest as there are currently clinical trials being performed to test ASO therapy against *SOD1* mutations in ALS patients (Miller, 2013).

Conclusions

Here I have reviewed what is known about the IMP family of RNA binding proteins in stem cells, development and during tumorigenesis. In addition, I have outlined how work on the IMP family has provided a foundation for studies of other

RNA binding proteins in human embryonic stem cells. Evolutionary conservation of expression and function regulating RNA translation, localization, and stability during embryogenesis of many different species ranging from drosophila to humans is indicative of the importance of these proteins. Multiple interaction partners were described to have an effect by co-regulating the same RNAs in IMP-RNPs or by aiding in localization of polarized cells.

The knowledge of only a handful of IMP mRNA targets greatly limits our ability to define which signaling pathways and mechanisms contribute to its regulation. The first published study to identify transcriptome-wide targets of IMP using an over-expression system revealed that it interacts with hundreds or even thousands of protein-coding transcripts (Hafner et al., 2010). These foundations motivated us to carefully define endogenous IMP binding sites throughout the human transcriptome in order to shed light on its network of direct and indirect mechanisms of regulation during development and tumorigenesis.

Methods

Evolutionary tree diagram

To design the evolutionary tree diagram I uploaded the various Uniprot identifiers for each of the different organisms, including: Human 1 IF2B1_HUMAN, 2 IF2B2_HUMAN, 3 IF2B3_HUMAN, Xenopus IF23A_XENLA, chick IF2B1_CHICK, Mouse IF2B1_mouse, IF2B2_mouse, IF2B3_mouse, IF2B1 drosophila, zebrafish IF2B3_DANRE to the Uniprot.org alignment tool and edited it in Adobe Illustrator.

Figures

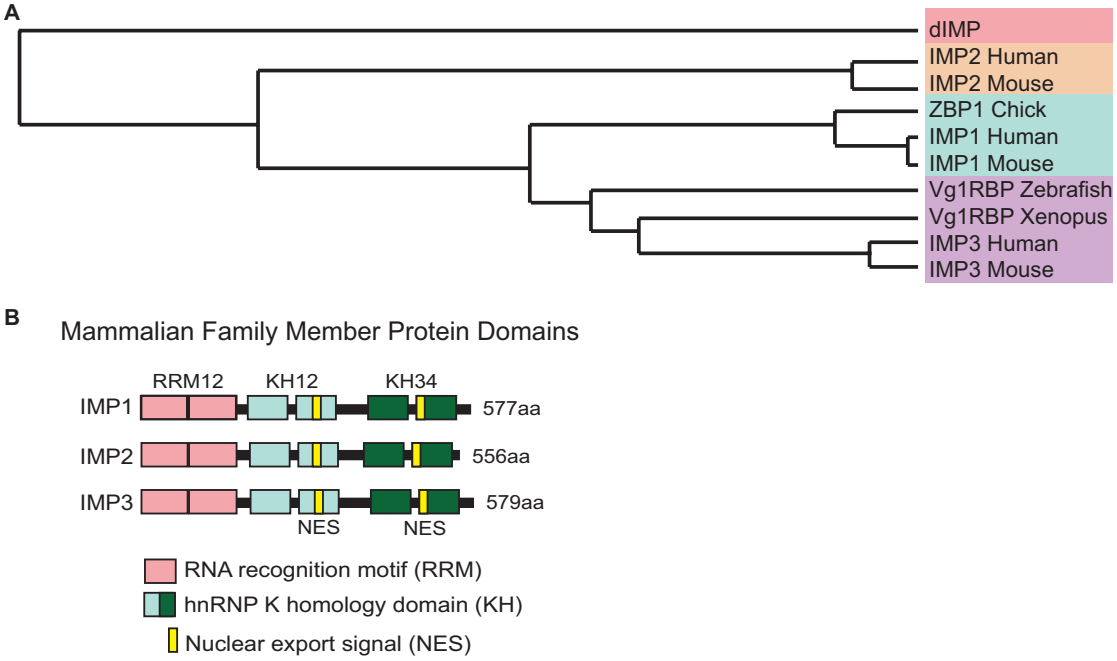

Figure 1. Evolutionary tree diagram of IMP RBP family members
(A) Evolutionary tree diagram showing divergence of IMP RBP family proteins. (B) Schematic of RNA binding domains located in the mammalian IMP family members.

Tables

TABLE 1. REVIEW OF IMP1 PROTEIN-PROTEIN INTERACTIONS.

Over-expressed?	Binding partner	RNA dependent?	Model/Cell Line	Reference
IMP1-FLAG tagged	ALY/REF	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	CBP80	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	DDX3X	N	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	DHX9	Y	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	EIF4AIII	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	ELAVL1	Y	U2Os cells	Weidensdorfer, 2009
G3BP- yes	G3BP	Y	P19 neuronal cells	Atlas, 2004
IMP1-FLAG tagged	HNRNPA1	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	HNRNPA2B1	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	HNRNPD	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	HNRNPE2	Y	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	HNRNPL	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	HNRNPQ	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	HNRNPR	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	HNRNPU	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	HNRNPU	Y	U2Os cells	Weidensdorfer, 2009
HuD- yes	HuD	Y	P19 Neuronal cells	Atlas, 2004
IMP1-FLAG tagged	ILF2	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	ILF3	Y	U2Os cells	Weidensdorfer, 2009
No	IMP1	Y	In vitro	Nielsen, 2004
No	IMP2	Y	In vitro	Nielsen, 2004
No	IMP3	Y	In vitro	Nielsen, 2004
LIN28- Flag-tagged	LIN28	Y	C2C12 myoblasts	Polesskaya, 2007
IMP1-FLAG tagged	PABP1	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	PABP2	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	PABP4	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	PABPC1	Y	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	PTBP2	Y	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	RPS6	NA	HEK293	Jonson, 2007
No	SMN1	NA	In vitro	Fallini, 2014
IMP1-FLAG tagged	STAU1	N	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	STAU2	Y	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	SYNCRIP	Y	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	UPF3B	NA	HEK293	Jonson, 2007
IMP1-FLAG tagged	YBX1	Y	U2Os cells	Weidensdorfer, 2009
IMP1-FLAG tagged	Y14	NA	HEK293	Jonson, 2007

CHAPTER 2 – IGF2BP/IMP RNA BINDING PROTEINS REGULATE A RNA NETWORK IN HUMAN PLURIPOTENT STEM CELLS TO MAINTAIN CELL SURVIVAL

Introduction

Human embryonic stem cells (hESCs) provide an invaluable model system to address mechanisms of early human development due to their ability to self-renew and differentiate into the majority of cell types in the mammalian embryo. While transcriptional networks controlled by DNA-binding transcription factors have been extensively studied in hESCs (Boyer et al., 2005; Chia et al., 2010), there has been a dearth of research investigating post-transcriptional RNA networks mediated by RNA binding proteins (RBPs). RBPs associate with RNAs to control diverse aspects of RNA metabolism such as splicing, polyadenylation, editing, stability, translation and localization. A limited number of studies have begun to reveal RNA regulatory networks controlled by RBPs, such as RBFOX2, LIN28A and MBNL proteins in human pluripotent stem cells (Han et al., 2013; Wilbert et al., 2012; Yeo et al., 2009). These studies demonstrate that RBPs play key roles in influencing the pluripotent state and emphasize the importance of exploring the RNA targets and functions of other RBPs in human pluripotent stem cells (hPSC).

The IGF2 mRNA binding proteins (IMP/IGF2BP) are a family of RBPs that are highly expressed in hPSC. IMP/IGF2BP family members share high homology and are conserved from insects to mammals (Hansen et al., 2004; Nielsen et al., 1999). The names used for the IMP homologues vary across organisms and include the

VICKZ family of zipcode-binding proteins (Vg1RBP/Vera, Xenopus), cMyc coding region determinant binding protein (CRD-BP, mouse), and Zipcode-binding protein (ZBP1, chicken) (Deshler et al., 1998; Havin et al., 1998; Leeds et al., 1997). Humans and mice have three IMPs (IMP1-3/IGF2BP1-3) that are expressed broadly during early development. Expression decreases in most tissues post-natally, with the exception of sustained expression in the germ line in adults (Hammer et al., 2005 1999; Hansen et al., 2004). Interestingly, IMP RBPs are targeted for degradation by the *let-7* family of miRNAs and often become re-expressed in many tumors where *let-7* expression is reduced (Boyerinas et al., 2008). Additionally, the IMPs are present in many different types of cancer including lung, liver, breast, and colon, and upregulation is tightly correlated with poor patient prognosis (Ross et al., 2001; Gu, et al., 2004; Hammer et al., 2005; Dimitriadis et al., 2007; Wachter et al., 2012). Furthermore, overexpression of IMP1 in adult murine breast tissue led to tumors, suggesting that IMP proteins can act as oncogenes (Tessier et al., 2004).

Molecular mechanisms of how IMP proteins bind and regulate their target RNAs have been studied predominantly in vitro. All IMP family members have six RNA-binding domains: two RNA recognition motifs (RRM) and two sets of hnRNP K homology domains (KH) (Figure 2A). Molecules of IMP1 protein bind RNA cooperatively and sequentially, dimerizing to form a stable complex with bound RNA (Nielsen et al., 2004). Both IMP2 and IMP3 are able to heterodimerize with the KH1-4 domains of IMP1 (Nielsen et al., 2004). Two nuclear export signals are located between the second and fourth KH domains that allow for IMP translocation from the

nucleus into the cytoplasm (Figure 2A) (Nielsen et al., 2003). All four KH domains contribute to RNA binding and are important for localization of IMP (Nielsen et al., 2002).

Several direct RNA targets of IMP proteins have been identified, and most correspond to genes that are important for development and differentiation. For example, IMP1 is required in axons of developing neurons for localized actin polymerization and growth cone turning in response to BDNF, as well as dendritic outgrowth (Perycz et al., 2011; Sasaki et al., 2010). IMP1 also controls the localization and translation of neuron-specific Tau mRNA, in a differentiation-dependent manner, during in vitro differentiation of murine P19 embryonal carcinoma cells to neurons (Atlas et al., 2007). Additionally, IMP proteins control RNA transcript stability, as is the case of MYC regulation in mesenchymal stem cells (Bernstein et al., 1992; Mahaira et al., 2014). The namesake target of the mammalian family, IGF2, is primarily regulated at the level of translation. Interestingly however, depending on the cellular context, IMP1 can either promote or repress translation of IGF2 in either NIH 3T3 cells or IMP1-/- MEFs, respectively (Nielsen et al., 1999; Dai et al., 2013). These few target examples illustrate the various ways IMPs can control aspects of RNA metabolism. Consistent with a broad expression pattern during development, IMP1 knockout mice have growth defects and a low survival rate, demonstrating that IMP1 is necessary for proper embryogenesis (Hansen et al., 2004). Additionally, IMP1 has been shown recently to be required for proper neural development in vivo as IMP1

depletion in IMP1-/- embryonic mouse brains leads to premature differentiation in the dorsal telencephalon (Nishino et al., 2013).

Although these studies in cell lines and model organisms have provided insight into IMP regulation of a small number of RNAs, our understanding of how the IMP-RNA target orchestra is conducted transcriptome-wide is incomplete. To provide insight into the role of IMP proteins during human development, UV crosslinking and immunoprecipitation, followed by high-throughput sequencing (CLIP-seq, also known as HITS-CLIP), was used to identify the endogenous targets of IMP1 and IMP2 in human PSCs. These data were combined with functional studies to link genes regulated by IMPs to processes regulated by these proteins. Here we demonstrate that loss of IMP1 leads to decreased cell survival and adhesion, and our data suggest that newly identified IMP1 targets BCL2 and ITGB5, respectively, contribute to these phenotypes. The survival and adhesion functions uncovered in hESCs provide insight into the roles of IMP proteins during development. In addition, our studies reveal how dysregulation of IMP protein expression could contribute to cellular transformation.

Results

Transcriptome-wide discovery of IMP1 and IMP2 RNA targets in human embryonic stem cells

Human IMPs 1-3 are highly expressed in PSCs (Figures 2B and 2C), and immunohistochemical staining demonstrated predominant cytoplasmic localization (Figures 2D and 3A). IMP1 and IMP3 protein expression was restricted to undifferentiated PSCs, while IMP2 expression was expressed in both induced

pluripotent stem cells (iPSCs) and parent fibroblasts (Figure 2C). To identify novel RNA targets that could reveal molecular pathways regulated by IMP proteins in PSCs and during development, CLIP-seq was performed in hESCs (H9 and HUES6) using antibodies that specifically recognize IMP1 or IMP2 (Figures 3B and 3C). Clusters of reads that passed both transcript and transcriptome-wide cutoffs for statistical significance ($p < 0.001$) were designated binding sites for each protein (Polymenidou et al., 2011; Zisoulis et al., 2010). We identified 23,985 binding sites for IMP1 in 7,371 annotated genes and 6,170 binding sites for IMP2 in 2,647 genes (Table 2). Surprisingly, despite significant differences in IMP1 and IMP2 expression across tissues, 79% of IMP2 target genes (2,100 of 2,647) were also IMP1 substrates (Figure 2E; $p < 0.0001$, by hypergeometric test).

To evaluate the cell-type specificity of IMP1 and IMP2 RNA targets, we compared our target RNAs in hESCs with those identified previously in HEK293 cells using PAR-CLIP with antibodies that recognize epitope tagged IMP proteins (Hafner et al., 2010). Our analysis revealed that an overwhelming majority (86%) of IMP2 substrates in hESCs are also bound in HEK293 cells (Figure 2F) and more than half of IMP1 targets in hESCs and HEK293 cells overlap (Figure 2G). Therefore, it appears that recruitment of IMP 1 or 2 to RNA substrates is largely independent of cell-type, even when ectopically expressed in somatic cells. Interestingly, we also found that RNA targets of IMP1 and IMP2 overlap significantly with those of human LIN28A ($p < 0.0001$, by Chi-square test), a RBP that is also highly expressed in PSCs (Wilbert et al., 2012)(Figure 2H). Similar to our previous report that LIN28A controls the

expression of splicing factors, targets of IMP1 in HEK293 and H9 hESCs are also enriched for factors that function in RNA processing (p = 2.7E-20, modified Fisher Exact test with Benjamini correction) and mRNA metabolism (p = 1.7E-19)(Table 3), as are common targets of IMP2 in these cells (RNA processing p = 2.7E-20; mRNA metabolic process p=1.7E-19), suggesting a possible overlap in function of LIN28A and IMP proteins in PSCs.

IMP1 and IMP2 binding is enriched at the 3' UTR of protein-coding genes

To evaluate the specificity with which IMP proteins interact with their direct RNA substrates, we examined the position of significant clusters of reads within IMP bound protein-coding transcripts. Similar to observations in HEK293 cells (Hafner et al., 2010), we found a striking enrichment of IMP1 and IMP2 binding within 3' untranslated regions (3' UTRs) of RNAs corresponding to protein-coding genes, compared with an expected pre-mRNA background (Figure 4A). In fact IMP binding was more enriched within 3' UTRs than LIN28-RNA interactions in hESCs (Wilbert et al., 2012), which exhibit higher coding exon preferences (Figure 4A), testament to the specificity of IMP binding. Within a 250 nucleotide window, we find that in 20% of pair-wise comparisons for all clusters, IMP2 is found within 10 bases of an IMP1 cluster, compared to 6% of randomly located clusters in the same regions (Figure 4B), a greater than 3-fold enrichment, supporting a model consistent with cooperative interactions between IMP1 and IMP2 with RNA substrates. Within only 3' UTR clusters, we find a statistically significant overlap of binding sites for IMP1 and IMP2

(~30%, compared to 3% expected with a control set of clusters; Z-score of 218; Figure 4C). Our analysis supports previous data demonstrating that IMP1 and IMP2 can dimerize and that IMP2 co-localizes and co-immunoprecipitates with tagged IMP1 in HEK293 cells (Jonson et al., 2007).

To investigate cooperative binding between IMPs 1 and 2 in more detail, we examined the beta-actin 3' UTR, which harbors the well-characterized IMP binding region, the "zipcode binding sequence" (Figure 4D)(Ross et al., 1997). We found increased densities of CLIP reads for IMP1 and IMP2 10-20 nucleotides 5' and overlapping the zipcode binding sequence in hESCs (Figure 4D), which agree with the clusters identified in HEK293 cells (Hafner et al., 2010) and correspond remarkably well to reads from our individual nucleotide-resolution CLIP (iCLIP) in K562 chronic myelogenous leukemia cells (Figures 4D, 3D and 3E). The iCLIP dataset provides additional validation of endogenous IMP targets using an independent system and cell type, thus strengthening our ability to report high-confidence target interactions. The different CLIP approaches used in multiple cell-lines also identified additional IMP binding sites downstream of the zipcode binding sequence (Figure 4D).

Using the de novo motif finding software HOMER, we calculated the most statistically significant overrepresented motifs found within our IMP1 and IMP2 clusters, relative to randomly located clusters in the appropriate genic backgrounds as control (Figures 4E, 4F, 5A and 5B). We identified consensus motifs 'GGACUN' ($p = 1E-76$) and 'CUGUAG' ($p = 1E-76$) within IMP1 binding sites and 'CUGUC' ($p = 1E-48$) and 'AGAAC' ($p = 1E-22$) within IMP2 binding sites. A search for

significantly enriched sequences also identified hexamers that contain 'ACUG' and stretches of 'AC' di-nucleotides in IMP1 clusters (Figure 5C). In addition, 'AAUAAA', which resembles the poly-A signal, was enriched around IMP2 sites (Figure 5D), consistent with the almost exclusive localization of IMP2 binding events in the 3' UTR (Figure 4A). Our CLIP-seq studies support previous structural evidence from Chao and colleagues for a bi-partite IMP1 motif with the sequence CGGACUG between 'AC'-rich motifs on either side (Chao et al., 2010), with our top consensus motif encompassed in the underlined portion. Within our experimentally determined IMP1 clusters, 81% of GGACU hexamers co-occurred with ACAC or CACA motifs within 100bp on either side, slight but significantly more than in matched control clusters ($p < 0.00001$; Z-score = 7.6). Thus, our results provide consensus IMP motifs and evidence that previous observation of an IMP bi-partite may apply to other RNA targets across the transcriptome.

Loss of IMP1 in hESCs leads to decreased cell survival and adhesion

To gain insight into roles for IMP1 in human development, IMP1 was depleted in hESCs using lentiviral transduction of short-hairpin RNAs (shRNA) specifically targeted against IMP1 (hereafter referred to as IMP knock-down, IMPKD, cells) (Figures 6A, 6B, 7A and 7B). Transduction with lentivirus carrying a non-targeting shRNA was used as a control. Knockdown of IMP1 did not affect the abundance of IMP2 or IMP3 proteins in hESCs, nor did depletion of IMP2 or IMP3 affect IMP1 levels (Figure 7A). Upon depletion of IMP1 in PSCs, we noticed a drastic reduction in

colony size coupled with irregularly shaped borders (Figures 6C and 7C). In addition, when induced to differentiate into embryoid bodies, EBs derived from IMPKD cells were smaller than those derived from controls (Figure 6D).

We performed a confluency assay with control and IMPKD cells over the course of 14 days, which confirmed a significant decrease in colony expansion ($p < 0.01$)(Figure 6E). In order to determine whether depletion of IMP1 led to a decrease in proliferation, which could explain a decrease in colony size, cell cycle profiling was conducted. IMPKD cells exhibited a significant, but small decrease in the S phase population ($p < 0.01$), along with an increase in the number of cells in G2 ($p < 0.05$) (Figure 6F). Therefore, IMP1 depletion had a modest effect on cell cycle distribution in hESCs, in contrast to previous reports that described a predominant role for IMP1 in proliferation (Ioannidis et al., 2005).

To determine whether lower levels of IMP1 in hESCs induced an increase in cell death, we measured the expression of cleaved caspase 3 by Western blot analysis. Indeed, a significant increase in levels of cleaved caspase 3 was observed upon reduction of IMP1 but not in control treated hESCs (Figures 6G and 7E). An increase in apoptosis of IMPKD cells was confirmed by Fluorescence activated cell sorting (FACS) analysis which detected an increase in Annexin V-positive IMPKD cells compared to controls ($p < 0.05$)(Figure 6H); therefore, these results strongly suggest that loss of IMP1 leads to an increase in cell death in hESCs. In addition to a decrease in the size of IMP1 KD hESC colonies and EBs, we observed that IMPKD cells exhibited a decreased plating efficiency, when compared to control cells, independent

of the passage number. To determine whether loss of IMP1 affected adhesion of hESCs, a quantitative adherence assay was performed following IMP1 depletion. Indeed, hESCs with decreased IMP1 expression did not adhere as well as control cells even within the first hour after plating ($p < 0.05$) (Figures 6I, 6J, and 7E).

IMP1 mediates expression of protein coding and long non-coding RNAs in hESCs

To begin to identify direct and changing targets of IMP1 that could be important for cell survival and adhesion in hPSCs, RNA libraries were generated and sequenced from IMPKD and control cells to identify RNA targets that are affected by loss of IMP1. Of the 28,255 expressed genes analyzed by RNA-seq analyses, 2,938 RNAs changed differentially upon IMP1 depletion ($p < 0.01$) (Table 4). Of those RNAs corresponding to protein-coding genes that change upon IMP1 depletion, 612 decreased and 850 increased significantly by at least 1.5 fold. Of these, 43% (264) up-regulated RNAs and 30% (257) down-regulated RNAs were IMP1 targets; 20% of IMP1 targets remained unchanged (Figure 8A). Therefore, a statistically significant proportion of differentially expressed transcripts were targets of IMP1 ($p < 0.0001$; up-regulated genes, chi-square=20; down-regulated genes, chi-square=128). Several RNAs that change at the mRNA level in response to IMP1 depletion may contribute to the cellular phenotypes that we observed. For example, the slight increase of IMP1 KD cells in G2 could be due to an upregulation of the IMP1 RNA target and cell cycle regulator p21 (CDKN1A) (3.4 fold upregulated, $p = 2.26E-16$). An increase in p21 has been associated previously with a loss of IMP1 in MCF7 breast cancer cells and

mesenchymal stem cells, suggesting that this may be an IMP1 target that is conserved across different cell types (Hafner et al., 2010; Ioannidis et al., 2005; Mahaira et al., 2014). IMP proteins are known to control gene expression through a number of mechanisms; therefore, the IMP1 protein-coding targets whose steady-state mRNA levels remain unchanged could be affected at the level of mRNA localization or translation.

Of the IMP1 bound genes 6,777 were protein coding and, an additional 519 were annotated as non-coding RNAs (ncRNAs) (Figure 8B and Table 5). Although several IMP1-ncRNA interactions have been reported, such as with the Y3 RNA (Sim et al., 2012), H19 long ncRNA (lncRNA) (Runge et al., 2000), and HULC lncRNA (Hammerle et al., 2013), the global population of ncRNAs bound by IMP1 have not been investigated. The 519 ncRNAs bound by IMP1 included 172 lncRNAs, 111 pseudogenes and 87 antisense transcripts (Figure 8C). Of these, 11% (19) of lncRNAs, 3% (3 of 111) of pseudogenes and 7% (6 of 87) of antisense transcripts changed in expression level upon depletion of IMP1 (Table 5). Therefore, IMP1-bound lncRNAs were more sensitive to IMP1 depletion, when compared to other types of ncRNAs (Tables 6 and 7). One lncRNA directly bound by IMP1 was the Differentiation Antagonizing Non-Protein Coding RNA (DANCR), also referred to as anti-differentiation ncRNA (ANCR) (NR_024031) (Figures 8D, 8E and Table 7). This lncRNA was previously shown to be downregulated during differentiation and is required to sustain progenitor cells in the basal layer of the epidermis (Kretz et al., 2012). Enriched densities of IMP1 and IMP2 CLIP and iCLIP reads and PAR-CLIP

clusters were identified within the 3' exon of the DANCR gene in hESCs, K562 and HEK293 cells, supporting a conserved interaction between IMP 1 and 2 proteins and DANCR lncRNA (Figure 8E). We confirmed this interaction in HUES6 hESCs using RNA immunoprecipitation (RIP) of IMP1, followed by qRT-PCR, to detect IMP bound DANCR and ACTB RNA (Figure 8D). A control lncRNA, HMNT, which does not contain IMP-binding sites identified by our CLIP-seq data, was not detected bound to IMP1 (Figure 8D). Furthermore, DANCR decreased significantly upon depletion of IMP1 in hESCs as confirmed by both RNA-seq ($p < 0.0001$) and qRT-PCR ($p < 0.01$)(Figure 8F). Our results demonstrate that IMP1 and IMP2 bind many non-coding RNAs and that IMP1 can control the levels of specific lncRNAs in hESCs.

IMP1 restricts the expression of pluripotency factors

Based on the markedly high expression of IMP1 in hESCs (Figure 1C), we hypothesized that downregulation of IMP1 in hESCs may result in a loss of pluripotency and initiation of differentiation. IMP1KD cells maintained high expression of the cell surface pluripotency marker stage-specific embryonic antigen 4 (SSEA4), and levels were indistinguishable from control cells treated with a non-targeting shRNA (Figures 9A and 9B). Surprisingly, increased expression of the OCT4, SOX2, and NANOG (OSN) proteins was observed upon depletion of IMP1 (Figure 9C). The increase in OSN was also observed in induced pluripotent stem cells from BJ fibroblasts also depleted for IMP1, indicating that the increase is not specific to hESCs (Figure 9E). Our CLIP-seq data revealed IMP1 binding sites in the OCT4

and SOX2 mRNAs, but not NANOG, suggesting both direct and indirect modes of regulation, respectively. In contrast to OCT4 and NANOG, however, SOX2 protein levels increased despite insignificant changes in the mRNA (Figures 9D and 9F). Within the SOX2 3' UTR we found the IMP1 binding motif 'GGACU' favorably positioned between two ACA rich sites (Figure 9G). In addition, an IMP2 binding site was located proximal to the polyadenylation signal, AAUAAA, downstream from the IMP1 site. These sites were also identified when tagged IMP proteins were overexpressed in HEK293 cells (Hafner et al., 2010). Independent validation of binding by IMP1 using native RNA immunoprecipitation (RIP) followed by RT-PCR confirmed binding of SOX2 mRNA by IMP1 but not IgG (Figure 9H). As a further indication of specificity of IMP binding sites, the clusters with the highest density of IMP1 and IMP2 CLIP reads were distinct from binding sites of LIN28, which also targeted SOX2 mRNA (Wilbert et al., 2012)(Figure 9G).

Upregulation of NANOG, OCT4, and SOX2 in human PSCs has been linked to enhanced differentiation towards specific embryonic lineages (Wang et al., 2012). For example, upregulation of OCT4 leads to mesendoderm, NANOG to definitive endoderm, and SOX2 to neuroectoderm (Wang et al., 2012). To determine whether levels of IMP1, and consequently increased OSN, influences the differentiation of IMP1KD cells into any of the three embryonic germ layers, we performed an undirected EB differentiation assay over a two week period and quantified expression of lineage markers by qRT-PCR. Downregulation of NANOG, OCT4, and SOX2 was observed in IMP1KD hESCs upon differentiation, as well as an upregulation of

lineage markers from each of the three embryonic germ layers, indicating that IMP1KD hESCs are able to differentiate efficiently (Figures 10B-E). Interestingly, however, there was a significant upregulation in lineage markers (PAX6, TH, EOMES, NODAL, SOX17, and BRACHYURY) detected in each of the three embryonic germ layers at various time points during differentiation, indicating a potential temporal role for IMP1 in influencing early cell fate decisions. This premature upregulation of differentiation markers has also been reported with loss of IMP1 in the developing mouse brain in vivo (Nishino et al., 2013). These results, combined with our findings that IMP1 regulates proper levels of OSN, suggest that a primary function of IMP1 is to maintain the critical balance among OSN that promotes the pluripotent state and indicate that a perturbation of this equilibrium could disrupt lineage decisions once differentiation is induced.

IMP1 controls stability of integrin mRNA

Gene ontology analysis of hESC-specific IMP1 targets revealed a significant enrichment for RNAs involved in cell-cell and cell-extracellular matrix (ECM) adhesion (Huang da et al., 2009a, b). Furthermore, of the RNAs that were most significantly decreased upon IMP1 knockdown in hESCs, those that were IMP1 CLIP-seq targets were enriched for the biological function of cell-cell adhesion ($p < 0.001$)(Figure 12). Although this small set of genes (n=30) was not sufficient to pass correction for genome-wide significance, this was indication of a direct role for IMP1 in controlling hESC adhesion and motility. To determine whether cell-cell interactions

were affected in IMPKD cells, we first performed immunofluorescence analysis to assess the level and localization of cell adhesion and cytoskeletal proteins. CTNNB1 (β-catenin) and CDH1 (E-cadherin) have been previously described as IMP1 targets regulated at the level of RNA stability (Gu et al., 2008). Although significant IMP1 and IMP2 binding sites were identified in the 3' UTR of CTNNB1 in our CLIP-seq dataset, we not did observe changes at the RNA level in IMP1KD cells (Figure 12E). However, disorganization of CTNNB1 localization was observed in IMPKD hESCs (Figure 12C). We also observed a general disorganization of the actin cytoskeleton, which is likely due to lack of ACTB localization by IMP1 (Figures 12B and 12D). Therefore, our results point to maintaining cytoskeletal integrity as an important function of IMP1 in hESCs.

To determine the molecular mechanism by which IMP1 promotes adhesion in hESCs, we hypothesized that RNAs coding for multiple integrins, which are key for adhesion, would be targeted by IMP1. Indeed, five integrins were significantly downregulated ($p < 0.05$) at the mRNA level, including integrins a2, aE, aV, B1 and B5. (Figure 11A). ITGB5, in particular, had increased read density in the 3' UTR, along with conserved binding sites in hESCs and HEK293 cells (Figure 11B). The interaction between IMP1 and ITGB5 was confirmed in hESCs using RIP, followed by qRT-PCR, to detect IMP1- associated ITGB5; ACTB was used as a positive control for a strong binding target. (Figure 11C). Down-regulation of ITGB1 and ITGB5, which have important roles in regulating stem cell maintenance, (Braam et al., 2008) was validated by qRT-PCR and Western blot analyses (Figures 11D and 11E).

To determine the effect of IMP1 on ITGB1 and ITGB5 RNA stability, transcription of newly transcribed RNAs was blocked using actinomycin D (ActD) treatment, and total RNA was collected after 60 and 120 minutes. Quantification of RNA levels by qRT-PCR revealed that ITGB5, but not ITGB1, was depleted more quickly in the IMP1KD cells, compared to cells treated with a control shRNA (Figure 11F). Thus, our results demonstrate that IMP1 promotes adhesion by stabilizing expression of ITGB5 in hESCs.

BCL2, a novel target of IMP1, enhances survival of IMP1-depleted hESCs

With respect to the mechanisms that underlie decreased cell survival when IMP1 is depleted (Figures 6G and 6H), we examined anti-apoptotic proteins as potential new targets of IMP1. We found that BCL2 (B-cell lymphoma 2) was decreased by 2-fold by RNA-seq analysis in IMP1KD cells (Figure 13A). The decrease in BCL2 was confirmed by both qRT-PCR (Figure 13A) and at the protein level by Western blot analysis (Figure 13B). RIP followed by RT-PCR was performed in HUES6 hESCs to confirm CLIP-seq data demonstrating an interaction between IMP1 and BCL2 RNA (Figure 13C). To test whether introduction of BCL2 can rescue the increased cell death of IMP1 KD cells, a doxycycline-inducible lentiviral system was used to ectopically express BCL2 in IMP1-depleted hESCs. Importantly, BCL2 induction partially rescued the survival of IMP1 depleted hESCs (Figures 13D and 13E) suggesting that the pro-survival function of IMP1 in hESCs is due, in part, to direct regulation of BCL2.

Discussion

Here we demonstrate that targeted depletion of IMP1 in hESCs leads to increased apoptosis and loss of cell adhesion in hPSCs. The combination of CLIP-seq and RNA-seq analysis of RNAs in hPSCs affected by loss of IMP1 identified integrin mRNAs as new IMP1 targets which are downregulated upon loss of IMP1 and further analysis demonstrated that ITGB5 is affected by IMP1 directly at the level of mRNA stability. We also identified BCL2, the founding member of the Bcl-2 family of proteins that control apoptosis, as a novel IMP1 mRNA target that decreases upon IMP1 depletion (Figure 7). Re-expression of BCL2 in hESCs partially rescues the cell death phenotype resulting from IMP1 depletion, suggesting that BCL2 is indeed a physiologically relevant RNA target of IMP1. In addition, we demonstrate that IMP1 affects the expression of the pluripotency factors NANOG, OCT4, and SOX2 via direct and indirect mechanisms, and in doing so, IMP1 likely controls the delicate transcriptional network necessary to maintain rigorous differentiation decisions. To our knowledge this is the first report of genome-wide targets for IMP1 and IMP2 in an endogenous system. PAR-CLIP (Hafner et al., 2010) and RIP (Jonson et al., 2007) studies have previously suggested many targets for IMP1, but these studies were based on the stable expression of Flag-tagged proteins in HEK293 cells, and others have now reported that the stable expression of IMP1 results in aberrant sedimentation in polysomal gradient centrifugation when compared with endogenous protein (Bell et al., 2013).

Systematic, transcriptome-wide mapping identified thousands of IMP1 and IMP2 binding sites within RNA targets in hESCs, demonstrating a largely overlapping set of IMP1 and IMP2 target RNAs. In addition, we found that IMP1 and IMP2 have a preference for interacting with 3' UTRs within protein-coding genes. Sequence motifs identified in our study were consistent with IMP target sequences described by a previous in vitro study (Chao et al. 2010), but distinct from motifs identified by (Hafner et al., 2010) which analyzed the binding of ectopically expressed IMP proteins in HEK293 cells. The stoichiometry of the IMP proteins, together with the ability to bind with up to six RNA binding domains and form protein-protein interactions, likely contribute to variations in the motifs reported for these RBPs.

The LIN28A RBP shares many features with the IMP family, including strong expression in PSCs and preferential binding within 3' UTRs of protein coding genes (Wilbert et al., 2012; Li et al., 2012; Hafner et al., 2013). We also observed LIN28A binding sites within the IMP1 and IMP2 mRNA, and vice versa, indicating cross-regulation among this regulatory circuit, similar to what we have observed for hnRNP proteins (Huelga et al., 2012). Among the genes targeted by both protein families in hESCs, we identified a large proportion of RNA processing factors, and genes involved in neurogenesis and cell motility. These results provide preliminary evidence for an overlapping RNA network controlled by LIN28A and IMP proteins. Indeed, LIN28 has been shown to co-immunoprecipitate with IMP proteins in proliferating muscle stem cells (Polesskaya et al., 2007). Future studies of gene expression changes

in response to both of these RBPs will illuminate possible cooperative or antagonistic effects in stem cells and during tumorigenesis.

Our data also provide evidence for broad regulation of non-coding RNAs by IMP proteins, which have not been previously appreciated. Expression changes of long ncRNAs upon IMP1 depletion in hESCs could have interesting implications for the role of IMP proteins during early embryogenesis. For example, XIST is transcribed exclusively from the X inactivation center (XIC) of the inactive X chromosome and is critical for initiating spreading of X-inactivation for dosage compensation during mammalian development (Penny et al., 1996). XIST increases 2.5 fold upon depletion of IMP1 in hESCs ($p = 1.49E-5$)(Table S6). Importantly, IMP-ncRNA interactions that do not change in response to IMP1 depletion could also be biologically relevant, as lncRNAs that were bound by IMP1 but not affected at the RNA level upon IMP1 depletion are potential candidates for co-factor lncRNAs that could function to stabilize IMP1-mRNA interactions, as is the case for HULC-IMP1-MYC (Hammerle et al., 2013). IMP1 binding to lncRNAs could also affect subcellular localization, such as the localization of H19 to lamellipodia in NIH3T3 cells (Runge et al., 2000).

Our findings have key implications for interpretation of developmental phenotypes observed in IMP1 mouse studies, as well as for human health. For instance, the IMP1 knockout mouse exhibits dwarfism and intestinal hypoplasia, which may be due to improper morphogenesis from loss of extracellular signaling cues provided by normal cell-cell adhesion (Hansen et al., 2004). Regulation of

adhesion through integrin expression is also associated with the epithelial to mesenchymal transition (EMT) (Zuk et al., 1994). EMT is required for normal tissue homeostasis in the developing embryo, as well as during invasion and metastasis in human tumors (Bolender and Markwald, 1979; Frixen et al., 1991). Consistent with these findings, IMP1 is often upregulated in a variety of epithelial tumors and is associated with poor patient prognosis (Dimitriadis et al., 2007; Gu et al., 2004). Stabilization of BCL2 expression in these tumors may endow IMP1 expressing tumor cells an additional advantage, even after treatments with radiation and/or chemotherapy (Hanahan and Weinberg, 2011). Further evaluation of IMP1-target interactions in specific tumor types will provide the necessary knowledge for designing anti-cancer therapeutics. Targeting antisense oligonucleotides against IMP1 specifically in cancer cells may be one such approach to combat this critical stage in cancer progression (King et al., 2014); (Coulis et al., 1999). In summary, our study reveals distinct and specific modes by which the IMP proteins control a convergent network of phenotypically relevant targets (Figure 7), which exist within coding and non-coding RNA regulatory circuits, providing a framework for future characterization of roles of RNA-binding proteins in pluripotent stem cells and cancer.

Methods

PAR-CLIP data

Published PAR-CLIP datasets of IGF2BP binding sites were obtained from Hafner et al., 2010. IMP binding sites were designated using the combined analysis of 4-thiouridine (4SU) labeled datasets. The liftOver tool from the UCSC Genome

Browser utilities (http://genome.ucsc.edu/) was used to map the published NCBI36/hg18 coordinates to the GRch37/hg19 genome build.

CLIP-seq and iCLIP

The CLIP-seq protocol was performed as previously described (Wilbert et al., 2012). Separate sample preparations of IMP1 CLIP-seq were generated using antibodies against the endogenous IMP1 protein (Cell signaling #2852, MBL RN007P) in the H9 hESCs. We performed CLIP-seq for IMP2 in the same manner as IMP1 in H9 cells using an antibody against endogenous IMP2 (MBL, RN008P). To improve the complexity to our IMP2 CLIP-seq dataset we repeated this experiment in a second hESC line, HUES6. The iCLIP protocol was used to identify binding sites in k562 cells, and was adapted from (Konig et al., 2011). Briefly, ten million UV crosslinked K562 cells were lysed, IMP1 and IMP2 protein-RNA complexes were immunoprecipitated using RN007P and RN008P polyclonal antibodies (MBL), respectively. For library generation, RNA associate with protein was trimmed using 2U RNase I (Low RNase treatment) and 40U (High RNase treatment) used for visualizing antibody specificity. Immunoprecipitation of proteins from cells which undergone no UV crosslinking was used to determine specificity of protein-RNA interactions. Radioisotope labeled ribonuclear particle protein complexes were resolved on SDS-PAGE, transferred to nitrocellulose membrane and visualized by autoradiogram. Two populations of RNAs associated with the molecular weight of protein and above the molecular weight of protein and from two different RNAse

treatments were isolated separately and reverse transcribed using barcoded RT primers. Final cDNA library was amplified for 25 cycles (IMP1) and for 27 cycles (IMP2), gel purified and sequenced on Hiseq2500 for 50 cycles of single end reading. Data were sequenced, and the demultiplexed using a randomer-aware algorithm that used only exact matches. Reads from the libraries were then trimmed of polyA tails, adapters, and low quality ends using cutadapt with parameters --match-read-wildcards --times 2e^-05 --quality-cutoff' 6 -m 18 -b TCGTATGCCGTCTTCTGCTTG -b ATCTCGTATGCCGTCTTCTGCTTG -b CGACAGGTTCAGAGTTCTACAGTCCGACGATC -b TGGAATTCTCGGGTGCCAAGG -b AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA -b TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT. Reads were then mapped against a database of repetitive elements derived from RepBase18.05. Bowtie version 1.0.0 with parameters -S -q -p 16 -e 100 -l 20 was used to align reads against an index generated from Repbase sequences (Langmead et al., 2009). Reads not mapped to Repbase sequences were aligned to the hg19 human genome (UCSC assembly) using STAR (Dobin et al., 2013) version 2.3.0e with parameters --outSAMunmapped Within –outFilterMultimapNmax 1 –outFilterMultimapScoreRange 1. Finally reads with the same start position and the same randomer were collapsed into one read.

hESC and k562 Cell culture

H9 and HUES6 human embryonic stem cell lines were grown on Matrigel (BD biosciences) using mTeSR1 medium (Stem Cell Technologies). Cells were routinely passaged using Dispase (2mg/ml) and scraping the colonies with a glass pipet. For assays requiring single-cell dissociation, Accutase (Innovative Cell Technologies, Inc) was used followed by culture medium supplemented with 10mM Rock Inhibitor Y-26732 (Calbiochem) for 24 hours. K562 cells were grown in RPMI1640 supplemented with 10% FBS and 1% PenStrep at 400,000 cells/ml.

Lentiviral Vectors, Production and hES infection

PLKO.1 lentiviruses (TRCN0000075149 for IMP1, TRCN00000255463 for IMP2, TRCN0000074675 for IMP3 and Sigma catalog number SHC002 for the untargeted control shRNA) were prepared as concentrated viruses by the Salk Gene Transfer, Targeting, and Therapeutics (GT3) core. Following titering, a dilution series was performed on hESCs to determine maximum shRNA efficiency with minimal cell death. Cells were single-cell dissociated to 200k cells/sample and incubated with concentrated virus for 1 hour at 37 degrees, 5% CO2 before plating out into 1 well of a 6 well plate. Medium was refreshed the following day and selection with 1mg/ml Puromycin (Sigma) began 48hours following transduction and continued for 5 days when the cells were collected for experiments. Viral supernatants were prepared for the GFP-shRNA and BCL2 rescue experiments in 293T cells using the packaging vectors MDL-gagpol, Rev-RSV, and CMV-VSVG and transfection reagent

Lipofectamine 2000 (Invitrogen). We would like to thank Matthew Inlay and the Weissman laboratory for the BCL2-GFP and rTTA-RFP lentiviral vectors (#408 and #329 respectively; Ardehali et al., 2011). The shRNA 2 GFP-shRNA vectors were prepared using the IMP1 shRNA #3 sequence from (Noubissi et al., 2006) and control shRNA sequence from Sigma, SHC002, in an iteration of the LV-GFP backbone (Tiscornia, 2006). Lentiviral supernatants were collected at 24 and 48 hours after transfection, pooled and passed through a 0.45 μM filter to remove cellular debris. Following filtration, virus supernatant was added to hESCs that had been plated out at ~40% confluency using single-cell dissociation the night before and removed by changing the media to regular mTeSR1 the next day.

Immunofluorescence Microscopy

Cells were fixed in 4% paraformaldehyde in PBS at 4°C for 10 minutes. hESCs were permeabilized at room temperature for 15 minutes in 1.0% Triton in PBS. All cells were blocked in 5% donkey serum with 0.1% Triton at room temperature for 30 minutes. The following primary antibodies and dilutions were used: mouse anti-Oct4 (Santa Cruz, #), 1:500; rabbit anti-IMP1 (Cell Signaling, #2852), 1:100; goat anti-IMP1 (Santa Cruz, #SC-5279), 1:50; rabbit anti-IMP2 (MBL, #RN008P), 1:200; rabbit anti-IMP3 (MBL, #RN009P), 1:200. Primary antibodies were incubated overnight at 4 degrees. Secondary antibodies were Alexa donkey 488, 555 and 647 anti-rabbit (Invitrogen), Alexa donkey 488 and 555 anti-mouse (Invitrogen), and Alexa donkey 488, 555, 568 and 594 anti-goat (Invitrogen); all were used at 1:200. To

visualize nuclei, slides were mounted with Vectashield + DAPI (Vector Labs). Images were acquired using an Olympus FluoView1000 confocal microscope.

Flow Cytometry

For BRDU cell cycle analysis, cells were incubated in mTeSR1 containing 10 μM BrdU for 30 minutes. Cells were dissociated using Accutase (Innovative Cell technologies, Inc), rinsed with PBS and fixed in 70% ethanol overnight. The cell pellet was resuspended in ice cold 0.1 M HCl/0.5% TritonX-100 for 10 minutes, after which the cells were boiled for 10 minutes in a water bath and transferred to ice for 5 minutes to cool. After a brief incubation in 0.5% Triton-X100 in PBS, cells were incubated with a rat-anti-BrdU antibody (1:100 dilution; Axyll) for 30 minutes followed by incubation with an Alexa 488 goat anti-rat secondary antibody (1:200; Life Technologies) for 20 minutes. The cell pellet was resuspended in PBS containing 5 μg/ml of Propidium Iodide and 100 μg/ml of RNaseA. Analyses were conducted using a FACScan (BD Bioscience) and data was analyzed using FlowJo software.

The AnnexinV apoptosis assay was performed using the AnnexinV-FITC Apoptosis Detection kit from BD Biosciences according to the manufacturers instructions and analyzed using the FACScan (BD Bioscience). Data was analyzed using FLowJo software.

Pluripotency cell surface marker analysis was performed using an SSEA4 (BD, 1:20) antibody. Cells were single-cell dissociated, washed twice with PBS and incubated with primary antibodies for 30 minutes at 4 degrees before analysis.

Analyses were conducted using a FACScan (BD Bioscience) and data was analyzed using FlowJo software.

Adhesion Assay

After virus transduction and puromycin selection, hES cells were plated out at 30,000 cells per well in a 96 well plate and incubated for 1 hour at 37 degrees, 5% CO2. They were then vortexed at 2000rpm for 15 seconds, washed three times, and fixed with 4% paraformaldehyde for 10mins at room temp. Following fixation, cells were washed and stained with Crystal Violet for 10 minutes. Cells were then rinsed with H20, and left to completely dry for 15 minutes. 2% SDS is added for 20 minutes and followed by absorbance reading on a plate reader.

Confluency Assay

Assay performed using the Celigo Imaging Cell Cytometer (Brooks Life Science Systems) Confluence with Texture Algorithm in 12 well plates. Cells were assayed every day over a 14 day time course with 3 biological replicates. ‘

Western Blot

Cells were washed with PBS and lysed with 10 mM Tris-HCl (pH 8), 150 mM NaCl, 1% Triton X100 and complete protease inhibitor mixture (Roche). Total protein extracts were used for SDS-PAGE, transferred to nitrocellulose membranes (Amersham Biosciences) and analyzed using primary antibodies. Primary antibodies

were incubated overnight at 4 degrees and secondary HRP conjugated antibodies (Jackson ImmunoResearch, 1:10,000) were incubated for 1 hour at room temp. Thermo Pierce ECL detection reagents were used. Antibodies used: anti-OCT4 (Santa Cruz, #sc-5279), 1:1000; anti-IMP1 (Cell Signaling, #2852), 1:1000; anti-IMP2 (MBL, #RN008P), 1:1000; anti-IMP3 (MBL, #RN009P), 1:1000; anti-NANOG (Cell Signaling, #4903S) 1:1000; anti-SOX2 (Cell Signaling, #3579S) 1:500; anti-BCL2 (BD, #610538); anti-ITGB1 (Cell Signaling, #4706S), 1:1000; anti-ITGB5 (Cell Signaling, # 4708P) 1:500; anti-HNRNPC (MBL, # RN052PW) 1:1000; anti-GAPDH (Abcam, #ab8245) 1:10,000; Cleaved-Caspase 3 (Cell Signaling, #9661) 1:500.

RNA extraction and real-time qPCR analysis

Total RNA was isolated using Trizol Reagent (Invitrogen) according to the manufacturer's recommendations, and cDNA synthesized using the SuperScript III Reverse Transcriptase kit for RT-PCR (Invitrogen). Real-time PCR was performed using the SYBR-Green FAST qPCR Master mix (Applied Biosystems). Values of gene expression were normalized using 18s and/or GAPDH (see figure legends) expression and are shown as fold change relative to the value of the sample control. All the samples were done in technical and biological triplicates. A list of the primers used for real time-PCR experiments are listed in Table 8.

Actinomycin D RNA stability Assay

H9 hESCs were treated with 10mg/ml Actinomycin D (Sigma) and treated with Trizol at time 0 (no treatment), 60 minutes and 120 minutes after treatment. RNA decay was measured using RT-qPCR normalized to the amount of RNA at time 0. See "RNA extraction and real-time qPCR analysis" above for RNA decay measurements. Values of gene expression were normalized using *RPLP0*, a gene determined not to change following addition of Actinomycin D.

RNA Immunoprecipitation (RIP) Assay in hESCs

RIP was performed using an input of 2 10 cm plates uncrosslinked HUES6 hES cells lysed with CLIP lysis buffer (Wilbert et al, 2012). 5mg of each antibody, Rb IgG (Santa Cruz, SC-2027) and IMP1 (MBL, RN007P), were coupled to Protein G Dynabeads (Invitrogen) and incubated with pre-cleared cell lysate overnight on rotation at 4 degrees. Immunoprecipitated RNA was isolated from beads using 1mL Trizol according to the manufacturers instructions.

CLIP-seq read processing and cluster analysis

Pre-processing of reads was used to as a quality control step to remove sequences of low quality and polynucleotide run-ons. Sequence resulting from adapters and barcodes are also removed at this step. This processing was performed using cutadapt (code.google.com/p/cutadapt). Clustering parameters included removing reads with the same start and stop site, use of an mRNA transcript length as background, searching within windows of 50-200bp, and setting the Poisson

significance cutoff at $p < 0.001$. Control clusters were generated by selecting a same sized sequence as real CLIP-seq clusters a random distance from the transcript start of the target gene, to control for differences in gene expression. Control clusters were also confined to the same genic regions as CLIP-seq clusters, i.e 3UTR/CDS/Intron/5UTR. For all datasets 10 iterations of randomly selected controls were generated unless otherwise specified.

Sample preparations of IMP1 and IMP2 CLIP-seq were combined at the level of clusters to define the final binding sites and transcript targets described herein. That is, individual CLIP-seq libraries were prepared from two sample preparations of IMP1 in H9 hESCs with anti-IMP1 Cell Signaling #2852 (samples 1) and 2nd MBL RN007P (sample 2), and IMP2 in H9 (sample 1) and HUES6 (sample 2) hESCs with MBL RN007P. These datasets were initially processed independently as described above to determine significant clusters of reads. Downstream analysis indicated that both of these sample pairs recapitulated the respective properties established for the IMP proteins, as determined by published and data presented here. For example, motifs enriched and binding site distributions were common for the separate CLIP-seq samples. After we had established these samples represented subsets of total cellular IMP-RNA maps we found the junction of cluster coordinates to generate a final dataset for each protein, IMP1 and IMP2.

RNA-seq data processing and differential expression analysis

Reads were mapped to the human genome build hg19 (hg19 http://genome.ucsc.edu) using RNAStar. Quality control confirmed the average length of mapped reads was 50bp as expected. Non-coding RNA analyses excluded rRNAs since these were removed by RiboMinus treatment (Life Technologies) during RNA-seq library preparation, and are therefore not appropriate to consider within these results. The results of reads mapped from RNA-seq experiments are found in Table S3.

Gene ontology analysis

The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID Bioinformatic Resources 6.7; http://david.abcc.ncifcrf.gov/) was used to generate gene ontology associations and assign functional categories to genes (Huang et al., 2009a,b). The set of transcripts from the human genome hg19 was used as background unless otherwise specified. In particular, when transcript datasets from hESCs (and no other cell type) were analyzed the H9 hESC transcriptome was used as background. The KEGG biochemical pathway figure (Figure 12) was created in DAVID using annotated gene functions (Kanehisa et al., 2000).

Human mRNA expression array data analysis

Primary fetal and adult human tissue samples and cell lines were subjected to custom microarray platforms profiling mRNA expression. Normalized probe intensity

values for replicate microarray experiments using GeneChip Human Exon 1.0 ST arrays on human tissues and cell lines were obtained from published (Yeo et al., 2007) and public sources-

http://www.affymetrix.com/support/technical/sample_data/exon_array_data.affx.

Mean and standard deviation of expression values were calculated using Perl scripts.

Acknowledgements

Chapter 2, in full, is an adaptation of material being submitted for publication as Conway, A.E., Wilbert. M.L.W., Landais, S., Sundaraman, B., Liang, T.Y., Pratt G., Essex, A., Hoon, S., Jones, D.L., Yeo, G.W., "IGF2BP/IMP family of RNA-binding proteins control an RNA network in pluripotent stem cells to maintain cell survival". The dissertation author is a primary investigator and author for this paper.

Figures

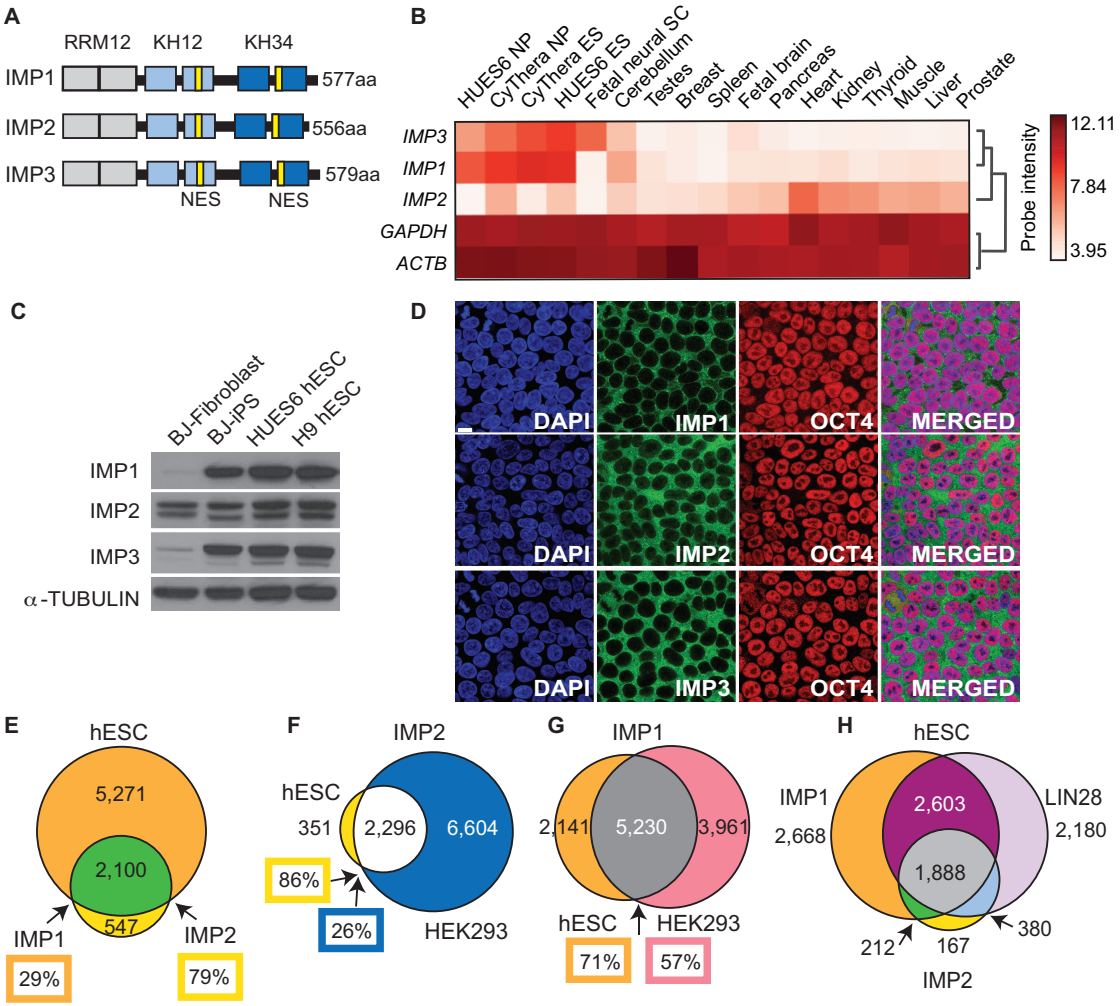

Figure 2. Expression and coding region targets of IMP1 and IMP2 in hESC

(A) Domain structure of IMP protein family members, with RNA-Recognition Motif (RRM) 1-2, hnRNPK-homology (KH) 1-2 and 3-4 domains, and nuclear export signal (NES). (B) *IMP1-3*, *GAPDH*, and *ACTB* mRNA expression in human embryonic stem cells (ES), fetal neural stem cells (SC) and progenitor cells (NP), and tissues. Expression levels are represented by probe intensity normalized across the array experiments. (C) IMP expression in somatic, induced pluripotent (Viswanathan et al.) and human embryonic stem cells (hESC) by Western blot analysis. (D) Immunofluorescence showing IMP localization in hESCs, scale bar represents 10μM. (E) Overlap of gene targets of the endogenous IMP1 and IMP2 determined by CLIP-seq in hESCs. (F-G) Comparison of hESC CLIP-seq data to published PAR-CLIP datasets of exogenously expressed IMP1-2 targets in HEK293 cells (Hafner et al., 2010). (H) Comparison of endogenous targets between IMP1-2 and LIN28 in hESC.

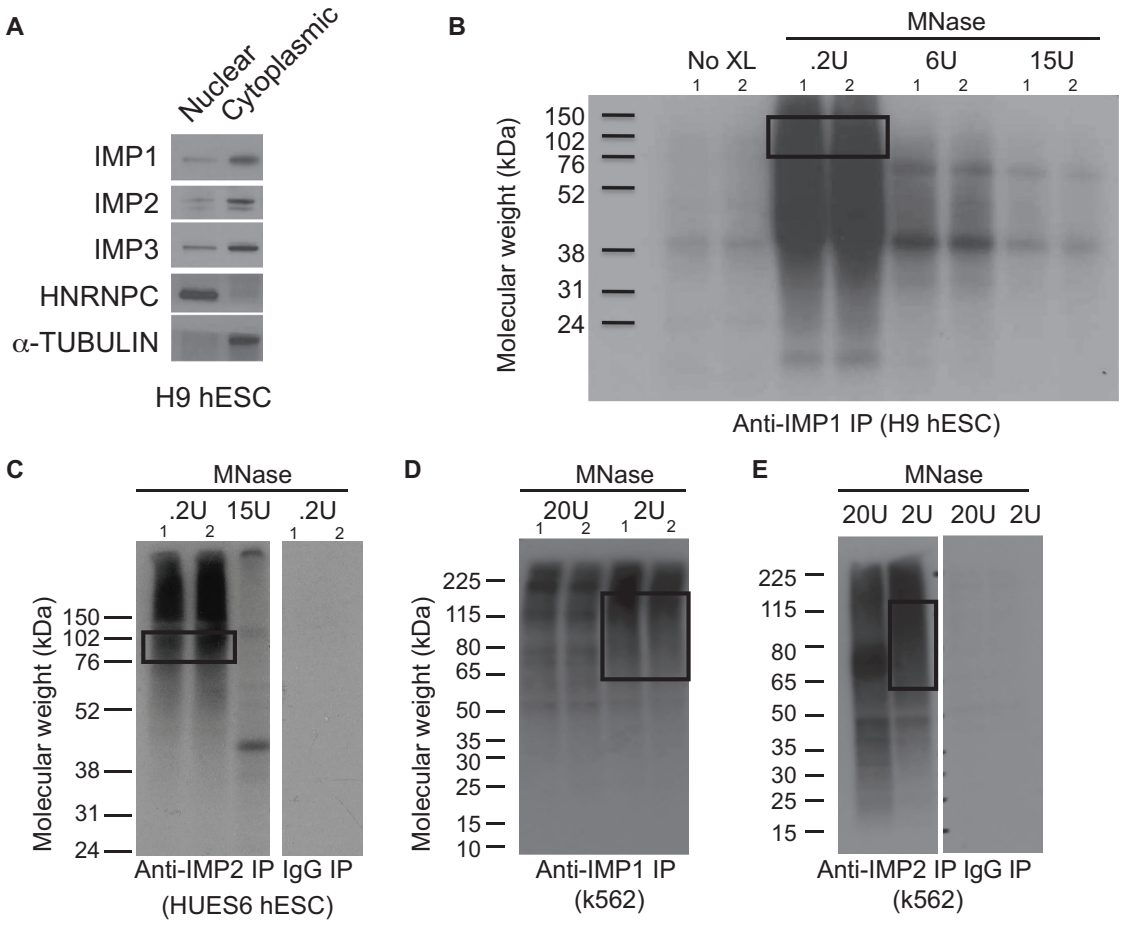

Figure 3. IMP1 and IMP2 are highly expressed and bind a variety of RNA targets in hESCs
(A) Western blot showing localization of the IMP proteins in H9 hESCs. (B) Autoradiograph showing distributions of RNA pulled down by IMP1 and (C) IMP2 in hESC. The boxed region is what was cut out and used for library preparations. (D) Autoradiograph showing distributions of RNA pulled down by IMP1 and (E) IMP2 in k562 cells. The boxed region is what was cut out and used for library preparations.

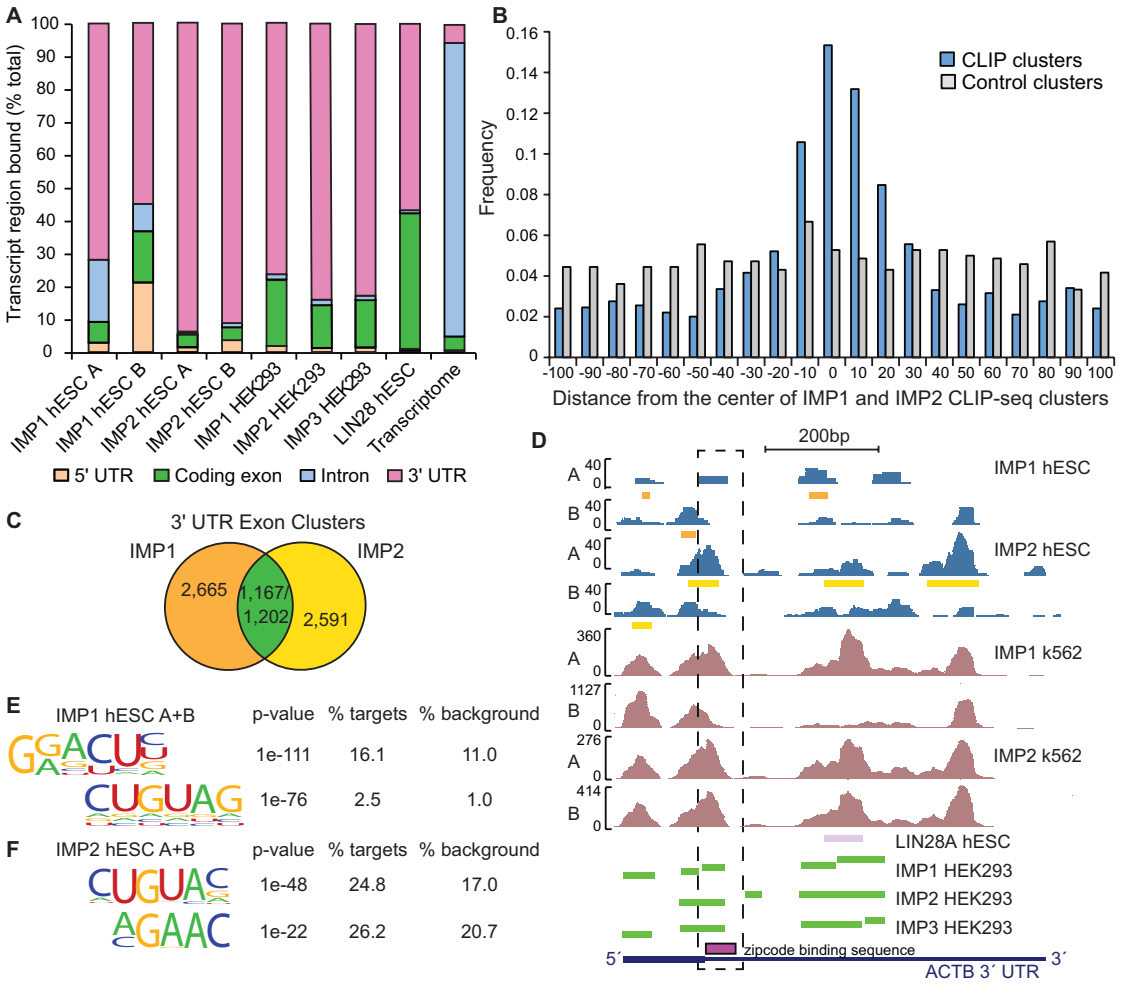

Figure 4. IMP1 and IMP2 bind target transcripts at closely related motifs
(A) Comparison of genic regions occupied by binding sites of IMP and LIN28A proteins in hESC and HEK293, relative to the transcriptome background. (B) Frequency of IMP1 binding sites within 100bp of IMP2 binding sites or randomly located binding sites. (C) Venn diagram depicting the number of IMP1 binding sites within 50bp of IMP2 binding sites within 3' UTRs, and vice versa. (D) Statistically significantly enriched motifs in IMP1 and IMP2 binding sites as calculated using the HOMER *de novo* motif discovery algorithm. (E) Genome browser view of CLIP-seq reads mapped to the 3' UTR of beta-actin. Significant binding sites (CLIP-seq clusters) of IMP1 (orange), IMP2 (yellow), LIN28 (light purple), and IMP1-3 (green; PAR-CLIP sites) are depicted. The zipcode binding sequence defined at the active site of IMP binding by previous studies is shown in purple.

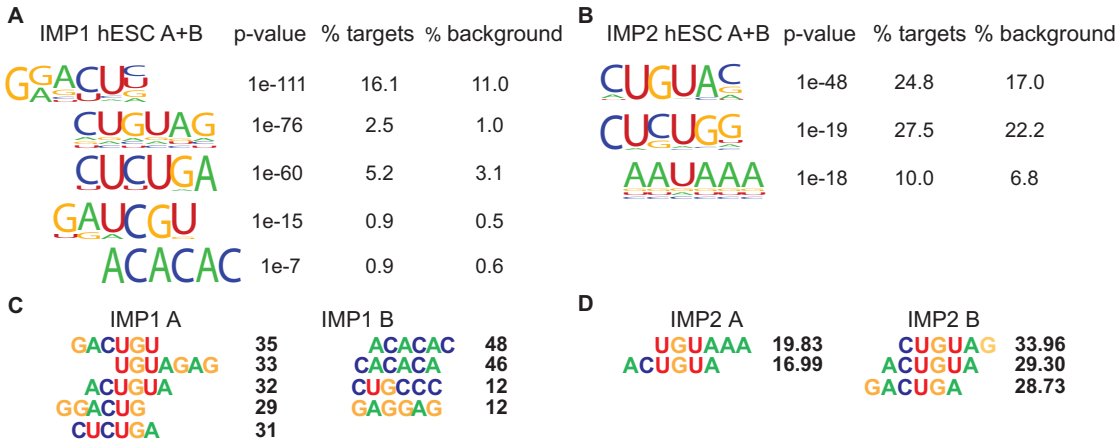

Figure 5. IMP1 and IMP2 bind an array of motifs in hESCs.
Sequence logos generated from HOMER using the combined CLIP clusters for IMP1 (A) and IMP2 (B), with associated p-values and percentage of clusters and randomly located regions of similar sizes (background) that contain the motifs. (C-D) Statistically significantly occurring hexamers in the CLIP clusters in the replicate IMP1 (C) and IMP2 (D) CLIPs are displayed, with associated Z-scores for enrichment.

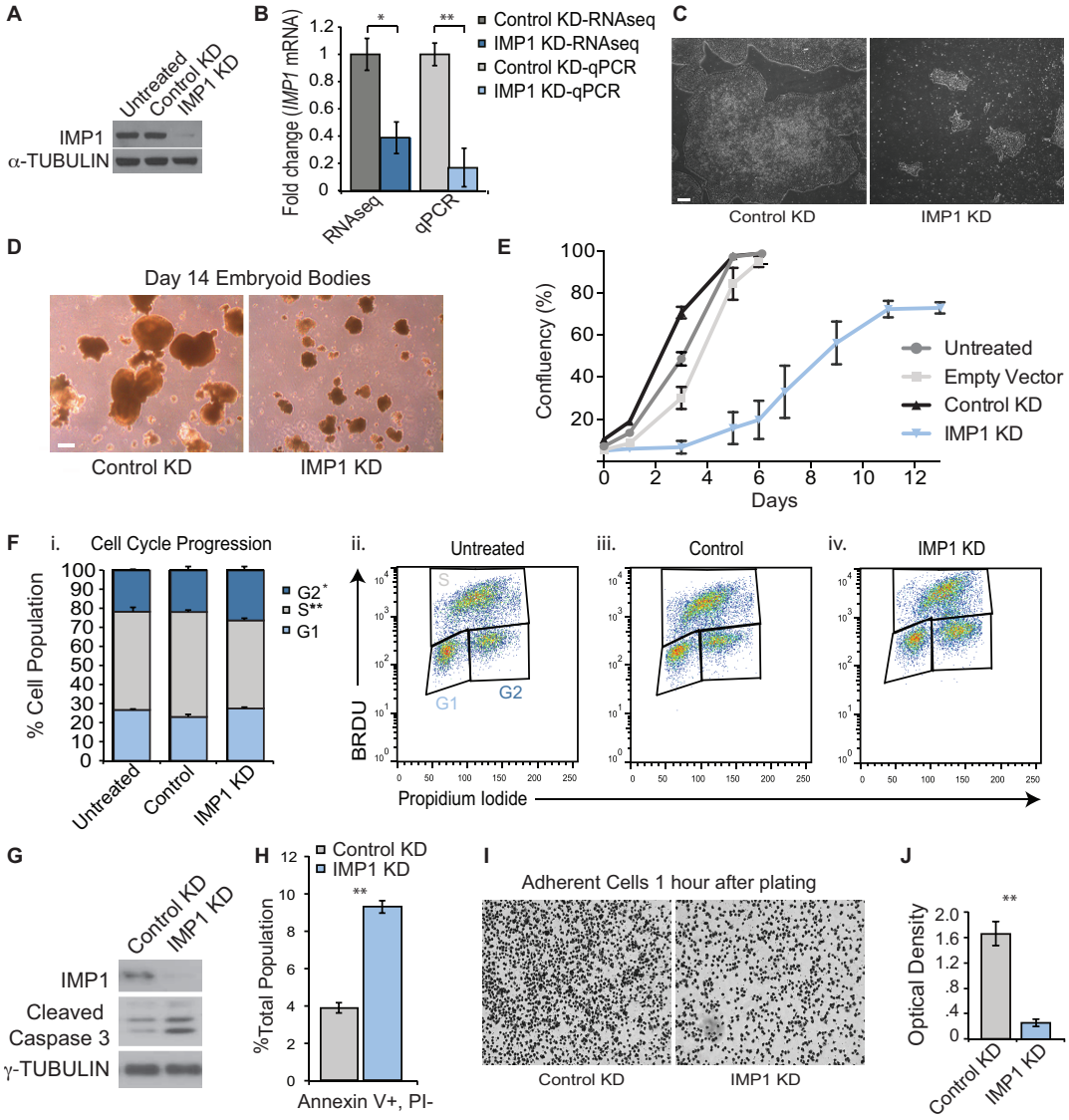

Figure 6. Loss of IMP1 in hESC leads to smaller colonies and less adherent cells
(A) IMP1 protein and (B) mRNA levels following shRNA-mediated depletion of IMP1. (C) Phase contrast images of H9 hESC following depletion of IMP1, scale bar represents 100mM. (D) Phase contrast images of HUES6 Embryoid Bodies (Kretz et al.) following 2 weeks undirected differentiation, scale bar represents 100mM. (E) Repeated measure confluency assay analyzing hESC growth rate over two weeks. (F) Bromodeoxyuridine (BRDU) and Propidium Iodide (PI) proliferation assay analyzed by FACS. (G) WB analyzing expression of cleaved caspase 3 with depletion of IMP1. (H) Flow cytometry analysis of AnnexinV expression following depletion of IMP1. (I) Phase contrast images and (J) Quantification of H9 hESC infected with non-targeting (Control KD) and IMP1-specific (IMP1 KD) lenti-shRNAs. Cells are stained with Crystal Violet one hour after plating. A single asterisk indicates significance of p<.05, two asterisks indicate significance of p<.01. Error bars represent mean +/- S.E.M.

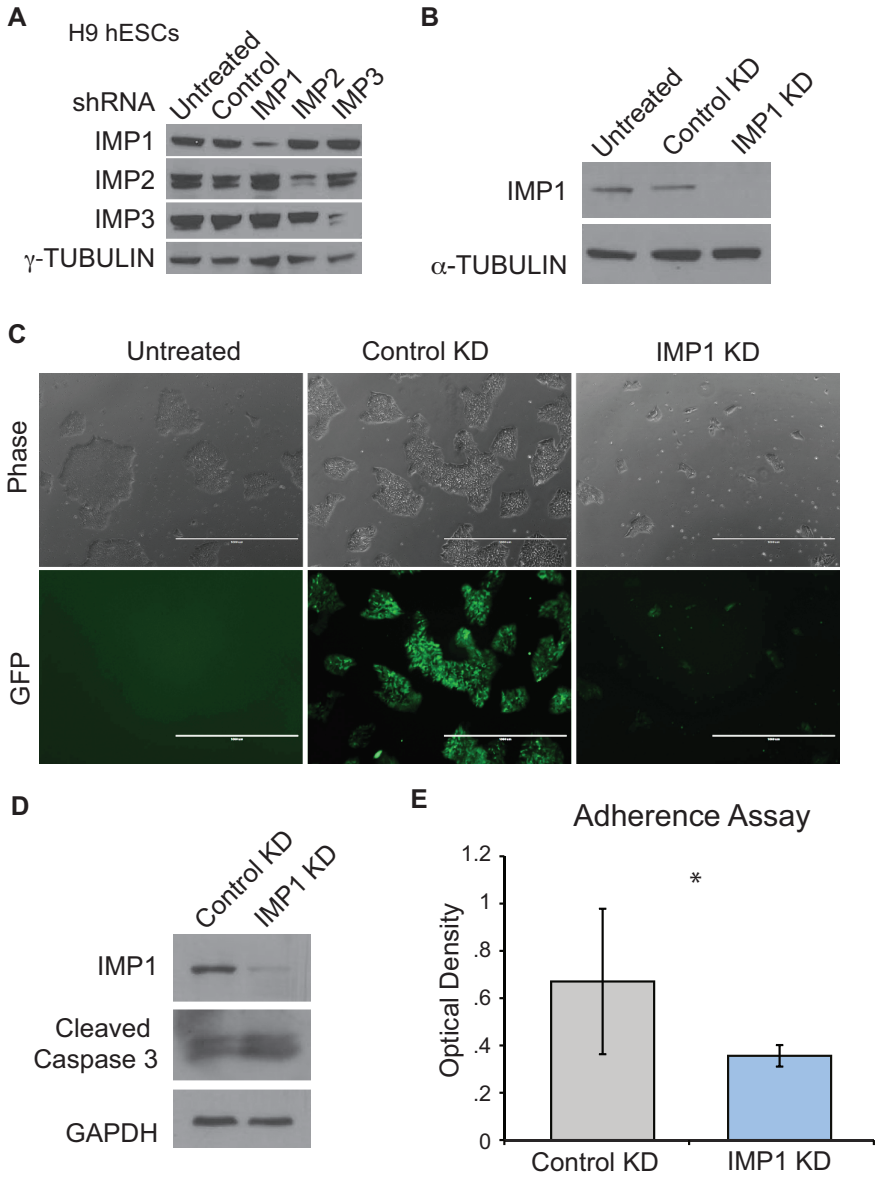

Figure 7. IMP1 protein depletion in hESCs by specific shRNAs.
(A) Transduction with shRNAs specifically targeting IMP1, IMP2, and IMP3 in H9 hESC. (B) Western blot for IMP1 protein levels in H9 hESCs transduced with shRNA 2. (C) Phase contrast and fluorescence images of H9 hESC transduced with shRNA 2. The scale bar represents 1mM. (D) Western blot for cleaved caspase 3 protein expression after transduction with shRNA 2. (E) Adhesion assay in H9 hESC following transduction with shRNA 2. Data are shown as mean +/- SEM, the asterisk denotes $p < 0.05$.

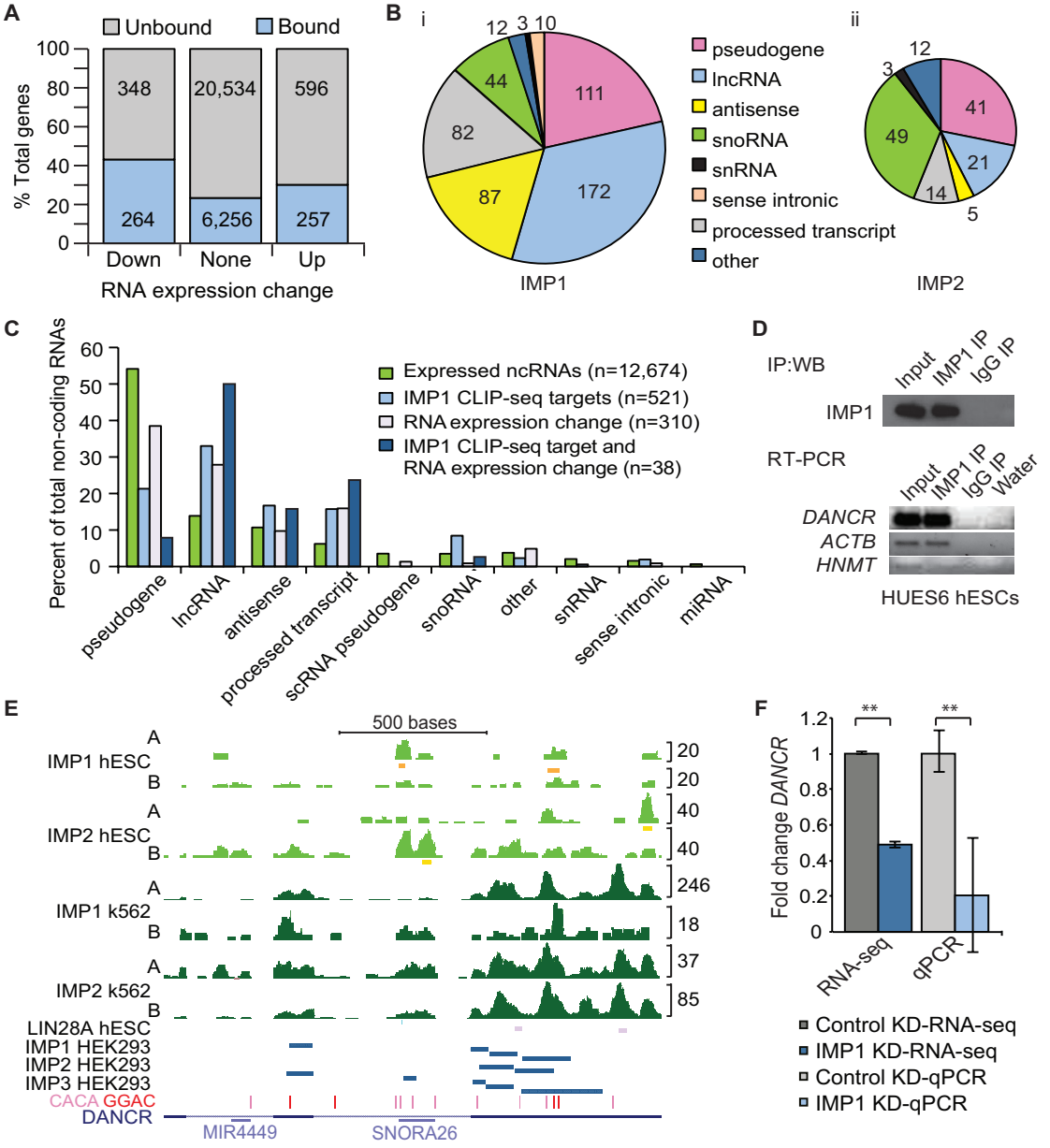

Figure 8. IMP1 and IMP2 interact with non-coding RNA in hESC

A) Fraction of IMP1 gene targets that are upregulated, unchanged or downregulated following depletion of IMP1 in hESC. (B) Pie chart showing the number and types of ncRNAs containing IMP1 and IMP2 binding sites. (C) Percentages of ncRNA classes that are expressed, bound and/or affected by IMP1 in hESCs. (D) RNA immunoprecipitation followed by qRT-PCR of DANCR in HUES6 hESCs. ACTB serves as a positively bound control and HNMT is an unbound control. (E) Genome browser view of CLIP-seq reads mapped to the DANCR lncRNA. Significant binding sites of IMP1 (orange), IMP2 (yellow) and LIN28 (light purple) in hESC, and IMP1-3 (dark-blue; PAR-CLIP sites in HEK293) are depicted. Densities of iCLIP or CLIP sequencing data for K562 and hESC displayed. Significantly enriched motifs are shown in pink (CACA) and red (GGAC). (F) Expression levels of the lncRNA DANCR following loss of IMP1. A single asterisk indicates significance of $p<.05$, two asterisks indicate significance of $p<.01$. Error bars represent mean +/- S.E.M.

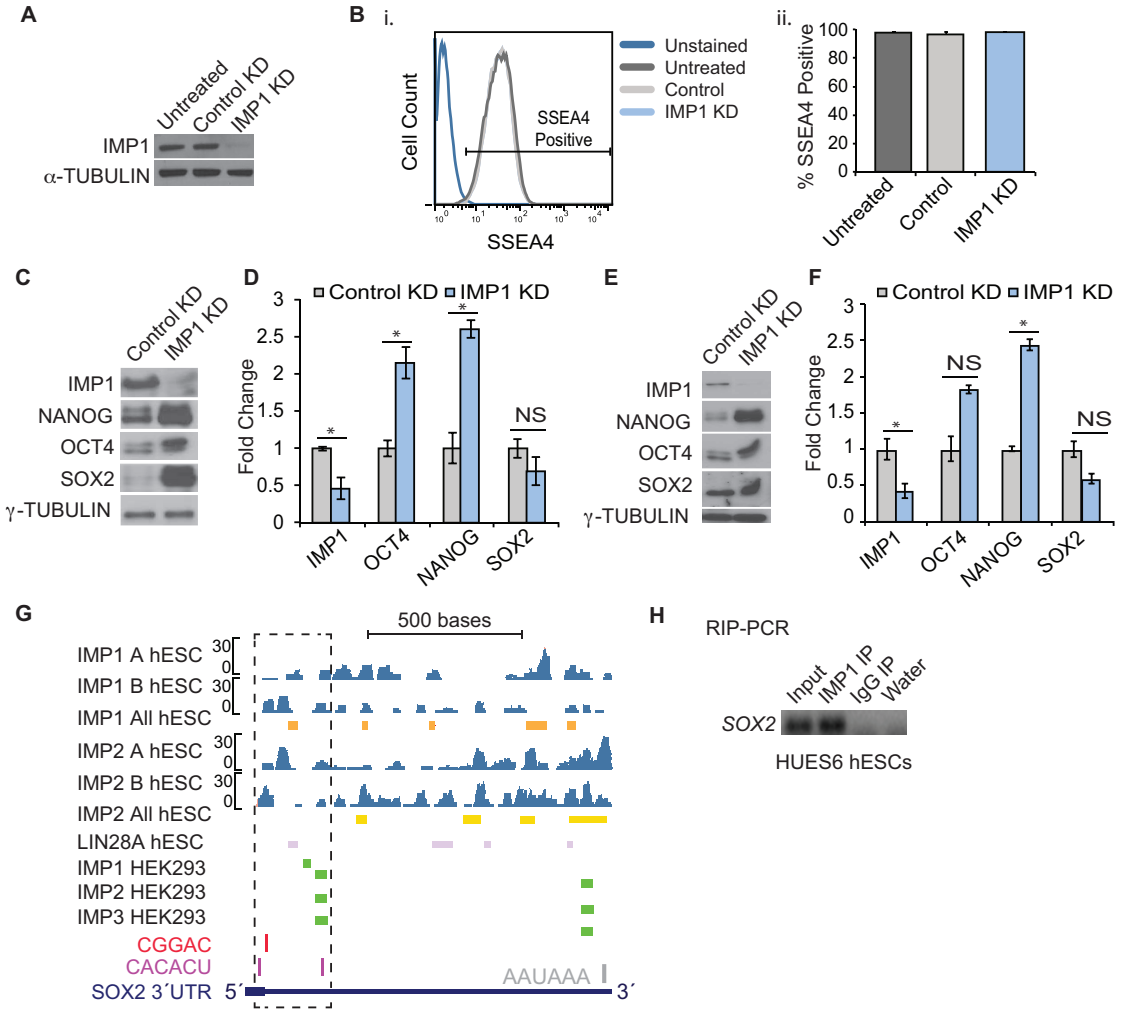

Figure 9. IMP1 affects levels of NANOG, OCT4, and OSX2 in human pluripotent stem cells
(A) Western blot showing IMP1 protein levels following loss of IMP1. (B) FACS analysis for the cell surface marker SSEA4 following loss of IMP1 in hESCs. Protein (Western blot analyses; C and E) and mRNA (qRT-PCR; D and F) levels of NANOG, OCT4 and SOX2 in HUES6 hESC (C and D) and BJ-derived (E and F) iPS cells following depletion of IMP1 and control shRNA treatment. (G) Genome browser view displaying IMP binding sites in the SOX2 3′ UTR. Significant binding sites (CLIP-seq clusters) of IMP1 (orange), IMP2 (yellow) and LIN28 (light purple) in hESC, and IMP1-3 (green; PAR-CLIP sites in HEK293 cells) are depicted. Densities represent CLIP-seq data in hESC. The boxed region highlights the "zipcode binding sequence" motif found in the SOX2 3′ UTR. (H) RNA immunoprecipitation (RIP) of *SOX2* in HUES6 hESC. A single asterisk indicates significance of $p<.05$. Error bars represent mean +/- S.E.M.

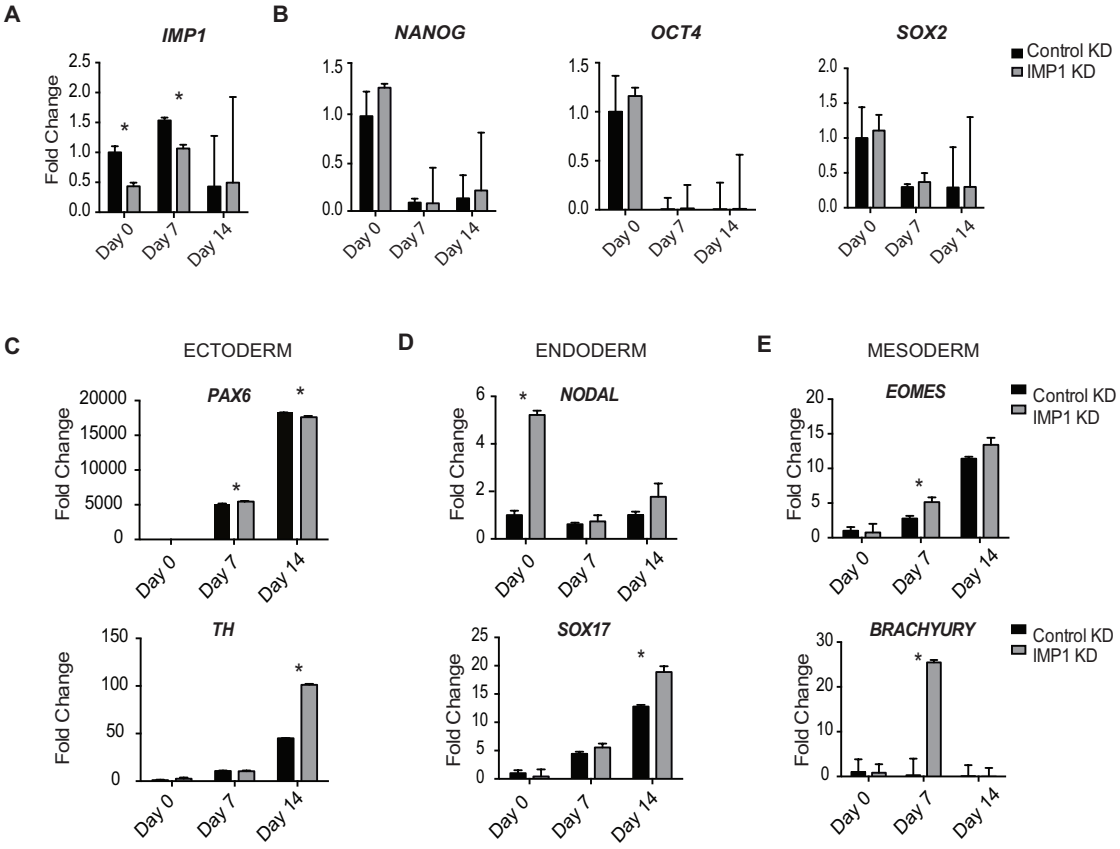

Figure 10. IMP1 KD hESC maintain pluripotency
(A) qRT-PCR for IMP1 over a 2 week undirected EB differentiation assay. (B) qRT-PCR for pluripotency markers over a two week undirected EB differentiation assay. (C) qRT-PCR for markers of the ectodermal lineage in a two week undirected EB differentiation assay. (D) qRT-PCR for markers of the endodermal lineage in a two week undirected EB differentiation assay. (E) qRT-PCR for markers of the mesodermal lineage in a two week undirected EB differentiation assay. A single asterisk indicates significance of $p < 0.05$. Data are shown as mean +/- S.E.M.

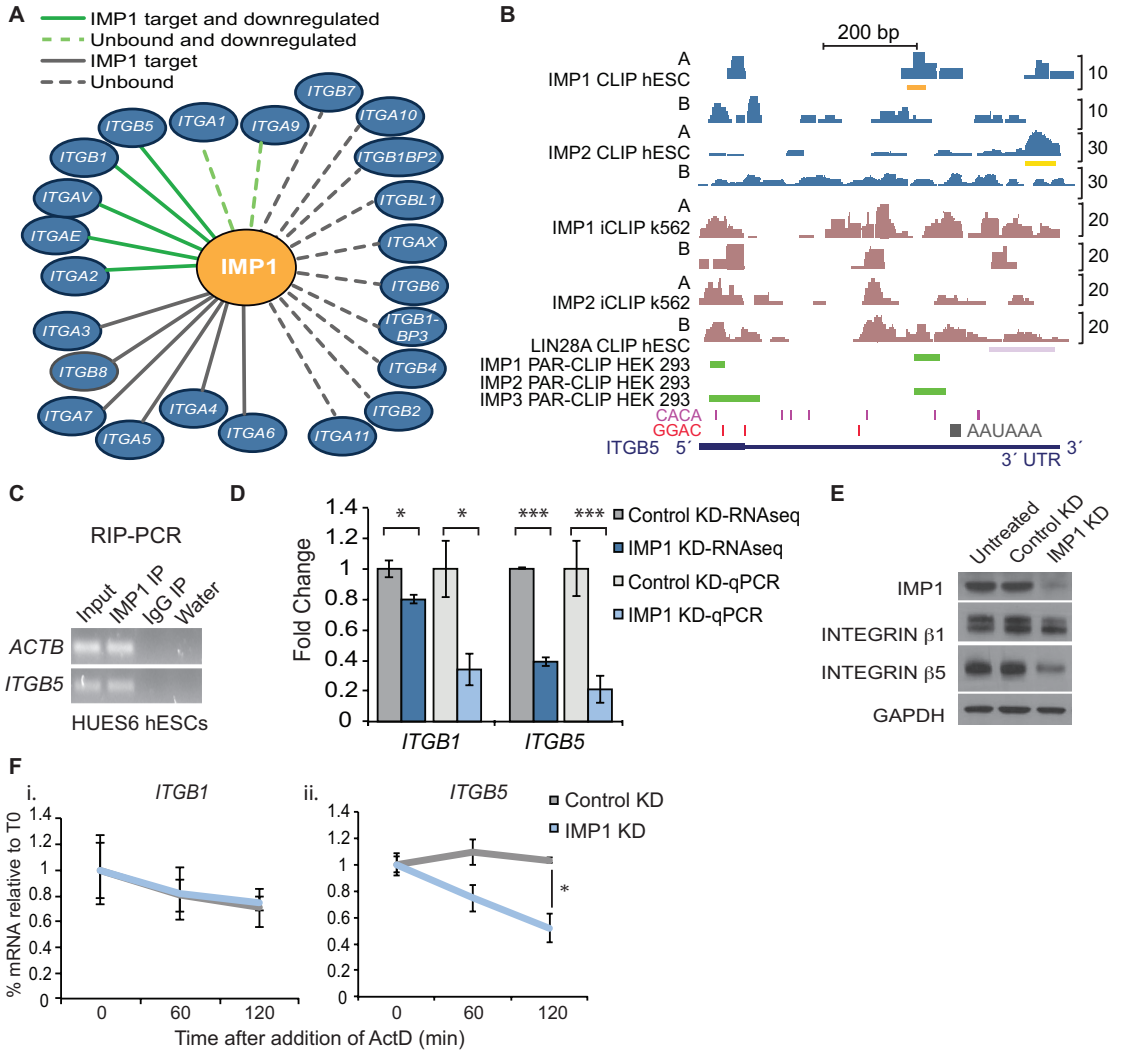

Figure 11. Figure 11: IMP1 controls Integrin RNA STABILITY
(A) Schematic of IMP1 interactions with the Integrin gene family. (B) Genome browser view depicting conserved IMP binding sites in the 3'UTR of *ITGB5*. Significant binding sites of IMP1 (orange), IMP2 (yellow) and LIN28 (light purple) in hESC, and IMP1-3 (green; PAR-CLIP sites in HEK293 cells) are depicted. Densities of CLIP or iCLIP sequencing data from hESC and K562 cells are displayed. Significantly enriched motifs are shown in purple (CACA) and red (GGAC). (C) RNA immunoprecipitation (RIP) for *ITGB5* in HUES6 hESCs. (D) Expression of Integrins b1 (ITGB1) and b5 (ITGB5) following depletion of IMP1 as measured by RNAseq and qRT-PCR, normalized to *18s* rRNA. (E) ITGB1 and ITGB5 protein levels following depletion of IMP1 in H9 hESCs. (F) Expression of *ITGB1* and *ITGB5* following addition of 10mg/ml Actinomycin D in H9 hESC as measured by qRT-PCR. A single asterisk indicates significance of p<.05, two asterisks indicate significance of p<.01, and three asterisks indicate a significance value of a p<.001. Error bars represent mean +/- S.E.M.

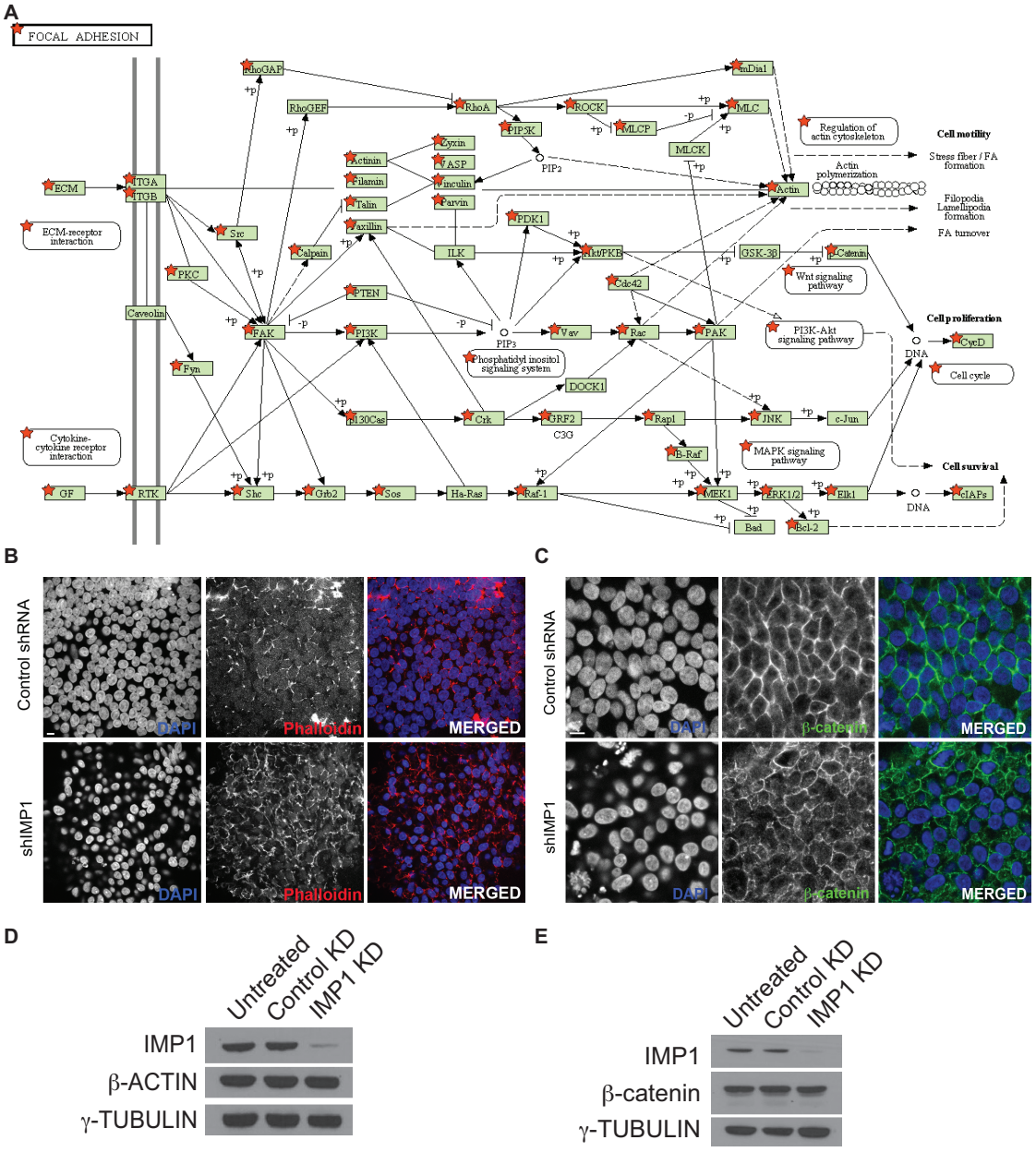

Figure 12. IMP1 binds and regulates cell adhesion and cytoskeletal pathways.
(A) KEGG Pathway diagram featuring the focal adhesion pathway. Gene targets (green boxes) bound by IMP1 in hESCs are designated with red stars. (B) Immunofluorescence microscopy for Phalloidin in H9 hESC following loss of IMP1, scale bar represents 10mM. (C) Immunofluorescence microscopy for b-catenin in H9 hESC following loss of IMP1, scale bar represents 10mM. (D, E) Western blots for detection of b-ACTIN and b-catenin (respectively) following depletion of IMP1 in H9 hESC.

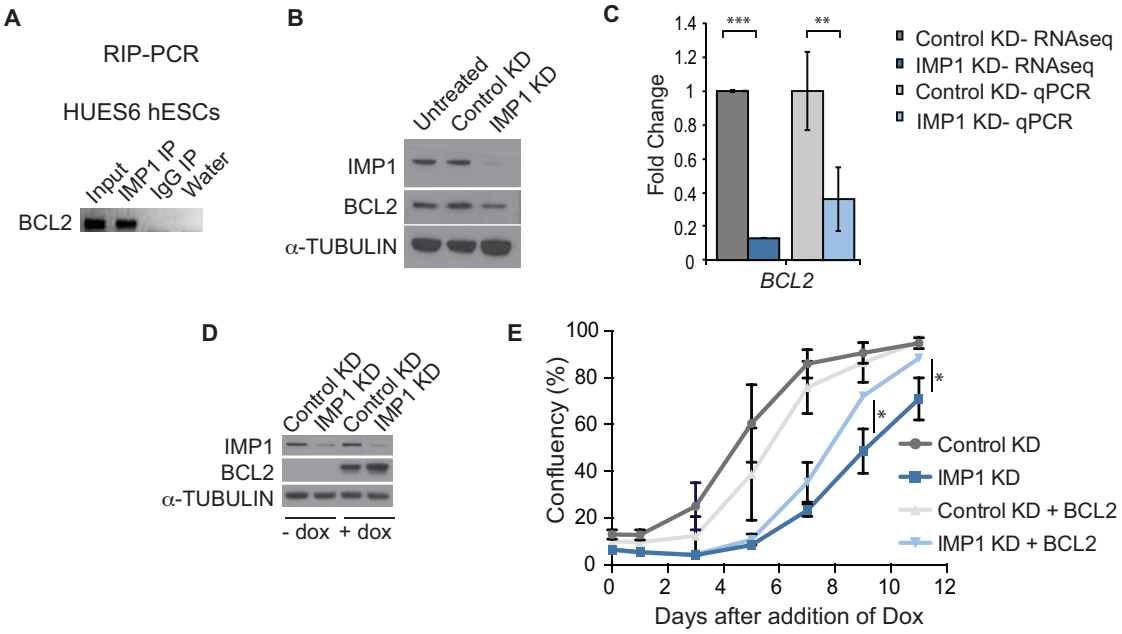

Figure 13. IMP1 promotes cell survival through regulation of BCL2.
A) RNA immunoprecipitation (RIP) for *BCL2* in HUES6 hESC using antibodies that target either IMP1 or IgG. (B) BCL2 protein expression with depletion of IMP1. (C) *BCL2* expression following depletion of IMP1 by RNAseq and qPCR validation. (D) Western blot of BCL2 over-expression in hESCs in the context of depleted levels of IMP1.(E) Confluency assay with BCL2 over-expressionin hESCs. A single asterisk indicates significance of $p<.05$, two asterisks indicate significance of $p<.01$, and three asterisks indicate a significance value of a $p<.001$. Error bars represent mean +/- S.E.M.

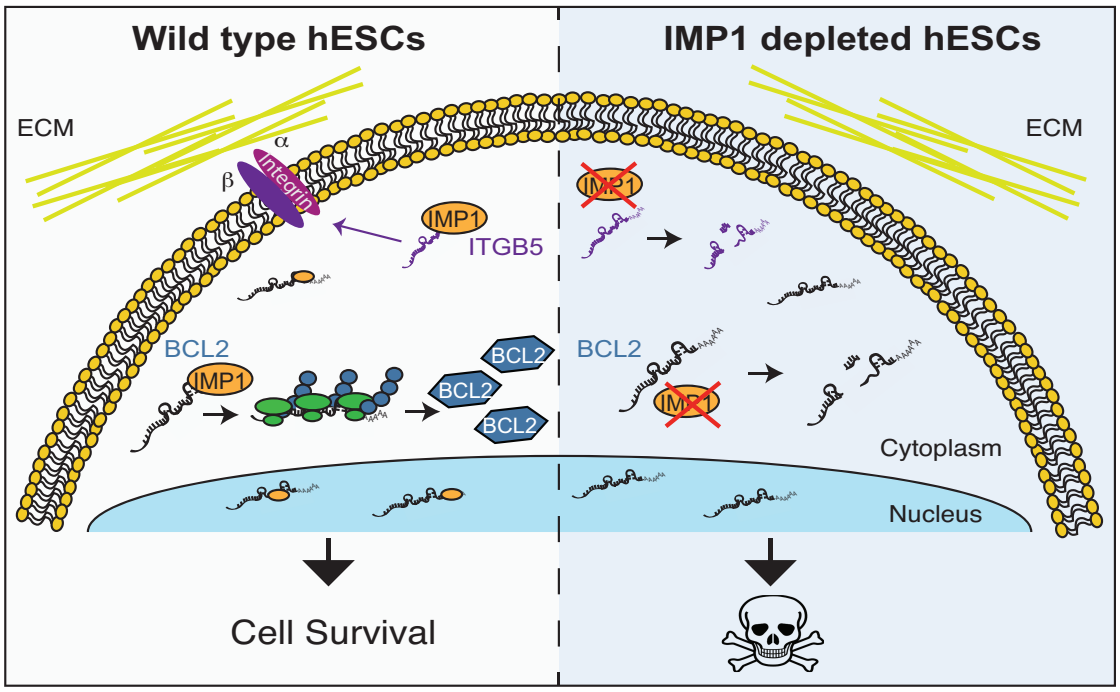

Figure 14. Model of IMP1 function in hESC.

Tables

Table 2. CLIP sequencing library processing

Sample	Seq run	Raw reads	Uniquely aligned reads	Non-redundant reads (same start)	Transcript targets	Cluster sites
IMP1 Sample A H9		26,109,132	18,326,992 (90%)		23,985	7,371
		238,013,222	141,972,098 (88%)			
	All	264,122,354	160,299,090	1,855,166		
IMP1 Sample B H9		24,258,957	4,245,720 (49%)			
		115,417,043	26,678,579 (70%)			
	ll	139,676,000	30,924,299	2,854,713		
IMP2 Sample A HUES6		27,584,445	19,179,734 (87%)		6,170	2,647
		19,697,794	12,759,724 (85%)			
		25,220,910	15,887,997 (86%)			
	All	72,503,149	47,827,455	880,025		
IMP2 Sample B H9		182,455,177	38,735,986 (74%)	4,187,247		
IMP1 Sample A K562		354,337,123	165,949,475 (46%)	13,260,243		
IMP1 Sample B K562		127,220,709	76,975,882 (60%)	1,391,414		
IMP2 Sample A K562		139,390,955	56,102,048 (40%)	5,564,402		
IMP2 Sample B K562		169,611,402	83,767,134 (49%)	8,751,702		

Table 3. RNA-seq processing and expression analysis

Sample	**Raw reads**	**Uniquely aligned reads (%)**	**Expressed Transcripts (RPKM >=1)**	**Differential expression IMP1 KD v Ctl KD (p<0.01)**
Ctl KD A	54210262	42398504 (78%)	29,159	Expressed: 28,255
Ctl KD B	25518766	19531864 (77%)	26,090	
IMP1 KD A	35846578	26931930 (75%)	27,946	Changed: 2,938
IMP1 KD B	37528602	28915082 (77%)	28,391	Up: 1,386
IMP1 KD C	36672029	27065313 (74%)	27,867	Down: 1,552

Table 4. Results of IMP1 CLIP-seq and RNA-seq for different transcript types

Type	Expressed transcripts	IMP1 CLIP-seq target	RNA change (p<0.01)	RNA change (p<0.01) & IMP1 CLIP-seq target
scRNA pseudogene	357	0	3	0
ncrna_host	6	0	0	0
Polymorphic pseudogene	16	4	1	0
Sense intronic	158	10	2	0
snoRNA	354	44	2	1
Non-coding	26	3	2	0
lincRNA	1403	172	63	19
Pseudogene	5473	111	87	3
snRNA	208	3	0	0
Not found	2574	0	83	0
TR V gene	6	0	0	0
3' prime overlapping ncRNA	7	0	0	0
Sense overlapping	8	0	1	0
Processed transcript	628	82	36	9
Antisense	1082	87	22	6
miRNA	67	0	0	0
Other	289	3	7	0
Total protein coding	16363	6711	2628	1273
Total non-coding	12662	519	310	38
Total transcripts	29025	7230	2938	1311

Table 5. lncRNAs most significantly affected by IMP1 depletion in hESCs.

ENSEMBL ID	lncRNA	Fold change (log_2)	P-value
EN00000175513	TSGA10IP	2.78	3.72E-14
ENSG00000247095	MIR210HG	2.76	3.40E-19
ENSG00000250337	RP11-46C20.1	2.55	1.06E-48
ENSG00000237087	AC068134.6	1.90	3.22E-05
ENSG00000244342	RP11-129K20.2	1.59	2.89E-11
ENSG00000246130	RP11-875O11.2	1.47	9.79E-05
ENSG00000258701	RP11-477I4.3	1.41	3.65E-03
ENSG00000254066	CTB-181F24.1	1.41	1.95E-03
ENSG00000225889	AC074289.1	1.38	4.74E-05
ENSG00000235706	DICER1-AS	1.37	1.49E-05
ENSG00000229807	XIST	1.30	5.59E-03
ENSG00000234028	AC062029.1	1.29	1.25E-03
ENSG00000247363	RP11-637A17.2	1.19	8.73E-03
ENSG00000185847	RP1-46F2.2	1.17	1.65E-10
ENSG00000153363	LINC00467	1.12	2.29E-03
ENSG00000250682	CTD-2340E1.3	0.96	3.75E-03
ENSG00000248927	CTD-2334D19.1	0.91	4.87E-04
ENSG00000255794	RMST	0.88	5.04E-12
ENSG00000215866	RP11-426L16.8	0.85	8.52E-04
ENSG00000215808	RP11-371I1.2	0.81	5.59E-06
ENSG00000233393	AP000688.29	0.80	5.05E-04
ENSG00000246792	RP11-68L18.1	0.80	4.52E-03
ENSG00000214174	AMZ2P1	0.63	6.31E-04
ENSG00000179743	RP11-169K16.9	0.62	6.53E-03
ENSG00000231889	TRAF3IP2-AS1	0.60	1.63E-03
ENSG00000248329	RP11-366M4.3	0.57	1.69E-04
ENSG00000227403	AC009299.3	0.55	9.18E-03
ENSG00000212978	AC016747.3	0.53	8.09E-03
ENSG00000223546	LL0XNC01-157D4.1	0.46	6.36E-03
ENSG00000250366	RP11-185P18.1	0.46	3.29E-05
ENSG00000196756	RP4-564F22.2	-0.37	4.93E-03
ENSG00000142396	AC020915.4	-0.44	7.45E-05
ENSG00000170846	AC093323.3	-0.46	3.20E-06
ENSG00000242516	RP11-413E6.8	-0.68	9.49E-03
ENSG00000249430	CTD-2231H16.1	-0.73	9.21E-03
ENSG00000230844	RP1-71L16.2	-0.74	9.62E-04
ENSG00000242808	SOX2-OT	-0.75	7.82E-05
ENSG00000250899	RP11-253E3.3	-0.77	9.32E-05
ENSG00000204466	DGKK	-0.87	4.13E-03
ENSG00000088970	PLK1S1	-0.92	1.89E-08
ENSG00000226031	RP1-260J9.2	-0.94	3.49E-03

Table 6. lncRNAs bound and affected by IMP1 in hESCs.

ENSEMBL ID	lncRNA	Fold change (log_2)	P-value	IMP1 CLIP-seq target	IMP2 CLIP-seq target
ENSG00000230937	CTA-55I10.1	1.94	7.50E-16	1	0
ENSG00000225783	MIAT	1.24	7.41E-11	1	1
ENSG00000182165	TP53TG1	1.15	1.42E-05	1	0
ENSG00000250903	RP1-80B9.2	0.92	3.34E-03	1	0
ENSG00000226383	AC093375.1	0.71	2.94E-03	1	0
ENSG00000245937	CTC-228N24.3	0.65	1.57E-03	1	0
ENSG00000237298	AC009948.3	0.51	3.92E-05	1	0
ENSG00000228223	HCG11	0.46	1.02E-03	1	0
ENSG00000177410	ZNFX1-AS1	0.40	1.23E-04	1	1
ENSG00000147676	MAL2	0.22	7.32E-03	1	1
ENSG00000250616	AC012645.1	-0.48	1.58E-03	1	0
ENSG00000215908	CROCCP2	-0.52	1.23E-03	1	0
ENSG00000228649	AC005682.5	-0.60	5.80E-04	1	1
ENSG00000247137	RP11-727A23.5	-0.67	3.92E-03	1	0
ENSG00000253230	RP11-403C10.2	-0.68	9.54E-05	1	0
ENSG00000228592	AP000459.4	-0.71	7.25E-10	1	0
ENSG00000226950	SNHG13	-1.04	2.83E-25	1	1
ENSG00000241732	RP11-38P22.2	-1.07	1.58E-09	1	1
ENSG00000250889	RP11-229C3.2	-1.86	1.26E-07	1	0

Table 7. IMP1 bound transcripts with greatest decrease upon IMP1 knockdown in hESCs.

ENSEMBL ID	Gene symbol	Gene name	Fold change (log2)
ENSG00000121454	LHX4	LIM Homeobox Protein 4	-2.378
ENSG00000206262	AK304483	uncharacterized	-2.055
ENSG00000154736	ADAMTS5	A Disintegrin And Metalloproteinase With Thrombospondin Motifs 5	-1.861
ENSG00000183770	FOXL2	Forkhead Transcription Factor FOXL2	-1.799
ENSG00000075223	SEMA3C	Sema Domain, Immunoglobulin Domain (Ig), Short Basic Domain, Secreted,(Semaphorin) 3C	-1.698
ENSG00000176641	RNF152	E3 Ubiquitin-Protein Ligase RNF152	-1.623
ENSG00000164488	DACT2	Dapper, Antagonist Of Beta-Catenin, Homolog 2 (Xenopus Laevis)	-1.597
ENSG00000106688	SLC1A1	Neuronal And Epithelial Glutamate Transporter	-1.596
ENSG00000182168	UNC5C	Unc-5 Homolog C (C. Elegans)	-1.570
ENSG00000151726	ACSL1	Fatty-Acid-Coenzyme A Ligase, Long-Chain 2	-1.554
ENSG00000180914	OXTR	Oxytocin Receptor	-1.548
ENSG00000165186	PTCHD1	Patched Domain Containing 1	-1.489
ENSG00000120156	TEK	Tunica Interna Endothelial Cell Kinase	-1.434
ENSG00000101445	PPP1R16B	TGF-Beta-Inhibited Membrane-Associated Protein	-1.408
ENSG00000170558	CDH2	Cadherin 2, Type 1, N-Cadherin (Neuronal)	-1.397
ENSG00000163032	VSNL1	Hippocalcin-Like Protein	-1.393
ENSG00000167508	MVD	Mevalonate Pyrophosphate Decarboxylase	-1.391
ENSG00000169851	PCDH7	Brain-Heart Protocadherin	-1.386
ENSG00000144681	STAC	Src Homology Three (SH3) And Cysteine Rich Domain	-1.366
ENSG00000082781	ITGB5	Integrin Beta-5	-1.325
ENSG00000079263	SP140	Lymphoid-Restricted Homolog Of Sp100	-1.320
ENSG00000169174	PCSK9	Neural Apoptosis-Regulated Convertase	-1.297
ENSG00000144824	PHLDB2	Protein LL5-Beta	-1.281
ENSG00000116833	NR5A2	Nuclear Receptor Subfamily 5, Group A, Member 2	-1.278
ENSG00000013293	SLC7A14	Probable Cationic Amino Acid Transporter	-1.277
ENSG00000164161	HHIP	Hedgehog-Interacting Protein	-1.256
ENSG00000148600	PCDH21	Cadherin-Related Family Member 1	-1.229
ENSG00000183722	AK021977	Lipoma HMGIC Fusion Partner	-1.207
ENSG00000130558	OLFM1	Neuronal Olfactomedin-Related ER Localized Protein	-1.207
ENSG00000113319	RASGRF2	Ras Guanine Nucleotide Exchange Factor 2	-1.206

Table 8. Primers used for qRT-PCR

Gene Name	Forward	Reverse
IMP1	AGGCCATCGAAACTTTCTCC	TTTCGGATTTGAATTTTCCG
18s	AGGCATTGACAACAGGGTTC	GTTGCACATCAGCAGCACTT
DANCR	AATGCAGCTGACCCTTACCC	GGCTTCGGTGTAGCAAGTCT
ACTB	GCACAGAGCCTCGCCTT	GTTGTCGACGACGAGCG
HNMT	GGAGCTTCAAAAGTGGGACT	TCTGAGATCAGGTGGTGCTG
ITGB1	GAGTCGCGGAACAGCAG	CAGTCCAATCCAGAAAATTGG
ITGB5	CCTTTCTGTGAGTGCGACAA	TGTAACCTGCATGGCACTTG
BCL2	CTGAGTACCTGAACCGGCA	GAGAAATCAAACAGAGGCCG
OCT4	GGGTTTTTGGGATTAAGTTCTTCA	GCCCCCACCCTTTGTGTT
PAX6	ACAGTCACAGCGGAGTGAATC	ACTTTTGCATCTGCATGGGTC
TH	TCACCAAGTTCGACCCTGAC	CGATCTCAGCAATCAGCTTCC
NANOG	AACCTCAGCTACAAACAGGTG	TGCGTCACACCATTGCTATTC
SOX17	TGAATGTGTCCCAAAACAGCTT	CACACCCAGGACAACATTTCTTT
EOMES	CTTCTACCCGCTGGAGAGTG	GACTGCCGGAAAACTTCTTG
RPLP0	TGCGTCACACCATTGCTATTC	TGCGTCACACCATTGCTATTC
PPIA	TGCGTCACACCATTGCTATTC	TGCGTCACACCATTGCTATTC
NODAL	CAGTACAACGCCTATCGCTGT	TGCATGGTTGGTCGGATGAAA
SOX2	CAAAAATGGCCATGCAGGTT	AGTTGGGATCGAACAAAAGCTATT
BRACHYURY	GCCCTCTCCCTCCCCTCCA	CGGCGCCGTTGCTCACAGACCACAGG

CHAPTER 3 – A ROLE FOR IMP PROTEINS IN RNA LOCALIZATION

Introduction

Given our transcriptome-wide identification of known and novel RNA substrates of the IMP protein family members in the previous chapter we decided to follow up on previous studies that discovered a role for IMP1 in RNA localization. The model for IMP1 (also known as ZBP1) RNA regulation is that it associates with its RNA cargo in the nucleus and subsequently determines the fate of target RNA in the cytoplasm (Nielsen et al., 2002; Oleykinov and Singer, 2003; Nielsen et al., 2003). Oleynikov and Singer used real-time high-speed imaging in living cells to demonstrate that IMP1 is rapidly shuttled out of the nucleus after it is is anchored to an RNA target. Additionally, targets of IMP1 binding have been shown to accumulate in the nucleus when IMP1 is depleted (Sim et al., 2012). Movement of IMP proteins is driven primarily by their highly conserved KH domains, which suggests that it is largely RNA dependent. Farina et al. first established that the last 2 KH domains (KH34) bound the 3'UTR zipcode region of the *ACTB* MRNA and were required for export into the nucleus (Farina et al., 2003). Additional studies confirmed the significance of the KH domains in IMP protein localization identified in this original report (Nielsen et al., 2002; Nielsen et al., 2003; Wachter et al., 2013).

Aside from a role for IMP1 in localizing *ACTB* mRNA in polarized cells, there is also a role for IMP1 in stress granules (SGs) (Stohr et al., 2006). SGs are RNPs involved in translational suppression of gene products and the sequestration of non-essential mRNA transcripts upon induction of cellular stress. This response is thought

to mediate any severe effects to cellular stress and is mediated by RBPs. IMP1 (ZBP1) accumulates in stress granules following arsenite or heat-shock treatment of U2OS human osteosarcoma cells (Stohr et al. 2006). This recruitment to SGs is specific as IMP1 is not recruited to P-bodies, a second type of RNA granule. Additionally, recruitment of the different IMP family members to SGs appears to be specific and is dependent on the 4 KH domains (Wachter et al., 2013). Deletion of KH1-4 in IMP1 and IMP2 severely demolished association with TIA1 positive SGs, while deletion of KH1-4 in IMP3 only slightly led to an enrichment in the cytoplasm. Similar to the RNA localization experiments, both granule formation and localization are unaffected by removal of the two RRMs, whereas deletion of the KH domains, which mediate RNA-binding, impairs these functions (Nielsen et al., 2012; Wachter et al., 2013). Therefore, RNA association and subcellular localization of the IMP proteins occurs in both a paralog and RNA substrate-dependent manner.

To test whether IMP1 and IMP2 have a role in RNA localization in hESCs, we first set out to determine the cellular localization of these proteins in hESC using cellular fractionation techniques. We next examined the association of endogenous IMPs with SGs to determine if this was a conserved function in hESCs. We also applied a protocol developed by Wang and Cody (2012) that combines subcellular fractionation with RNA-seq. The following pages describe unpublished results that aim to advance what is already known about the interaction between IMP proteins and localized RNAs such as *ACTB* in polarized cells to a genome-wide level in a model of mammalian development.

Results

In the previous two chapters I discussed trafficking and localization of IMP proteins and how it can vary from cell type to cell type. In our hESC model endogenous IMPs are localized primarily to the cytoplasm of the cell (Figures 2D and 3A). However, in another cancer cell model used in the lab, IMP2 specifically has an enriched localization in the nucleus compared to the other two family members (Figure 15). This experiment raises questions about specific functions of each paralog in localization of target RNAs.

Additionally, we were interested to know whether IMP1 was able to co-localize with stress granules in hESCs as had been previously reported in other cell types. Indeed, IMP1 is able to co-localize with TIA1 positive RNPs in hESCs upon the induction of oxidative stress with 0.5mM sodium arsenite (SA) (Figure 16A). While there appears to be increased IMP1 signal at the level of immunofluorescence, a WB analyzing IMP1 protein levels in SA treated cells suggests otherwise (Figure 16B). IMP1 protein levels appear to remain the same, they just localize with larger RNPs which changes their distribution in the cell. I found that IMP1 also localizes to SGs in SA treated 293 cells as well as NPCs, suggesting this function is conserved across cell types. IMP localization to stress granules during embryonic development could have broad affects on translational regulation should the embryo become exposed to external stresses. Furthermore, IMP1 is known to co-IP with members of SGs in other cell types that are implicated in neurodegenerative disease (Stohr et al., 2006; Wachter

et al., 2013). Given their expression in developing neurons these data imply there may be a yet-undiscovered role for IMP proteins in SG-associated neurodegeneration.

Lastly, an advantage of CLIP-seq over other methods of detecting genome-wide target RNAs is the ability to identify the precise RNA sequences bound by the protein of interest. These data coupled with an analysis of where the target RNAs are localized could provide a much-needed map of potential RBP-RNA interactions throughout the cell. Recently, Wang and colleagues developed such an approach to analyze the global localization profile of MBNL target transcripts and also identify the functional consequences of MBNL depletion on the transcriptome (Wang et al., 2012). Using their protocol we set out to address these questions for IMP1 and IMP2 in hESCs.

To test whether IMP1 and IMP2 have a general effect on RNA localization in hESCs, we performed IMP (1 or 2) KD (Figure 20A) followed cellular fractionation and high throughput sequencing. Fractionation of hESCs leads to four different cellular compartments, the nuclear, cytoplasmic, membrane, and insoluble fractions in addition to a total RNA sample for 5 RNA samples total/condition (Figure 17A). Before making cDNA libraries and sending these samples off for sequencing we first wanted to validate that known RNAs were properly fractionated to the correct cellular compartment. In order to do this we performed RT-qPCR for markers specific to each of the cellular compartments and compared our data to the Wang et al published dataset. Indeed, our fractionations looked good and we were able to proceed onto library preparation (Figure 17B).

Our very preliminary results suggest that there is perhaps a mild affect on global RNA localization with IMP1 and IMP2 depletion in hESCs, however these results still need to be filtered to examine only target RNAs, and perhaps even only target RNAs that have specific binding motifs to increase specificity (Figures 18A-B)(Wang et al., 2012). While in-depth analysis of these data is still currently being performed, our goal is that the results garnered here will help decipher the RBP-RNA localization code for the IMP family of RBPs.

Discussion

The data presented in this chapter are still in their initial stages consequently there are still many experiments to perform. In terms of the IMP family localization pattern, I propose a more thorough analysis across a wide range of cell types- both immortalized and primary cells, control and disease conditions, with validated antibodies by both cellular fractionation and immunofluorescence/histochemistry would be beneficial to the field. In the very few cell types I've analyzed in my thesis work I've found that the localization patterns vary depending on the cell type. This has large implications for functional analyses and determining the molecular mechanism by which these RBPs regulate their target RNAs.

IMP1 association with stress granules in hESCs is not necessarily novel in and of itself, but I think wider applications of these studies will be relevant for different disease phenotypes as well as in establishing a more standardized model to understand how RBPs regulate their target RNAs in different cellular conditions. For example, during mammalian development the embryo endures in a generally hypoxic state;

additionally, hypoxia is one mechanism that cancer cells overcome during tumorigenic transformation. It's possible that stress granules induced by hypoxia during development may shed light on how cancer cells are able to escape this type of cellular stress. If we are able to understand how stress granules form and the RNAs that are affected in normal conditions we will more likely to respond when things go awry in disease. Additionally, once a standard model is developed for assembly and disassembly of SGs, it will be important to test whether different types of stress affect the model. For example, in my studies I used oxidative stress to induce SGs, but many other agents including translational inhibitors such as puromycin, as well as UV irradiation and heat shock treatments can induce SGs. It may also be that RBPs associate with different RNA targets with different types of stress. To test this hypothesis one could perform a variation of CLIP-seq for different RBPs known to play a role in stress granules and then compare the target genes bound in untreated conditions to targets after various types of stress. Lastly, performing co-immunoprecipitation experiments during all of these conditions will be vital to interpretation of the results because it may be that presence of a single RBP-RNA interaction could affect the proper entire assembly and/or disassembly process.

Lastly, we're combining CLIP-seq and localization-seq to first identify which RNAs are mis-localized with loss of IMP RBPs and then determine the molecular mechanism as well as if there is a general IMP "zipcode" present for localization of target RNAs. This procedure could also be performed in different cell types where alterations in RNA localization may be even more pronounced. For example,

migrating cells and neurons are examples of polarized cells that have requirements for acute RNA localization not present in other cell types. We could isolate RNA from different cellular compartments in neurons growing in chambers or only from the leading edge of migrating cells and perform qPCR for CLIP-seq targets or sequencing for a genome-wide approach. These alterations in cell types and approaches will help paint a picture for specific requirements of RNA localization. While the characterization of localization zipcodes remains a challenging task, studies such as ours characterizing the properties of RBP–RNA interactions, combined with recent genome-wide approaches to elucidate RBP binding specificities in different areas of the cell should pave the way for rapid progress in the field.

Methods

Cell culture and oxidative stress induction

H9 and HUES6 human embryonic stem cell lines were grown on Matrigel (BD biosciences) using mTeSR1 medium (Stem Cell Technologies). Cells were routinely passaged using Dispase (2mg/ml) and scraping the colonies with a glass pipet. For assays requiring single-cell dissociation, Accutase (Innovative Cell Technologies, Inc) was used followed by culture medium supplemented with 10mM Rock Inhibitor Y-26732 (Calbiochem) for 24 hours. K562 cells were grown in RPMI1640 supplemented with 10% FBS and 1% PenStrep at 400,000 cells/ml.

Oxidative stress induction was performed by addition of 0.5mM sodium arsenite for 1 hour and then cells were immediately fixed with 4% Paraformaldehyde.

Lentiviral Transduction of hESCs

PLKO.1 lentiviruses (TRCN0000075149 for IMP1, TRCN00000255463 for IMP2, TRCN0000074675 for IMP3 and Sigma catalog number SHC002 for the untargeted control shRNA) were prepared as concentrated viruses by the Salk Gene Transfer, Targeting, and Therapeutics (GT3) core. Following titering, a dilution series was performed on hESCs to determine maximum shRNA efficiency with minimal cell death. Cells were single-cell dissociated to 200k cells/sample and incubated with concentrated virus for 1 hour at 37 degrees, 5% CO2 before plating out into 1 well of a 6 well plate. Medium was refreshed the following day and selection with 1mg/ml Puromycin (Sigma) began 48hours following transduction and continued for 5 days when the cells were collected for experiments.

Immunofluorescence microscopy

Cells were fixed in 4% paraformaldehyde in PBS at 4°C for 10 minutes. hESCs were permeabilized at room temperature for 15 minutes in 1.0% Triton in PBS. All cells were blocked in 5% donkey serum with 0.1% Triton at room temperature for 30 minutes. The following primary antibodies and dilutions were used: rabbit anti-IMP1 (Cell Signaling, #2852), 1:100; goat anti-IMP1 (Santa Cruz, #SC-5279), 1:50; rabbit anti-IMP2 (MBL, #RN008P), 1:200; rabbit anti-IMP3 (MBL, #RN009P), 1:200; goat anti-TIA-1 (Santa Cruz, #SC-1751). Primary antibodies were incubated overnight at 4 degrees. Secondary antibodies were Alexa donkey 488, 555 and 647 anti-rabbit (Invitrogen), Alexa donkey 488 and 555 anti-mouse (Invitrogen), and

Alexa donkey 488, 555, 568 and 594 anti-goat (Invitrogen); all were used at 1:200. To visualize nuclei, slides were mounted with Vectashield + DAPI (Vector Labs). Images were acquired using an Olympus FluoView1000 confocal microscope.

Nuclear/Cytoplsamic fractionation

Nuclear/Cytoplasmic fractionation was performed with the NE-PER nuclear and cytoplasmic fractionation kit according to the manufacturer's instructions (Pierce, cat#78833). Either one well of a fresh 6 well plate of hESCs (~2 million) or 2 million k562 cells was used as input material. Western blots were performed as described previously (p. 55) with the following antibodies: IMP1 (MBL, RN007P), IMP2 (MBL, MBL, #RN008P), IMP3 (MBL, #RN009P) all 1:1000; HNRNPC (MBL, #RN052PW 1:1000); α-TUBULIN, (Abcam, Ab4074-100) 1:10,000.

Cellular Fractionation followed by high-throughput sequencing

Cellular fractionation was performed according to Wang et al., 2012. H9 hESCs (Scrbl shRNA, IMP1 KD, IMP2 KD, 2 biological replicates each) were grown in monolayers on 10 cm2 tissue culture dishes. Lysis buffers contained 10 mM PIPES, 0.25 M sucrose, 1 mM EGTA, 5 mM MgCl2, and 25 mM NaCl, and the cytosolic lysis buffer additionally contained 0.015% digitonin, the membrane lysis buffer 0.5% Triton X-100, and the cytoskeletal lysis buffer 1 M NaCl. Cells were washed with PBS, and incubated for 10 min at 4°C with 2 ml cytosolic lysis buffer, with gentle rocking. The cytosolic fraction was removed and saved, and the membrane lysis buffer

was added for 5 min at 4°C with gentle rocking. The membrane fraction was removed and saved, and the cytoskeletal buffer was added. The plate was quickly rocked 4 times back and forth, and the cytoskeletal fraction was removed and saved. Protein from each fraction was prepared by heating in SDS page buffer (Invitrogen). RNA from each fraction was prepared by adding guanidinium isothiocyanate to 8 M, vortexing until clear, and subsequent phenol/chloroform extraction. RNeasy mini columns (QIAGEN) with DNase treatment were used to further purify the RNA prior to standard Illumina library construction. For fractionation experiments, Ribo-Zero columns (Epicentre Biotechnologies) were used to reduce ribosomal RNA abundance instead of oligo dT beads. Library preparation was performed with 1μg RNA input material according to the manufacturer's instructions in the TruSeq Stranded Total RNA kit and sequenced on a HiSeq (Illumnia) SE50 flow cell run.

Bioinformatics Analyses

The percent relative contribution for each gene in each first computed by normalizing the RPKM of gene in a fraction again the sum of all fractions. Gene expression was then visualized in a simplex space, using a simplex centered at the origin with corner coordinates of [1,0], [0,1], [1,1] , [0,0] in Cartesian space. The x and y coordinates of each gene were calculated by $\cos(0) + \cos(\pi/2)$, $+\cos(\pi) + \cos(3/2\ \pi)$ and $\sin(0) + \cos(\pi/2)$, $+ \sin(\pi) + \sin(3/2\ \pi)$ respectively. Genes that were predicted to be bound or not bound via CLIP-seq were colored in red, and green respectively.

Acknowledgements

I would like to thank Gabriel Pratt for his invaluable assistance in analyzing the sequencing datasets, Eric Van Nostrand for teaching me the technique of large-scale sequencing library preparation, Julia Nussbacher also for help with the library preparation protocol and lastly Neal Cody for his assistance in performing the cellular fractionations.

Figures

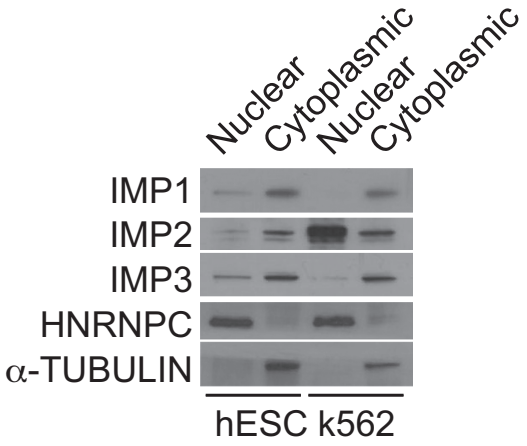

Figure 15. IMP Family RBP expression in fractionated cells
hESCs and k562 cells were fractionated using the Pierce NE-PER fractionation kit. Matched amount of protein were analyzed for proper segregation by western blot, 20μg lysate was run per lane on the gel.

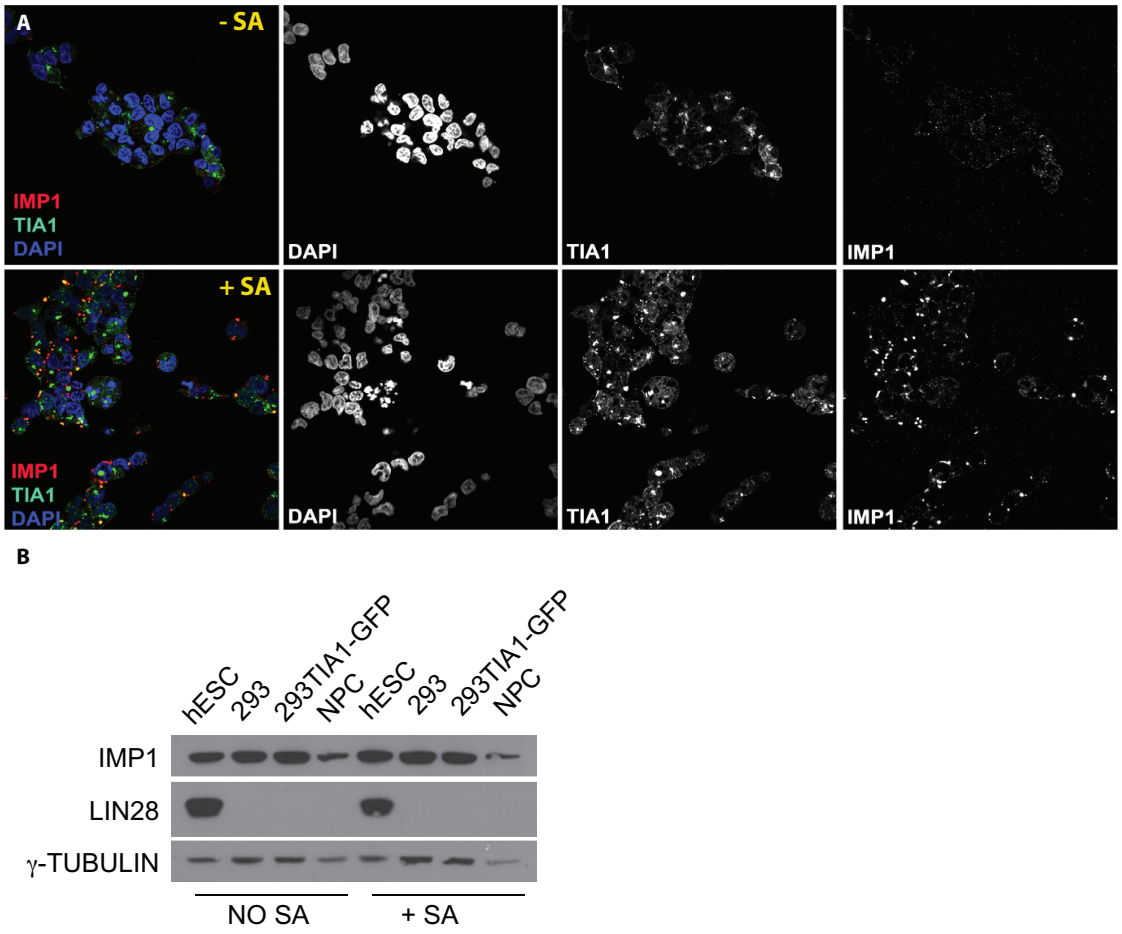

Figure 16. IMP1 is expressed in stress granules in hESC.
(A) Immunofluorescence microscopy for TIA1 (green) and IMP1 (Hammer et al.) in hESCs +/- 0.5mM sodium arsentire (SA). (B) Western blot analyzing total levels of IMP1 and LIN28 protein in H9 hESC +/- SA treatment.

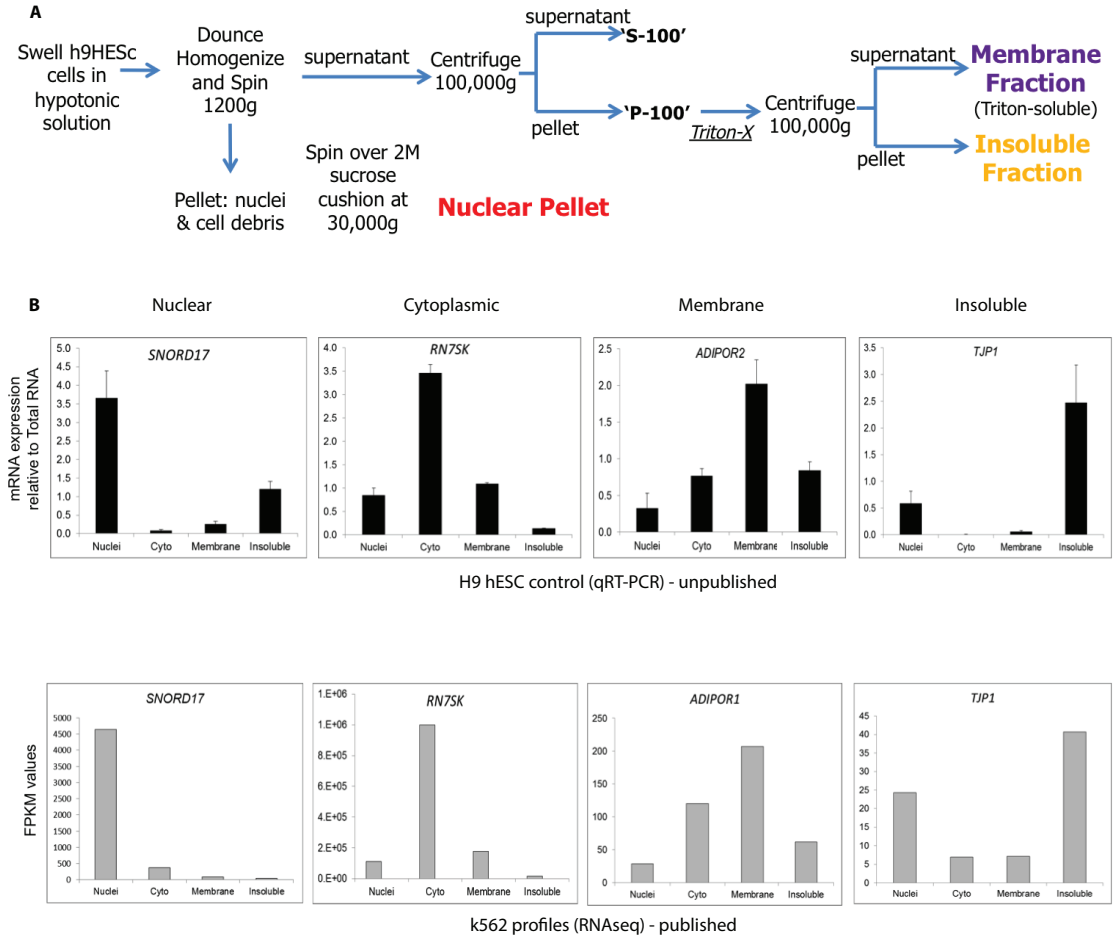

Figure 17. hESC fractionations validate well compared to k562 cells
(A) Protocol for isolation of all cellular fractions pre-sequencing.(B) qRT-PCR validation of previously known markers expressed in only one cellular organelle; IMP1 KD hESCs are very similar to the previously published k562 cells.

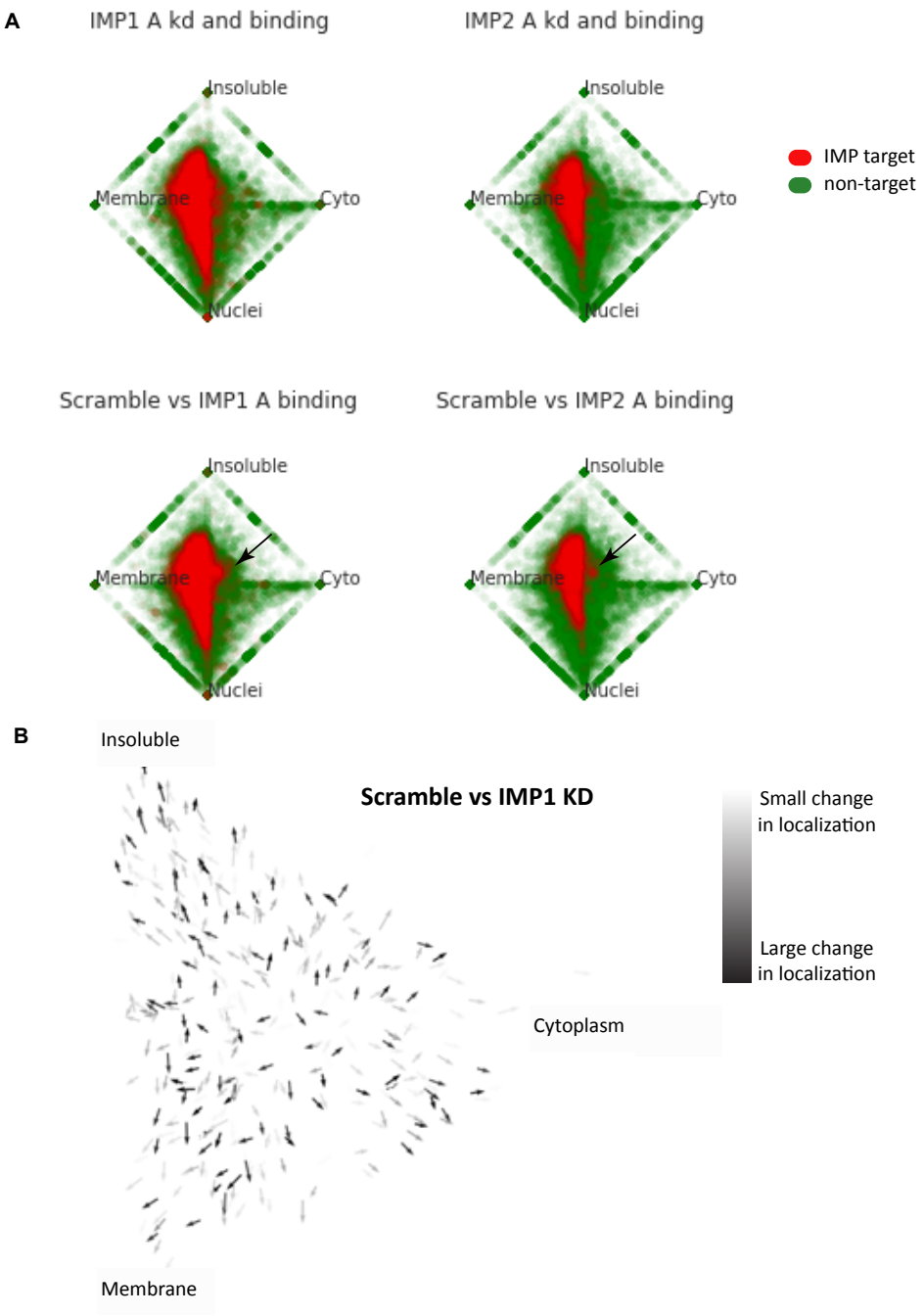

Figure 18. Moderate RNA localization changes with loss of IMP1 and IMP2 in hESC
(A) Simplex diagrams denoting localization changes with loss of IMP1 or IMP2 in hESCs. Red dots are CLIP-seq targets and green dots are not bound. Black arrows show a shift to the cytoplasm following loss of IMP. (B) Simplex diagram for only IMP1 KD cells compared to control, with each arrow denoting a single RNA target. Gradient scale shows level of change in localization.

CHAPTER 4 – A ROLE FOR IMPS IN DIRECTING CELL FATE

Introduction

My dissertation to this point has focused primarily on hESCs and a role for IMP proteins in the pluripotent state. I would like to remind the reader however that initial studies examining IMP family RBP expression during early development *in vivo* reported substantial evolutionarily conserved IMP1 expression in the developing nervous system. In addition, early functions for IMP proteins were discovered in polarized cells using neurons as a model system. While the initial studies focused on roles for IMP proteins in localization and translation, only recently did investigators specifically identify a role for IMP RBPs in determining murine neural cell fate (Nishino et al., 2013).

Analyzing expression of all three human family members in *in vitro* derived stem cells as well as progenitor cells (Figure 2B), *IMP1* was expressed at the highest levels compared to the other two family members. Remarkably though, *IMP3* was consistently expressed at higher levels in human fetal NPCs suggesting perhaps that the stage of differentiation differs between the two samples (Figure 2B)(Also see Patterson et al., 2012). All of the mammalian IMP family members have been implicated in playing a role in various types of brain tumors while not being associated with normal brain tissue (Janiszewska et al., 2012; Suvasini et al., 2011). While this is not necessarily surprising due to the number of other tumor types that these genes are upregulated in, their function in brain tumors has not been well studied. It appears that IMP family protein expression in brain tumors gives the

transformed cells a significant survival advantage since all of the family members are upregulated and implicated in this disease.

To test whether the IMP proteins have a role in human neural development *in vitro* we will use a neural differentiation paradigm developed by colleagues at the Salk Institute that generates primarily glutamatergic neurons, but also some GABAergic and dopaminergic neurons (Brennand et al., 2011). Due to the majority of reports in this area being about IMP1, will focus primarily on the function of this family member for the following studies.

Results

IMP1 expression is maintained during progenitor cell differentiation in vitro

To determine whether IMP1 has a role during early differentiation from the pluripotent state, I first performed an embryoid body (EB) differentiation assay to see whether it is expressed in spontaneously differentiating cells. Using reduction in expression of the pluripotency markers OCT4, NANOG and SOX2 (ONS) as a proxy to signify the cells were indeed differentiating, I could determine whether IMP1 expression is maintained. Indeed, IMP1 is expressed throughout the two week time course while ONS are depleted suggesting that it may have a role in progenitor cell populations (Figure 19A). Interestingly, IMP1 expression appears to increase in levels in the differentiating cells compared to expression in the pluripotent state. This was a very surprising result due to literature reports that IMP expression decreases with differentiation.

I next wanted to test whether IMP1 expression is maintained during directed differentiation to the neural lineage. To test this I assayed IMP protein and RNA expression from BJ fibroblasts, iPS cells derived from fibroblasts (BJ-iPS), Neural progenitor cells (NPCs), and 1, 2, 4, and 6 week differentiated neurons (Figure 19B). Interestingly, IMP protein expression was maintained in NPCs, but decreased upon induction of neural differentiation (Figure 19C). This expression pattern is consistent with published reports in that IMP1 protein is maintained only in early progenitor cells and then declines with differentiation (Nishino et al., 2013). To further define the role of IMP1 in early human neural development *in vitro*, I also assayed for expression of IMP1 at the mRNA level. In contrast to protein level expression there was an increase in the mRNA level from the pluripotent to the NPC state (Figure 19C). The increase in mRNA level without a concomitant increase in the protein level suggested that *IMP1* mRNA levels may be regulated post-transcriptionally.

A previous study had identified the IMP family RBPs as being mRNA targets of the *let-7* family of miRNAs (Boyerinas et al., 2008). Additionally, IMP1 was recently shown to be targeted by *let-7* during neural development *in vivo* (Nishino et al., 2013). To test whether *let-7* may be responsible for the incongruent pattern of IMP1 protein and RNA expression during neural development *in vitro*, I tested for *let-7a* expression during the same *in vitro* neural differentiation time course. Incredibly, *let-7a* displayed an expression pattern that was directly inverted compared to *IMP1* mRNA; that is to say that *let-7a* was expressed highest in the BJ fibroblasts and decreased with reprogramming, then increased again upon neural differentiation

(Figure 19C). The NPC time point is of interest because at this time point is when IMP1 RNA and protein levels are no longer correlated and it is the time point that *let-7a* begins to increase. Although the validation experiments to show that it is a direct interaction have not yet been completed, it is very likely that *IMP1* becomes targeted for degradation by *let-7* family members in *in vitro* derived human NPCs upon upregulation of *let-7a*.

IMP1 may affect cell fate decisions in in vitro derived NPCs

There is precedent set in the literature for RNA binding proteins having a role in directing cell fate decisions and in the case of human neural development, one of the RBPs to do this is LIN28. LIN28 has a *let-7* independent role in promoting neurogenesis over gliogenesis during human neural development *in vitro* (Balzer et al., 2010). To test the function of IMP1 on directing cell fate decisions in NPCs, I performed qRT-PCR for neural lineage markers in NPCs with gain or loss of IMP1. Preliminary data suggest that IMP1 does indeed have the same affect as LIN28 in promoting neurogenesis and repressing gliogenesis *in vitro* (Figure 19D). While these data are preliminary and need to be repeated, it provides further evidence for the LIN28 and IMP families of RBPs to have similar and/or overlapping functions during development.

IMP1 promotes NPC survival

Upon performing the KD and OE qPCR experiments one of the phenotypes that was quite apparent was that there were fewer NPCs in the KD/OE conditions compared to control (Figure 19E). To test whether some of the phenotypes we see for IMP1 KD in hESC hold true for NPCs, we set out to determine whether IMP1 KD NPCs are undergoing apoptosis. Increased cleaved caspase 3 expression in the IMP1 KD NPCs suggest that this is indeed the case (Figure 19F). Given that we saw upregulation of differentiation markers with both KD and OE it is possible that the cells are differentiating in both cases. Further experiments examining cell cycle progression with BRDU would determine whether the cells are indeed exiting the cell cycle or undergoing arrest. Taken together, these data provide additional evidence that *let-7* gene targets, and more specifically IMP1, regulate neural progenitor cell function and that the regulation of these targets at different stages by *let-7* contributes to alterations in stem cell homeostasis.

Discussion

These preliminary experiments in NPCs suggest that not only do the IMP proteins have a role maintaining the pluripotent state, but they may also have a role in progenitor cell populations. This begs the question if IMP proteins are expressed NPCs, are they also expressed in other progenitor cell populations during *in vitro* differentiation from hESCs? To test this I performed directed differentiation to each of the three embryonic germ layers using defined medium and growth factors, then assayed IMP1 protein expression by WB. Intriguingly, IMP1 expression was induced

in all 3 conditions! This data suggests that IMP1 and likely IMP2 and IMP3 also have specific roles in progenitor cell populations and not just the pluripotent state. Future experiments in this area could identify roles for binding and regulating RNA targets throughout the stages of differentiation.

Additional questions raised by these results are whether IMP family RBPs have a general or "housekeeping" affect on the transcriptome in highly proliferative cells and whether they remain targeted to bound RNAs as the cell cycle slows down during differentiation. To address the first question, it's true that these proteins are generally expressed in cells with high proliferation rates during embryogenesis and tumorigenesis, but they are not expressed in all proliferative cell types. For example, they are not endogenously expressed in all cancer cell lines such as MCF7 breast cancer cells and U2OS osteosarcoma cells. There have been few studies addressing the second question, that is whether IMP RBPs target the same target RNAs throughout embryonic development or differentiation *in vitro*. One way to test this is by performing CLIP-seq on cells during a directed differentiation protocol where endogenous IMP protein is maintained. A tricky aspect to this experiment would be finding the correct time points where the cells have initiated terminal differentiation, but IMP protein expression is maintained. The P19 EC model of *in vitro* neural differentiation might be a good model system for this type of experiment because components of IMP RNPs are already known at various stages of differentiation, but there has not yet been a comprehensive analysis of bound RNA targets. Future

experiments will elucidate the mechanisms of how IMP family-associated RNPs are regulated throughout development.

Methods

Neural Differentiation paradigm

hiPSCs grown in HUES media on MEFs were incubated with Collagenase (1 mg/ml in DMEM) at 37°C for one to two hours until colonies lifted from the plate and were transferred to a nonadherent plate (Corning). Embryoid Bodies were grown in suspension in N2 media (DMEM/F12-Glutamax (Invitrogen), 1x N2 (Invitrogen)). After seven days, EBs were plated in N2 media with 1 µg/ml Laminin (Invitrogen) onto polyornithine (PORN)/Laminin-coated plates. Visible rosettes formed within one week and were manually dissected onto PORN/Laminin-coated plates. Rosettes were cultured in NPC media (DMEM/F12, 1x N2, 1x B27-RA (Invitrogen), 1 µg/ml Laminin and 20 ng/ml FGF2) and dissociated in TrypLE (Invitrogen) for three minutes at 37°C. NPCs are maintained at high density, grown on PORN/Laminin-coated plates in NPC media and split approximately 1:4 every week with Accutase (Millipore). For neural differentiations, NPCs were dissociated with Accutase and plated in neural differentiation media (DMEM/F12, 1x N2, 1X B27-RA, 20 ng/ml BDNF (Peprotech), 20 ng/ml GDNF (Peprotech), 1 mm dibutyrl-cyclicAMP (Sigma), 200 nm ascorbic acid (Sigma) onto PORN/Laminin-coated plates. Density is critical for neural differentiaitons and six well plates were split 200,000 cells/well. hiPSC derived-neurons were differentiated for 1–6 weeks.

Western Blot

Cells were washed with PBS and lysed with 10 mM Tris-HCl (pH 8), 150 mM NaCl, 1% Triton X100 and complete protease inhibitor mixture (Roche). Total protein extracts were used for SDS-PAGE, transferred to nitrocellulose membranes (Amersham Biosciences) and analyzed using primary antibodies. Primary antibodies were incubated overnight at 4 degrees and secondary HRP conjugated antibodies (Jackson ImmunoResearch, 1:10,000) were incubated for 1 hour at room temp. Thermo Pierce ECL detection reagents were used. Antibodies used: anti-OCT4 (Santa Cruz, #sc-5279), 1:1000; anti-IMP1 (Cell Signaling, #2852), 1:1000; anti-NANOG (Cell Signaling, #4903S) 1:1000; anti-SOX2 (Cell Signaling, #3579S) 1:500; anti-GAPDH (Abcam, #ab8245) 1:10,000; Cleaved-Caspase 3 (Cell Signaling, #9661) 1:500.

RNA extraction and real-time qPCR analysis

Total RNA was isolated using Trizol Reagent (Invitrogen) according to the manufacturer's recommendations, and cDNA synthesized using the SuperScript III Reverse Transcriptase kit for RT-PCR (Invitrogen). Real-time PCR was performed using the SYBR-Green FAST qPCR Master mix (Applied Biosystems). Values of gene expression were normalized using 18s and/or GAPDH (see figure legends) expression and are shown as fold change relative to the value of the sample control. All the samples were done in technical and biological triplicates. Taqman primers Pax6, #Hs00240871_m1, Olig2, #Hs00377820_m1, NeuroD1, #Hs00159598_m1,

IMP1 #Hs00198023_m1, SOX2 #Hs01053049_s1, GAPDH #4333764F. All samples were normalized to the Untreated control.

Acknowledgements

I would like to thank Leah Boyer and Kristen Brennand of the Gage laboratory and members of the Salk stem cell core staff for their assistance with the neural differentiation protocol and getting me up and running when I first started working there. I would also like to thank Kristen for initially providing me with some of her extra samples to determine whether IMP1 was even expressed during *in vitro* neural differentiation.

Figures

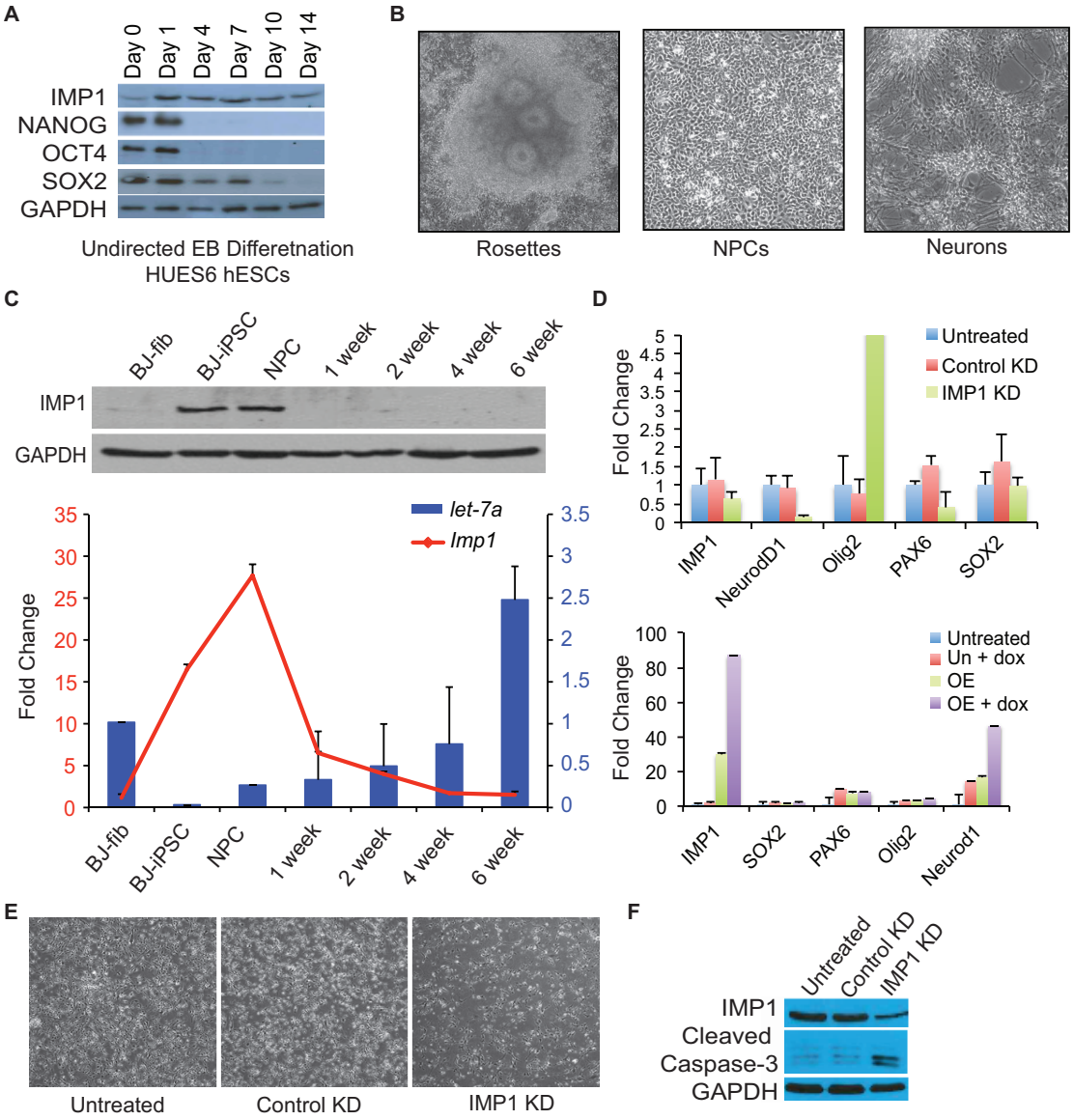

Figure 19. IMP1 is required for neural progenitor cell survival
(A) Western Blot for IMP1 during undirected Embryoid Body differentiation (EB). (B) Phase contrast images of cells from neural differentiation time course. (C) Western blot and qRT-PCR for IMP1 protien and RNA levels, respectively, during *in vitro* neural differentiation assay. (D) qRT-PCR for neural differentiation markers with loss or gain of IMP1. (E) Phase contrast images of NPCs following loss of IMP1. (F) Western blot for Cleaved-caspase 3 in IMP1 KD NPCs.

CHAPTER 5 - CLOSING REMARKS

Yet-to-be-solved Mysteries

Redundancy between IMP family members

There are many examples in the literature and in my work where IMP family members acting redundantly could explain a phenotype or an aberrant result. For example, I tried in earnest to try and knock down both IMP1 and IMP2 expression at the same time in hESCs using lentiviral shRNA vectors for the localization-seq experiments, but unfortunately I couldn't get it to work. IMP1 protein would be depleted just fine, but IMP2 remained expressed, and if anything slightly increased (!!) even though the IMP2 shRNA alone worked just fine (Figure 20B). With the recent discovery of the CRISPR/Cas9 genome editing system these experiments are now much more feasible. Also, additional evidence from IMP2 and IMP3 knockout mice as well as from double and triple knockout mice could further clarify these issues. COME ON, will someone make these mice already?!?!

Secondly, I think that these types of questions may be best addressed at the single cell level. The development of single cell RNAs-sequencing and ongoing improvements to this method will better allow for investigations of redundancy and heterogeneity within progenitor cell populations. Preliminary data from the lab suggest that there are some interesting relationships between IMP RBP family members throughout neural differentiation *in vitro* (Figures 21, 22). Future experiments and validations in this area will surely lead to exciting discoveries on how the IMP RBPs regulate neural differentiation.

IMP1 and somatic cell reprogramming

One of the initial aims I set out to conquer when I first started my PhD was to elucidate the function of IMP1 in somatic cell reprogramming. I was very excited about this aim because it was the first IMP1 KD functional experiment I got to work! And it made sense! (Or so I thought….). The rationale behind this set of experiments is as follows: *MYC* was one of the first described RNA targets for IMP1/CRD-BP (Bernstein et al., 1992) and *MYC* is also an important factor for somatic cell reprogramming (Sridharan et al., 2009); based on the model for IMP1 regulation of *MYC* in the literature, loss of IMP1 should lead to loss of *MYC* and then a reduction in somatic cell reprogramming. My results showed a loss of reprogramming efficiency, but not necessarily a reduction of *MYC* expression. Furthermore, when I over-expressed IMP1 I didn't see any colonies at all. I don't doubt that there is a role for IMP1 in the somatic cell reprogramming process, but from what I know now it is likely to be very specific and act in a temporal manner. Around the time I was performing these experiments it was suggested that I "focus" my attention to one project so that I would finish my PhD in a reasonable amount of time and I never got around to pursuing this question any further. If there is a beginning PhD student out there reading this, I think this would be an interesting question to investigate!!

IMP2 and Diabetes

Perhaps one of the most interesting things about the IMP RBPs is their expression pattern. It is very specific in that they are expressed during embryogenesis

and generally not in adult tissues with the exception of the gonads. However, there are more and more reports coming out that IMP2 is unique and may be expressed at the protein level in more adult tissues than was originally thought (See Bell, 2013 for a review). An additional very unique trait exclusive to IMP2 is the association of IMP2 with Type 2 Diabetes (Christiansen et al., 2009 for review). From our work performing CLIP-seq in hESCs we know that IMP1 and IMP2 largely overlap in the set of targets they bind, however, there must be either a specific set of targets or specific type of regulation that permits IMP2 to have this special function. One such function could be the mechanism by which IMP2 regulates its target RNAs; IMP2 bound its target RNAs almost exclusively in the 3' UTR. 3' UTR shortening or alternative polyadenylation could impact normal regulation by IMP2 on its target RNAs involved in metabolism and later on, trigger diabetes. As the set of target RNAs IMP2 binds in different conditions becomes more established and mechanisms are developed for how IMP2 regulates its target RNAs, I propose the function of IMP2 in Type 2 diabetes should be a low hanging fruit and high priority problem to investigate because solving it could lead to a large impact on human health.

IMPs and LIN28s

It is now becoming commonly known that families of RBPs bind and can regulate each other's RNAs. This is certainly the case for IMP RBPs and it is as well for another family of RBPs, LIN28A and LIN28B. The LIN28s have a similar oncofetal expression pattern as the IMPs and as I mentioned previously they are also

involved in early neuronal development *in vitro* (Balzer et al., 2010). Furthermore, LIN28 acts upstream of the *let-7* family of miRNAs and represses processing. Therefore, the model would be that as LIN28 expression increases, *let-7* is reduced and then IMP1 (and/or any other heterochronic or "LOG" gene regulated by *let-7*) becomes upregulated. Because of these connections our hypothesis was that loss of IMP1 and/or loss of LIN28 would lead to depletion of the other RBP. This did not turn out to be the case, at least in hESCs. Expression at the protein level remained unchanged. However, we were able to identify a large set of overlapping CLIP-seq target RNAs shared between LIN28A and the IMPs suggesting that they at least regulate a similar network of target genes and therefore may have a similar function (Figure 2H). If I had more time, I would want to investigate the inter-relationship between these families of RBPs and how manipulation of the LIN28—*let-7*—IMP feedback loop could alleviate disease.

Some notes to biologists on collaborations with bioinformaticians (and vice-versa)

I was very fortunate to collaborate with some great bioinformaticians during the latter few years of my PhD and wanted to share some lessons I learned while working to analyze high-throughput sequencing datasets. First of all, bioinformatics is a whole different language- learn it! Just as science itself constitutes its own language and culture, so does bioinformatics. One issue that came up time and time again while analyzing our datasets was different iterations (versions) of gene annotations. Turns out, those gene diagrams visible on the genome browser that look to be set in stone aren't so, it is very much still a work in progress. By learning the language of

bioinformatics and understanding how these things can vary, a biologist can get a lot more out of their data. Second most important lesson- don't be afraid to ask questions! Even though you may not know exactly how to articulate your question in bioinformatics-speak, keep trying until you get an answer. Often times there are multiple ways to do things, frequently within a field or for a particular type of dataset and even sometimes within a single lab! It is important to understand how your dataset is analyzed to the best of your ability, and what the caveats/alternatives were to the particular type of analysis used, especially if you're not the one doing it yourself. Lastly, communication is key! Just like with wet lab experiments there are sometimes assumptions made for one type of analysis that don't stand true for another (or even just a newer version of the same analysis) and the end result is trying to jam a square peg in a round hole. Be sure to ask your collaborator to walk you through every.single.step. of the analysis pipeline to be sure the correct assumptions are being made for your experiment. All that being said, this information is likely to become antiquated pretty quick because biologists and bioinformaticians of the future are likely to become one and the same.

Conclusion

Throughout my PhD journey I've developed from a single-gene model scientist to thinking about networks and "omics" within cells. This is likely a sign of the times and also a nod to the direction academic science is moving. Unfortunately the NIH funding situation for basic research is moving the opposite direction and is decreasing year upon year. This downward trend of public funding is accompanied by

more industry-academic collaborations as well as an influence of private donations and philanthropy, partly out of necessity to keep high-caliber research programs prospering in the United States. A side effect (be it for better or worse) of funding academic labs with private money is that there is an increasing emphasis on translational disease-oriented research. While this is certainly a noble cause, I'm concerned that the free-will of scientific exploration will disappear if the end goal of any proposal is only ever to "find a cure". It is thus imperative to preserve this naivety for future generations of scientists. My concerns now stated, there hasn't ever been a more exciting time to do biomedical research. During the last few years I've been in grad school the $1000 genome has arrived, the prospect of gene therapy is back with a vengeance with the advent of CRISPR/Cas9, and induced pluripotent stem cells have made it a possibility to do large scale syngenic cell transplants. I look to the future with a cautious eye, but am excited to see what it will bring!

Methods

Western Blot

Cells were washed with PBS and lysed with 10 mM Tris-HCl (pH 8), 150 mM NaCl, 1% Triton X100 and complete protease inhibitor mixture (Roche). Total protein extracts were used for SDS-PAGE, transferred to nitrocellulose membranes (Amersham Biosciences) and analyzed using primary antibodies. Primary antibodies were incubated overnight at 4 degrees and secondary HRP conjugated antibodies (Jackson ImmunoResearch, 1:10,000) were incubated for 1 hour at room temp. Thermo Pierce ECL detection reagents were used. Antibodies used: anti-IMP1 (Cell

Signaling, #2852), 1:1000; rabbit anti-IMP2 (MBL, #RN008P), 1:1000; and rabbit anti-γ-TUBULIN (Sigma, T6557) 1:10,000.

Motor Neural Differentiation Protocol

Human iPSCs were differentiated according to a modified version of Chambers *et al*. 2009 and Burkhardt *et al*. 2013. Dorsomorphin dihydrochloride (Tocris) at 1 μM was used as SMAD inhibitor instead of recombinant hNOGGIN, cells were dissociated at day 18 instead of day 11 and maturation of motor neurons was carried out in Maturation Media: d-MEM/F12 + GLUTAMAX, 2% N-2 Supplement, 4% B-27 Serum-Free Supplement (Invitrogen), 9.0 mM d-Glucose, 0.1 mM Ascorbic Acid in addition to 2 nG/mL each of Ciliary Neurotrophic Factor (CNTF), Brain-Derived Neurotrophic Factor (BDNF), and Glial Cell-Derived Neurotrophic Factor (GDNF). The neurons were ventralized with 200 nM Smoothened Agonist (EMD Biosciences) instead of recombinant SHH. Following plating, cells were maintained in Maturation Media and 1.5 μM Retinoic Acid, 200 nM Smoothened Agonist, 2 μM DAPT for a period of 4 days. Media was changed daily.

Single Cell Sequencing Analysis

RNA sequencing reads were generated from cDNA libraries of individual cells from one of three types: induced pluripotent stem cells, neural progenitor cells (NPC), and motor neurons (MN). 48 cells were sequenced on an Illumina HiSeq 2500 flow cell for an average of 10-20 million reads per cell. Standard quality control metrics

were applied (pass filter, Q<20 bases masked)

The paired-end 100 nt RNA-seq reads were mapped to the reference human transcriptome (hg19 annotated with GENCODEv18) using RNA STAR with default settings (allowing up to 10 mismatches). The log expression of each gene *g* was computed as follows: $\log(R_{g,c})$ where *c* is the cell and $R_{g,c}$ is the number of reads per kilobase per million mapped (RPKM value, depicted by shades of blue in Figure 22) obtained from counting the reads mapped to gene *g*. The distribution of RPKM values for the cells in each population was tabulated in horizontal "violin" plots (depicted in black on Figure 21).

Acknowledgements

I would like to thank Boyko Kakaradov for the analysis of the IMP RBPs in the single cell differentiation datasets.

Figures

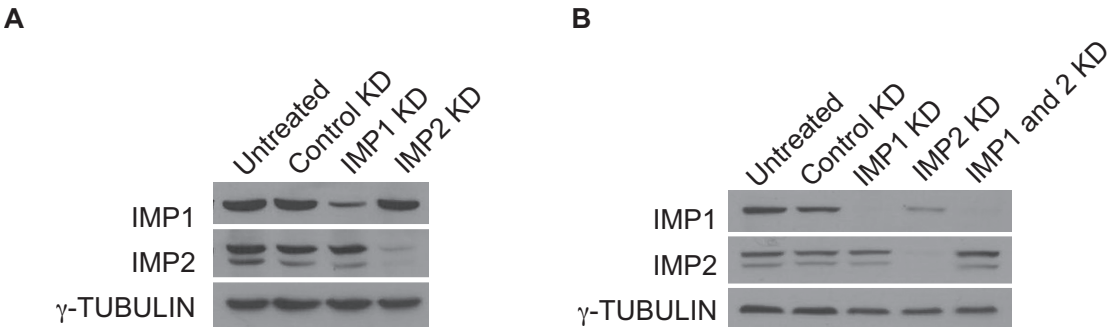

Figure 20. IMP family members may act redundantly in hESCs
(A) Western blot for IMP1 and IMP2 in localization-sequencing samples pre-fractionation. (B) Test for double knock-down of IMP1 and IMP2 in H9 hESCs.

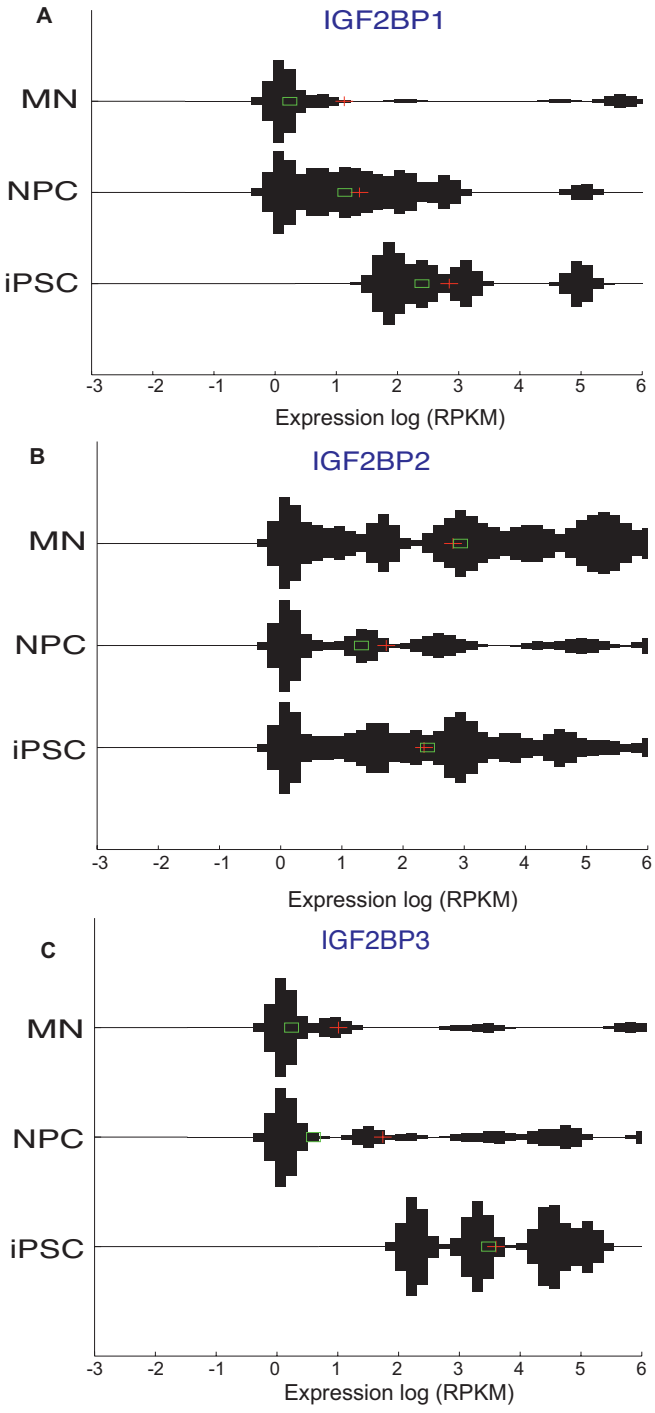

Figure 21. IMP family expression at the single cell level during neural differentiation
Analysis of single cell gene expression during motor neuron differentiation. Each black square represents values from a single cell. Graphs display expression pattern of IMP1 (A), IMP2 (B), and IMP3 (C) at the three stages of the *in vitro* differentiation time course.

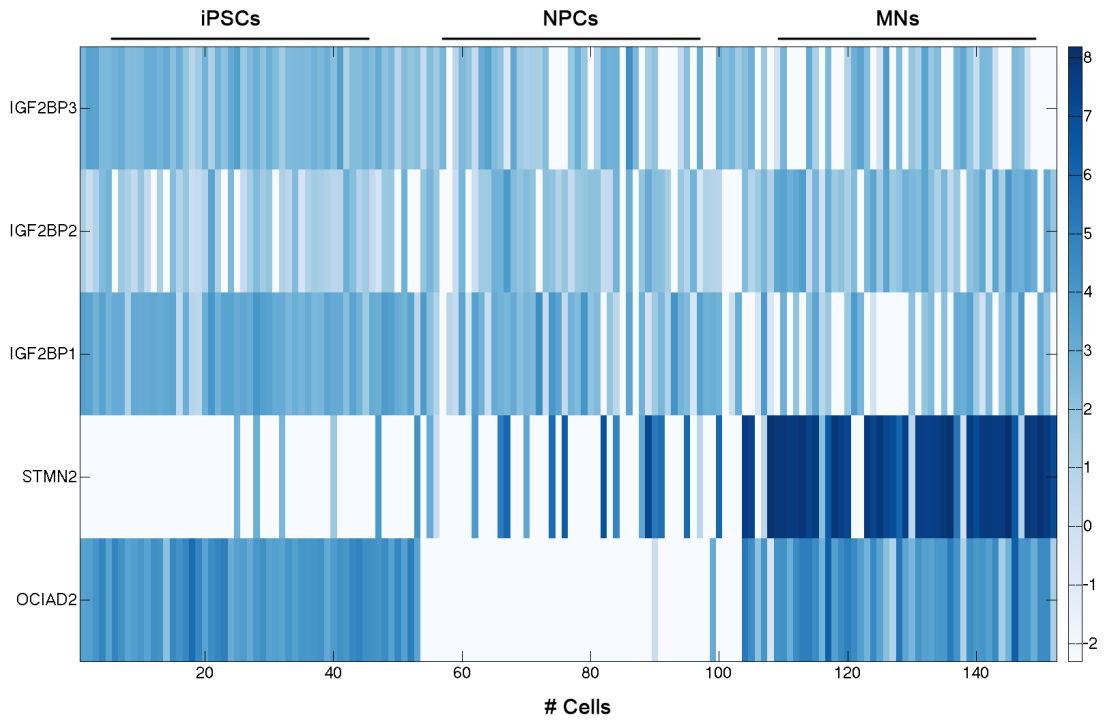

Figure 22. IMP family members may act redundantly at the single cell level
Analysis of single cell gene expression levels during motor neuron differentiation *in vitro*. Each vertical line represents gene expression levels from a single cell.

REFERENCES

Abbott AL, Alvarez-Saavedra E, Miska EA, Lau NC, Bartel DP, Horvitz HR, Ambros V. The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. Dev Cell. 2005 Sep;9(3):403-14.

Alajez NM, Shi W, Wong D, Lenarduzzi M, Waldron J, Weinreb I, Liu FF. Lin28b promotes head and neck cancer progression via modulation of the insulin-like growth factor survival pathway. Oncotarget. 2012 Dec;3(12):1641-52.

Ardehali, R.I., M.A.; Ali, S.R.; Tang, C.; Drukker, M.; Weissman, I.L (2011). Overexpression of BCL2 enhances survival of human embryonic stem cells during stress and obviates the requirement for serum factors. PNAS *108*, 3282-3287.

Atlas, R., Behar, L., Elliott, E., and Ginzburg, I. (2004). The insulin-like growth factor mRNA binding-protein IMP-1 and the Ras-regulatory protein G3BP associate with tau mRNA and HuD protein in differentiated P19 neuronal cells. Journal of neurochemistry *89*, 613-626.

Atlas, R., Behar, L., Sapoznik, S., and Ginzburg, I. (2007). Dynamic association with polysomes during P19 neuronal differentiation and an untranslated-region-dependent translation regulation of the tau mRNA by the tau mRNA-associated proteins IMP1, HuD, and G3BP1. Journal of neuroscience research *85*, 173-183.

Balzer, E., Heine, C., Jiang, Q., Lee, V.M., and Moss, E.G. (2010). LIN28 alters cell fate succession and acts independently of the let-7 microRNA during neurogliogenesis in vitro. Development *137*, 891-900.

Bell, J.L., Wachter, K., Muhleck, B., Pazaitis, N., Kohn, M., Lederer, M., and Huttelmaier, S. (2013). Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? Cellular and molecular life sciences : CMLS *70*, 2657-2675.

Bernstein, P.L., Herrick, D.J., Prokipcak, R.D., and Ross, J. (1992). Control of c-myc mRNA half-life in vitro by a protein capable of binding to a coding region stability determinant. Genes & development *6*, 642-654.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., and Young, R.A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyerinas, B., Park, S.M., Shomron, N., Hedegaard, M.M., Vinther, J., Andersen, J.S., Feig, C., Xu, J., Burge, C.B., and Peter, M.E. (2008). Identification of let-7-regulated oncofetal genes. Cancer research *68*, 2587-2591.

Boylan, K.L., Mische, S., Li, M., Marques, G., Morin, X., Chia, W., and Hays, T.S. (2008). Motility screen identifies Drosophila IGF-II mRNA-binding protein--zipcode-binding protein acting in oogenesis and synaptogenesis. PLoS genetics *4*, e36.
Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., McCarthy, S., Sebat, J., and Gage, F.H. (2011). Modelling schizophrenia using human induced pluripotent stem cells. Nature *473*, 221-225.

Braam, S.R., Zeinstra, L., Litjens, S., Ward-van Oostwaard, D., van den Brink, S., van Laake, L., Lebrin, F., Kats, P., Hochstenbach, R., Passier, R., Sonnenberg, A., and Mummery, C.L. (2008). Recombinant vitronectin is a functionally defined substrate that supports human embryonic stem cell self-renewal via alphavbeta5 integrin. Stem cells *26*, 2257-2265.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209-1222.

Chao, J.A., Patskovsky, Y., Patel, V., Levy, M., Almo, S.C., and Singer, R.H. (2010). ZBP1 recognition of beta-actin zipcode induces RNA looping. Genes & development *24*, 148-158.

Chia, N.Y., Chan, Y.S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.S., Huss, M., Soh, S.S., Kraus, P., Li, P., Lufkin, T., Lim, B., Clarke, N., Bard, F., and Ng, H.H. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316-320.

Christiansen J, Kolte AM, Hansen Tv, Nielsen FC. IGF2 mRNA-binding protein 2: biological function and putative role in type 2 diabetes. 2009. J Mol Endocrinol. Nov;43(5):187-95.

Coffee, B., Zhang, F., Warren, S. and Reines, D. Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. Nat Genet, 1999. 22(1): p. 98-101.

Coulis CM, Lee C, Nardone V, Prokipcak RD. Inhibition of c-myc expression in cells by targeting an RNA-protein interaction using antisense oligonucleotides. Mol Pharmacol. 2000 Mar;57(3):485-94.

Dai, N., Christiansen, J., Nielsen, F.C., and Avruch, J. (2013). mTOR complex 2 phosphorylates IMP1 cotranslationally to promote IGF2 production and the proliferation of mouse embryonic fibroblasts. Genes & development *27*, 301-312.

Dai, N., Rapley, J., Angel, M., Yanik, M.F., Blower, M.D., and Avruch, J. (2011). mTOR phosphorylates IMP2 to promote IGF2 mRNA translation by internal ribosomal entry. Genes & development *25*, 1159-1172.

Deshler, J.O., Highett, Martin I., Abramson, T., Schnapp, B.J. (1998). A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates. Current Biology *8*, 489-496.

Dimitriadis, E., Trangas, T., Milatos, S., Foukas, P.G., Gioulbasanis, I., Courtis, N., Nielsen, F.C., Pandis, N., Dafni, U., Bardi, G., and Ioannidis, P. (2007). Expression of oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer. International journal of cancer Journal international du cancer *121*, 486-494.

Elisha Z, Havin L, Ringel I, and Yisraeli JK. Vg1 RNA binding protein mediates the association of Vg1 RNA with microtubules in Xenopus oocytes. EMBO J. 1995 Oct 16;14(20):5109-14.

Fallini, C., Rouanet, J.P., Donlin-Asp, P.G., Guo, P., Zhang, H., Singer, R.H., Rossoll, W., and Bassell, G.J. (2013). Dynamics of survival of motor neuron (SMN) protein interaction with the mRNA-binding protein IMP1 facilitates its trafficking into motor neuron axons. Developmental neurobiology.

Farina, K.L., Huttelmaier, S., Musunuru, K., Darnell, R., and Singer, R.H. (2003). Two ZBP1 KH domains facilitate beta-actin mRNA localization, granule formation, and cytoskeletal attachment. The Journal of cell biology *160*, 77-87.
Frixen, U.H.B., J.; Sachs, M.; Eberle, G.; Voss, B.; Warda, A.; Lochner, D.; Birchmeir, W. (1991). E-cadherin-mediated Cell-Cell Adhesion Prevents Invasiveness of Human Carcinoma Cells The Journal of cell biology *113*, 173-185.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., Nedelec, S., Wichterle, H., Woltjen, K., Hughes, T.R., Zandstra, P.W., Nagy, A., Wrana, J.L., and Blencowe, B.J. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell *147*, 132-146.

Git, A., Standart, N. The KH domains of Xenopus Vg1RBP mediate RNA binding and self-association. RNA. 2002 Oct;8(10):1319-33.

Gu, L.S., K.; Ohama, K. (2004). Increased expression of IGF II mRNA-binding protein 1 mRNA is associated with an advanced clinical stage and poor prognosis in patients with ovarian cancer. International Journal of Oncology *24*, 671-678.

Gu, W., Wells, A.L., Pan, F., and Singer, R.H. (2008). Feedback regulation between zipcode binding protein 1 and beta-catenin mRNAs in breast cancer cells. Mol Cell Biol *28*, 4963-4974.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., Ulrich, A., Wardle, G.S., Dewell, S., Zavolan, M., and Tuschl, T. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Hafner, M., Max, K.E., Bandaru, P., Morozov, P., Gerstberger, S., Brown, M., Molina, H., and Tuschl, T. (2013). Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. Rna *19*, 613-626.

Hammer, N.A., Hansen, T., Byskov, A.G., Rajpert-De Meyts, E., Grondahl, M.L., Bredkjaer, H.E., Wewer, U.M., Christiansen, J., and Nielsen, F.C. (2005). Expression of IGF-II mRNA-binding proteins (IMPs) in gonads and testicular cancer. Reproduction *130*, 203-212.

Hammerle, M., Gutschner, T., Uckelmann, H., Ozgur, S., Fiskin, E., Gross, M., Skawran, B., Geffers, R., Longerich, T., Breuhahn, K., Schirmacher, P., Stoecklin, G., and Diederichs, S. (2013). Posttranscriptional destabilization of the liver-specific long noncoding RNA HULC by the IGF2 mRNA-binding protein 1 (IGF2BP1). Hepatology *58*, 1703-1712.

Han, H., Irimia, M., Ross, P.J., Sung, H.K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., Wang, E., Trcka, D., Thompson, T., O'Hanlon, D., Slobodeniuc, V., Barbosa-Morais, N.L., Burge, C.B., Moffat, J., Frey, B.J., Nagy, A., Ellis, J., Wrana, J.L., and Blencowe, B.J. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241-245.
Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hansen, T.V.O., Hammer, N.A., Nielsen, J., Madsen, M., Dalbaeck, C., Wewer, U.M., Christiansen, J., and Nielsen, F.C. (2004). Dwarfism and Impaired Gut Development in Insulin-Like Growth Factor II mRNA-Binding Protein 1-Deficient Mice. Molecular and Cellular Biology *24*, 4448-4464.

Havin, L., Git, A., Elisha, Z., Oberman, F., Yaniv, K., Schwartz, S.P., Standart, N., and Yisraeli, J.K. (1998). RNA-binding protein conserved in both microtubule- and microfilament-based RNA localization. Genes & development *12*, 1593-1598.

Hu, S., Wu, X., Zhou, B., Xu, Z., Qin, J., Lu, H., Lv, L., Gao, Y., Deng, L., Yin, J., and Li, G. (2014). IMP3 combined with CD44s, a novel predictor for prognosis of patients with hepatocellular carcinoma. Journal of cancer research and clinical oncology *140*, 883-893.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., Ares, M., Jr., and Yeo, G.W. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell reports *1*, 167-178.

Huttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G.J., Condeelis, J., and Singer, R.H. (2005). Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. Nature *438*, 512-515.

Ioannidis P, Mahaira L, Papadopoulou A, Teixeira MR, Heim S, Andersen JA, Evangelou E, Dafni U, Pandis N, Trangas T. 2003. 8q24 Copy number gains and expression of the c-myc mRNA stabilizing protein CRD-BP in primary breast carcinomas. Int J Cancer. Mar 10;104(1):54-9.

Ioannidis P, Kottaridi C, Dimitriadis E, Courtis N, Mahaira L, Talieri M, Giannopoulos A, Iliadis K, Papaioannou D, Nasioulas G, Trangas T. 2004. Expression of the RNA-binding protein CRD-BP in brain and non-small cell lung tumors. Cancer Lett. Jun 25;209(2):245-50.

Ioannidis, P., Mahaira, L.G., Perez, S.A., Gritzapis, A.D., Sotiropoulou, P.A., Kavalakis, G.J., Antsaklis, A.I., Baxevanis, C.N., and Papamichail, M. (2005). CRD-BP/IMP1 expression characterizes cord blood CD34+ stem cells and affects c-myc and IGF-II expression in MCF-7 cancer cells. The Journal of biological chemistry *280*, 20086-20093.

Janiszewska, M., Suva, M.L., Riggi, N., Houtkooper, R.H., Auwerx, J., Clement-Schatlo, V., Radovanovic, I., Rheinbay, E., Provero, P., and Stamenkovic, I. (2012). Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells. Genes & development *26*, 1926-1944.

Jonson, L., Vikesaa, J., Krogh, A., Nielsen, L.K., Hansen, T. vO., Borup, R., Johnsen, A., Christiansen, J., Nielsen, F.C. (2007). Molecular Composition of IMP1 Ribonucleoprotein Granules. Molecular and Cellular Proteomics *6*, 798-811.

Kim, H.J., Kim, G.E., Lee, J.S., Lee, J.H., Nam, J.H., and Choi, C. (2014). Insulin-like growth factor-II mRNA-binding protein 3 expression in effusion cytology: a marker for metastatic adenocarcinoma cells and a potential prognostic indicator in gastric adenocarcinoma. Acta cytologica *58*, 167-173.

King, D.T., Barnes, M., Thomsen, D., and Lee, C.H. (2014). Assessing specific oligonucleotides and small molecule antibiotics for the ability to inhibit the CRD-BP-CD44 RNA interaction. PloS one *9*, e91585.

Kislauskis EH, Li Z, Singer RH, Taneja KL. Isoform-specific 3'-untranslated sequences sort alpha-cardiac and beta-cytoplasmic actin messenger RNAs to different cytoplasmic compartments. J Cell Biol. 1993 Oct;123(1):165-72.

Kislauskis EH, Zhu X, Singer RH. Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. J Cell Biol. 1994 Oct;127(2):441-51.

Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP-transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. J Vis Exp. 2011 Apr 30;(50). pii: 2638. doi: 10.3791/2638.

Kremer, E.J., Pritchard M., Lynch M., Yu S., Holman K., Baker E., Warren S.T., Schlessinger D., Sutherland G.R., and Richards R.I. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science, 1991. 252(5013): p. 1711-4.

Kretz, M., Webster, D.E., Flockhart, R.J., Lee, C.S., Zehnder, A., Lopez-Pajares, V., Qu, K., Zheng, G.X., Chow, J., Kim, G.E., Rinn, J.L., Chang, H.Y., Siprashvili, Z., and Khavari, P.A. (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. Genes & development *26*, 338-343.

Latham VM Jr, Kislauskis EH, Singer RH, Ross AF. Beta-actin mRNA localization is regulated by signal transduction mechanisms. J Cell Biol. 1994 Sep;126(5):1211-9.

Leeds, P.K., B; Boylan, J; Betz, N; Steer, C; Gruppuso, P; Ross, J (1997). Developmental regulation of CRD-BP, an RNA-binding protein that stabilizes c-myc mRNA in vitro. Oncogene *14*, 1279-1286.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frézal, J., Cohen, D., Weissenbach, J., Munnich, A. and Melki, J. Identification and characterization of a spinal muscular atrophy- determining gene. Cell, 1995. 80(1): p. 155-65.

Lewis, T.L., Jr., Mao, T., Svoboda, K., and Arnold, D.B. (2009). Myosin-dependent targeting of transmembrane proteins to neuronal dendrites. Nature neuroscience *12*, 568-576.

Li, L., Bennett, S.A., and Wang, L. (2012). Role of E-cadherin and other cell adhesion molecules in survival and differentiation of human pluripotent stem cells. Cell adhesion & migration *6*, 59-70.

Mahaira, L.G., Katsara, O., Pappou, E., Iliopoulou, E., Fortis, S., Antsaklis, A., Fotinopoulos, P., Baxevanis, C.N., Papamichail, M., and Perez, S.A. (2014). IGF2BP1 expression in human mesenchymal stem cells significantly affects their proliferation and is under the epigenetic control of TET1/2 demethylases. Stem cells and development.

Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M. and Thornton, C. Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. Science, 2000. 289(5485): p. 1769-73.

Miller TM, Pestronk A, David W, Rothstein J, Simpson E, Appel SH, Andres PL, Mahoney K, Allred P, Alexander K, Ostrow LW, Schoenfeld D, Macklin EA, Norris DA, Manousakis G, Crisp M, Smith R, Bennett CF, Bishop KM, Cudkowicz ME. An antisense oligonucleotide against SOD1 delivered intrathecally for patients with SOD1 familial amyotrophic lateral sclerosis: a phase 1, randomised, first-in-man study. Lancet Neurol. 2013 May;12(5):435-42. doi: 10.1016/S1474-4422(13)70061-9.

Mueller-Pillasch F, Pohl B, Wilda M, Lacher U, Beil M, Wallrapp C, Hameister H, Knöchel W, Adler G, Gress TM. Expression of the highly conserved RNA binding protein KOC in embryogenesis. Mech Dev. 1999 Oct;88(1):95-9.

Munro, T.P., Kwon, S., Schnapp, B.J., and St Johnston, D. (2006). A repeated IMP-binding motif controls oskar mRNA translation and anchoring independently of Drosophila melanogaster IMP. The Journal of cell biology *172*, 577-588.

Nalavadi, V.C., Griffin, L.E., Picard-Fraser, P., Swanson, A.M., Takumi, T., and Bassell, G.J. (2012). Regulation of zipcode binding protein 1 transport dynamics in axons by myosin Va. The Journal of neuroscience : the official journal of the Society for Neuroscience *32*, 15133-15141.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell stem cell *4*, 487-492.

Nielsen, F.C., Nielsen J., Kristensen, Mette A., Koch, Grete, Christiansen, Jan (2002). Cytoplasmic trafficking of IGF-II mRNA-binding protein by conserved KH domains. Journal of Cell Science *115*, 2087-2097.

Nielsen, J., Kristensen, M.A., Willemoes, M., Nielsen, F.C., and Christiansen, J. (2004). Sequential dimerization of human zipcode-binding protein IMP1 on RNA: a cooperative mechanism providing RNP stability. Nucleic acids research *32*, 4368-4376.

Nielsen J., C., J., Lykke-Andersen, J., Johnsen, A.H., Wewer, U.M. and Nielsen, F.C. (1999). A Family of Insulin-Like Growth Factor II mRNA-Binding Proteins Represses Translation in Late Development. Molecular and Cellular Biology *19*, 1262-1270.

Nielsen, J.A., Sidsel K.; Rajpert-De Meyts, Ewa; Lykke-Andersen, Jens; Koch, Grete; Christainsen, Jan; Nielse, Finn C. (2003). Nuclear transit of human zipcode-binding protein IMP1. Biochem Journal *376*, 383-391.

Nishino, J., Kim, S., Zhu, Y., Zhu, H., and Morrison, S.J. (2013). A network of heterochronic genes including Imp1 regulates temporal changes in stem cell properties. eLife *2*, e00924.

Noubissi, F.K., Elcheva, I., Bhatia, N., Shakoori, A., Ougolkov, A., Liu, J., Minamoto, T., Ross, J., Fuchs, S.Y., and Spiegelman, V.S. (2006). CRD-BP mediates stabilization of betaTrCP1 and c-myc mRNA in response to beta-catenin signalling. Nature *441*, 898-901.

Noubissi, F.K., Nikiforov, M.A., Colburn, N., and Spiegelman, V.S. (2010). Transcriptional Regulation of CRD-BP by c-myc: Implications for c-myc Functions. Genes & cancer *1*, 1074-1082.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F. and Mandel, J.L. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science, 1991. 252(5009): p. 1097-102.

Patel, V.L., Mitra, S., Harris, R., Buxbaum, A.R., Lionnet, T., Brenowitz, M., Girvin, M., Levy, M., Almo, S.C., Singer, R.H., and Chao, J.A. (2012). Spatial arrangement of

an RNA zipcode identifies mRNAs under post-transcriptional control. Genes & development *26*, 43-53.

Pellizzoni, L., B. Charroux, and G. Dreyfuss, SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. Proc Natl Acad Sci U S A, 1999. 96(20): p. 11167-72.

Penny, G.K., G; Sheardown, S; Rastan, S; Brockdorff, N (1996). Requirement for Xist in X Chromosome inactivation. Nature *379*, 131-137.

Perycz, M., Urbanska, A.S., Krawczyk, P.S., Parobczak, K., and Jaworski, J. (2011). Zipcode binding protein 1 regulates the development of dendritic arbors in hippocampal neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience *31*, 5271-5285.

Polesskaya, A., Cuvellier, S., Naguibneva, I., Duquet, A., Moss, E.G., and Harel-Bellan, A. (2007). Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. Genes & development *21*, 1125-1138.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Huelga, S.C., Moran, J., Liang, T.Y., Ling, S.C., Sun, E., Wancewicz, E., Mazur, C., Kordasiewicz, H., Sedaghat, Y., Donohue, J.P., Shiue, L., Bennett, C.F., Yeo, G.W., and Cleveland, D.W. (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nature neuroscience *14*, 459-468.

Ross, A.F., Oleynikov, Y., Kislauskis, E.H., Taneja, K.L., Singer, R.H (1997). Characterization of a beta-actin mRNA zipcode-binding protein. Molecular and Cellular Biology *17*, 2158-2165.

Ross, J.L., I., Berberet, B. (2001). Overexpression of an mRNA-binding protein in human colorectal cancer. Oncogene *20*, 6544-6550.

Runge, S., Nielsen, F.C., Nielsen, J., Lykke-Andersen, J., Wewer, U.M., and Christiansen, J. (2000). H19 RNA binds four molecules of insulin-like growth factor II mRNA-binding protein. The Journal of biological chemistry *275*, 29562-29569.

Sasaki, Y., Welshhans, K., Wen, Z., Yao, J., Xu, M., Goshima, Y., Zheng, J.Q., and Bassell, G.J. (2010). Phosphorylation of zipcode binding protein 1 is required for brain-derived neurotrophic factor signaling of local beta-actin synthesis and growth cone turning. Neuroscience *30*, 9349-9358.

Schulman BR, Esquela-Kerscher A, Slack FJ. Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. Dev Dyn. 2005 Dec;234(4):1046-54.

Schwartz SP1, Aisenthal L, Elisha Z, Oberman F, Yisraeli JK. A 69-kDa RNA-binding protein from Xenopus oocytes recognizes a common motif in two vegetally localized maternal mRNAs. Proc Natl Acad Sci U S A. 1992 Dec 15;89(24):11895-9.

Sim, S., Yao, J., Weinberg, D.E., Niessen, S., Yates, J.R., 3rd, and Wolin, S.L. (2012). The zipcode-binding protein ZBP1 influences the subcellular location of the Ro 60-kDa autoantigen and the noncoding Y3 RNA. Rna *18*, 100-110.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. 2009. Role of the murine reprogramming factors in the induction of pluripotency. Cell. Jan 23;136(2):364-77.

Stohr, N., Lederer, M., Reinke, C., Meyer, S., Hatzfeld, M., Singer, R.H., and Huttelmaier, S. (2006). ZBP1 regulates mRNA stability during cellular stress. The Journal of cell biology *175*, 527-534.

Sundell CL, Singer RH. Actin mRNA localizes in the absence of protein synthesis. J Cell Biol. 1990 Dec;111(6 Pt 1):2397-403.

Sundell CL, Singer RH. Requirement of microfilaments in sorting of actin messenger RNA. Science. 1991 Sep 13;253(5025):1275-7.

Suvasini, R., Shruti, B., Thota, B., Shinde, S.V., Friedmann-Morvinski, D., Nawaz, Z., Prasanna, K.V., Thennarasu, K., Hegde, A.S., Arivazhagan, A., Chandramouli, B.A., Santosh, V., and Somasundaram, K. (2011). Insulin growth factor-2 binding protein 3 (IGF2BP3) is a glioblastoma-specific marker that activates phosphatidylinositol 3-kinase/mitogen-activated protein kinase (PI3K/MAPK) pathways by modulating IGF-2. The Journal of biological chemistry *286*, 25882-25890.

Szarvas, T., Tschirdewahn, S., Niedworok, C., Kramer, G., Sevcenco, S., Reis, H., Shariat, S.F., Rubben, H., and Vom Dorp, F. (2014). Prognostic value of tissue and circulating levels of IMP3 in prostate cancer. International journal of cancer *135*, 1596-1604.

Tamada, A., Kawase, S., Murakami, F., and Kamiguchi, H. (2010). Autonomous right-screw rotation of growth cone filopodia drives neurite turning. The Journal of cell biology *188*, 429-441.

Tessier, C.R.D., G.A.; Clark, B.A.; Pitot, H.C.; Ross, J (2004). Mammary Tumor Induction in Transgenic Mice Expressing an RNA-Binding Protein. Cancer research *64*, 209-214.

Tiscornia, G.S., O.; Verma, I.M. (2006). Design and cloning of lentiviral vectors expressing small interfering RNAs. Nature protocols *1*, 234-240.

Viswanathan, S.R., Powers, J.T., Einhorn, W., Hoshida, Y., Ng, T.L., Toffanin, S., O'Sullivan, M., Lu, J., Phillips, L.A., Lockhart, V.L., Shah, S.P., Tanwar, P.S., Mermel, C.H., Beroukhim, R., Azam, M., Teixeira, J., Meyerson, M., Hughes, T.P., Llovet, J.M., Radich, J., Mullighan, C.G., Golub, T.R., Sorensen, P.H., and Daley, G.Q. (2009). Lin28 promotes transformation and is associated with advanced human malignancies. Nature genetics *41*, 843-848.

Verkerk, A.J., Pieretti, P., Sutcliffe,J.S., Fu, Y.H., Kuhl, D., Pizzuti, A., Reiner, O., Richards, S., Victoria, M., Zhang, F., Eussen, B., van Ommen, G-B., Blonden, L., Riggins, G., Chastain, J., Kunst, C., Galjaard, H., Caskey, T., Nelson, D., Oostra, B., Warren, S. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell, 1991. 65(5): p. 905-14.

Wachter, D.L., Kristiansen, G., Soll, C., Hellerbrand, C., Breuhahn, K., Fritzsche, F., Agaimy, A., Hartmann, A., and Riener, M.O. (2012). Insulin-like growth factor II mRNA-binding protein 3 (IMP3) expression in hepatocellular carcinoma. A clinicopathological analysis with emphasis on diagnostic value. Histopathology *60*, 278-286.

Wachter, K., Kohn, M., Stohr, N., and Huttelmaier, S. (2013). Subcellular localization and RNP formation of IGF2BPs (IGF2 mRNA-binding proteins) is modulated by distinct RNA-binding domains. Biological chemistry *394*, 1077-1090.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell *10*, 440-454.

Weidensdorfer, D., Stohr, N., Baude, A., Lederer, M., Kohn, M., Schierhorn, A., Buchmeier, S., Wahle, E., and Huttelmaier, S. (2009). Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. Rna *15*, 104-115.

Welshhans, K. and Bassell, GJ. Netrin-1-induced local β-actin synthesis and growth cone guidance requires zipcode binding protein 1. J Neurosci. 2011 Jul 6;31(27):9800-13.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., Krainer, A.R., Darnell, R.B., and Zhang, C. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell reports *6*, 1139-1152.

Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., Kazan, H., Vu, A.Q., Massirer, K.B., Morris, Q., Hoon, S., and Yeo, G.W. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Molecular cell *48*, 195-206.

Wisdom, R. and Lee, W. The protein-coding region of c-myc mRNA contains a sequence that specifies rapid mRNA turnover and induction by protein synthesis inhibitors. Genes Dev. 1991 Feb;5(2):232-43

Yao, J., Sasaki, Y., Wen, Z., Bassell, G.J., and Zheng, J.Q. (2006). An essential role for beta-actin mRNA localization and translation in Ca2+-dependent growth cone guidance. Nature neuroscience *9*, 1265-1273.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130-137.

Yeo, G.W., Xu, X., Liang, T.Y., Muotri, A.R., Carson, C.T., Coufal, N.G., and Gage, F.H. (2007). Alternative splicing events identified in human embryonic stem cells and neural progenitors. PLoS computational biology *3*, 1951-1967.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, II, and Thomson, J.A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, M., Mulley, J.C., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. Fragile X genotype characterized by an unstable region of DNA. Science, 1991. 252(5009): p. 1179-81.

Zisoulis, D.G., Lovci, M.T., Wilbert, M.L., Hutt, K.R., Liang, T.Y., Pasquinelli, A.E., and Yeo, G.W. (2010). Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nature structural & molecular biology *17*, 173-179.

Zuk, A.a.H., E. (1994). Expression of B1 Integrins Changes During Transformation of Avian Lens Epithelium to Mesenchyme in Collagen Gels. Developmental Dynamics *201*, 378-393.