UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Role of UPF3B in Neurodevelopment

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of Science

in

Biology

by

Kendall Todd Higgins

Committee in Charge:

Professor, Miles Wilkinson, Chair
Professor, Jens Lykke-Andersen
Professor, Amy Pasquinelli

2016

The Thesis of Kendall Todd Higgins is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

Co-Chair

Chair

University of California, San Diego

2016

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Nonsense Mediated mRNA Decay – NMD

Premature Termination Codon – PTC

Termination Codon – TC

Guanosine Triphosphate – GTP

Eukaryotic Release Factor – eRF

Exon Junction Complex – EJC

Untranslated Region – UTR

Messenger Ribonucleic Protein –mRNP

Copy Number Variant – CNV

Human Embryonic Stem Cells – hES

Induced Pluripotent Stem Cells – iPSC

Clustered Regular Interspersed Repeats – CRISPR

CRISPR associated – Cas

Protospacer Adjacent Motif – PAM

CRISPR RNAs – crRNAs

trans-activating RNAs – tracrRNA

non-homologous end joining – NHEJ

homology directed repair – HDR

Human Accelerated Regions – HARs

Ethylenediaminetetraacetic acid – EDTA

Translation Start Site – TSS

Reads Per Kilobase of transcript per Million mapped reads – RPKM

Fragments Per Kilobase of transcript per Million mapped reads – FPKM

Gene Ontology – GO

Transcripts Per Million – TPM

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Miles Wilkinson for his continued guidance and support and allowing me the opportunity for education. I would also like to thank Ada Shao for her careful instruction and patience as teacher. Lastly I would like to thank Terra Plank-Brannan for her support throughout the project.

ABSTRACT OF THE THESIS

Role of UPF3B in Neurodevelopment

by

Kendall Todd Higgins

Master of Science in Biology

University of California, San Diego, 2016

Professor Miles Wilkinson, Chair
Professor Jens Lykke-Andersen, Co-Chair

The Nonsense Mediated Decay (NMD) pathway was originally discovered for its role in degrading mRNA transcripts bearing a premature termination codon (PTC). Recently, it has been shown that the NMD pathway also degrades a subset of normal transcripts. This raises the possibility that NMD regulates normal biological events,

including development. Indeed, increasing evidence suggests that NMD regulates various developmental events, including governing early cell fate in the specification of the three germ layers: endoderm, mesoderm, and ectoderm, from the pluripotent stem cell. In particular, NMD promotes formation of mesoderm while inhibiting the formation of endoderm.

As support of NMD holding developmental importance, to date 11 patients bearing mutations in one NMD factor, UPF3B, have been reported, all of which exhibit neurodevelopmental disorders including intellectual disability, autism spectrum disorder, and schizophrenia. In addition, reports of individuals with copy number variants in NMD factors UPF2, UPF3A, RBM8A, EIF4A3, RNPS1, and SMG6, have been shown to be associated with neurodevelopmental disorders. This evidence suggests a role for NMD in governing neurodevelopment. Using both *in vitro* and *in silico* methods, I have elected to identify the role of one such NMD factor, UPF3B in early cell fate. Specifically, focusing on its regulation of the neural induction process from the pluripotent stem cell state. As well as providing speculation on a potential evolutionary significance of the NMD pathway and its implications in neurodevelopment in the transition from early hominid to modern human.

INTRODUCTION

History of Nonsense Mediated Decay

The impact of nonsense mutations on mRNA stability was first found in *S. Cerevisiae*, where nonsense mutations in transcripts reduced the amount of mRNA present while still maintaining the same rate of synthesis (Losson and Lacroute, 1979). Subsequent work done in yeast established the identity of the "up-frameshift" genes, *UPF1, 2*, and *3*, found to be necessary for NMD (Leeds P, 1991) (He F and Jacobson A, 1995) (Lee and Culbertson, 1995). In *C. elegans* identification of the suppressors with morphogenic effects on genitalia, or *SMG* genes, were identified with SMG2-4 being homologs for the UPF factors already identified in yeast (Pulak and Anderson, 1993). These core NMD factors (UPF1/2/3, and SMG1, SMG5/6/7) are conserved in most metazoans while UPF3 remains the least conserved as a vertebrates contain two paralogs, UPF3A and UPF3B (Karousis ED, 2016).

Current Models of NMD

Current models regarding the way in which these factors function to elicit NMD are still being studied however here is an overview of those models. Previous work establishes that NMD is a branched pathway. Where some of the factors are necessary for proper NMD and others are dispensable depending on the branch acting on the transcripts. NMD can be elicited by inefficient translation termination where a termination codon (TC) is recognized by a translating ribosome along with eukaryotic release factor 1 and 3 (eRF1) and (eRF3) bound together with GTP (A, 2012) (Schweingruber C, 2013). At this point interaction with UPF1 can occur either in an exon junction complex (EJC) independent or

dependent manner. In the independent manner UPF1 bound non-specifically to messenger ribonucleoproteins (mRNPs) will interact with eRF3 and recruit both or either UPF2 and UPF3B. The pathway converges with the EJC dependent pathway where similar protein associations take place however in the recruitment of UPF2 and UPF3B, UPF3B is bound directly to an EJC complex and UPF2, then UPF2 is able to bind UPF1 and eRF3. After the association with either or both UPF2 and UPF3B with UPF1, SMG1,8, and 9 are recruited to the transcript and SMG1 phosphorylates UPF1 displacing SMG8 and SMG9. The subsequent RNA degradation can occur in two different fashions. In an endonucleolytic fashion where SMG6 is recruited, or in an exonucleolytic fashion where phosphorylated UPF1 recruits the decapping complex remove the 5' 7 methylguanosine cap then SMG5 and SMG7 recruitment facilitates interaction with the CCR4-NOT complex to promote deadenylation of the transcript (Karousis ED, 2016).

NMD acting as a branched pathway is further complicated by the finding that it also acts in a cell type specific manner where branched feedback loops maintain NMD factor levels in a self-governing fashion, making the NMD factors themselves targets of the pathway in different contexts dependent on cell type (Huang L, 2011).

Specific transcript features have been shown to make a normal transcript likely to be regulated by NMD. These are a 5' upstream open reading frame, where the proximity of the stop codon in the 5' end of the transcript gives it potential to be recognized as a PTC (Oliveira and McCarthy, 1995). Normal transcripts bearing an intron in the 3'UTR more than 55 nucleotides away from the normal stop codon can be targeted for NMD as the EJC downstream has the potential to prevent the removal of the ribosome eliciting NMD (Maquat L, 1994). Alternative splicing events where the alternatively spliced exon

relocates the stop codon to an area greater than 55 nucleotides away from the last EJC, in a similar fashion as previously mentioned (Ni JZ, 2007). The final feature, a long 3' UTR, does not have clear mechanism by which these are able to elicit NMD while evidence exists for an enrichment of long 3'UTRs in yeast and mammalian cells (Kebaara and Atkin, 2009) (Yepiskoposyan H, 2011) (Boehm V, 2014) conflicting data suggesting that many normal transcripts are able to evade NMD as well as cis acting elements within the 3'UTR can also inhibit a transcript with a long 3' UTR from being targeted for decay by NMD (Toma KG, 2015).

Role of NMD in Development

With the molecular mechanism which NMD operates still being uncovered recent work has shed light on the physiological importance of NMD, knock out models of NMD factors UPF1, UPF2, and SMG6 in mice, result in embryonic lethality, (Medghalchi SM, 2001) (Weischenfeldt J, 2008) (Li T, 2015) This lethality exists in other organisms, UPF1 and UPF2, knock out in fruit flys, and reduction in zebrafish, result in embryonic lethality (Avery P, 2011) (Wittkopp N, 2009). Reduction of other NMD factors, NAG and DHX34, in nematode and zebrafish also result in embryonic lethality (Anastasaki, 2011) (Longman D, 2007). As further support, reduction or loss of EJC components which associate with UPF and SMG proteins to elicit NMD (Karousis ED, 2016), can also have similar effects, knockout mice of EJC component Magoh result in embryonic lethality (McMahon JJ, 2014)) Taken together, strong evidence that proper functioning NMD is necessary for development.

Role of NMD in Neurodevelopment

Specifically, in neurodevelopment NMD and EJC core components have been shown to be necessary, as loss or reduction of many of these components results in neurodevelopmental disorders. Identification of copy number variants (CNVs) in *UPF2*, either gain or loss, result in intellectual disability. While loss of function mutations in *UPF3B* result in variable neurodevelopmental disorders such as autism spectrum disorders, intellectual disability, and schizophrenia (Nguyen, 2013) (Tarpey PS, 2007) (Lyncha SA, 2012) (Laumonnier F, 2010) (Addington AM, 2011). CNV deletion or duplication of other NMD factors *UPF3A* and *SMG6* as well as EJC components *EIF4A3, RNSPS1, MAGOH, MAGOHB, MLN51, and RBM8A*, have all been found with patients with intellectual disability (Nguyen, 2013). However, the strongest association with ID in humans and NMD are mutations in *UPF3B* (Nguyen, 2013). NMD's role in both early cell fate specification and its implications on the early stages of the neural differentiation process still remain relatively uncharacterized.

INTRODUCTION

Traditional cell lines of human embryonic stem cells (hES) as well as induced pluripotent stem cells (iPSCs) are derived from, or mimic, the inner cell mass (ICM) of the human post implantation embryo (Weinberger L, 2016) (Thomson JA, 1998) (Takahashi K, 2007) (Yu J, 2007). Such cells have the capacity to differentiate into the three primary germ layers (Thomson JA, 1998) (Takahashi K, 2007). The process which the ICM is converted to the three germ layers, starts with cellular organization into the epiblast and the hypoblast. Then the developing human embryo undergoes gastrulation to form the three primary germ layers. Gastrulation begins with formation of the primitive streak which is an invagination in the bilaminar disc of the epiblast that is the center of cellular reorganization that will give rise to the three germ layers. As cells migrate inward at the primitive streak the inner most cells migrating into the invagination become the endoderm, with those following afterward becoming the mesoderm. Those cells on the outside of the invagination will ultimately become the ectoderm (Niakan KK, 2012) The inner most germ layer derived is the endoderm which gives rise to respiratory organs and digestive organs, The inner most germ layer—the endoderm—gives rise to respiratory organs and digestive organs. The middle layer—the mesoderm—gives rise to connective tissue, heart, muscle, bone, and the urogenital system. The outermost layer—the ectoderm—is the precursor to the epithelium and the nervous system (Solnica-Krezel L, 2012).

The extent to which hES cell models can recapitulate germ layer specification is somewhat unclear. As they correspond with ICM cells, differentiation into epiblast must first occur followed by generation of the three primary germ layers (Niakan KK, 2012). However much of the process of gastrulation depends on cell-cell interactions, changes in

tissue structure, and morphogen gradients, to proceed through development (Mammoto T & Ingber DE, 2010). One assay of pluripotency is germ layer formation assays either in the form of embryoid body formation or teratoma analysis. The former where pluripotent stem cells are grown in non adherent conditions and allowed to self organize and form the three germ layers (Rungarunlert S, 2009) The latter, where pluripotent stem cells are injected to a single area in a mouse and allowed to grow tumors. Then upon sacrifice the tumor is assayed for markers of all three germ layers (Zhang WY, 2012). One way around this is modulation of signaling pathways in order to promote formation of primary germ layers. However this can often be convoluted as pluripotent stem cells are allowed to proceed directly to the formation of the germ layer of interest. In a simplistic sense the earliest germ layer forms that hES models can capture are early formation of mesoendoderm within the invagination of the primitive streak, and the neuroectoderm on the exterior of the primitive streak (Wang L & Chen YE, 2016). Which is most commonly done through altering the levels of Activin, FGF2, and BMP class signaling molecules (Vallier L, 2009). This constraint of the model is important to consider when looking at early germ layer specification considering that hES may be modeling an intermediate cell pool that would otherwise not exist *in vivo.*

Previous work outlining the impact of NMD in the early embryo found that NMD downregulation occurs in order for endoderm lineage and that this downregulatory response is critical for efficient endoderm specification to occur specification to occur. While NMD promotes the mesoderm germ layer. While the mechanism of NMD action in endoderm vs. mesoderm differentiation decisions is not known, perturbation of NMD (through depletion of *UPF1*) was found to dysregulate several transcripts involved in

several signaling pathways, including the TGF-β and BMP signaling pathways that are critical for regulating these events (Lou CH, 2016).

Patients with mutations in UPF3B result in varable intellectual disability, 4 patients have been reported with nonsense mutations in the transcript encoding for UPF3B and 7 with missense mutations. The context in which a core component in the NMD pathway could be lost, while in patients bearing a nonsense in UPF3B ultimately have no UPF3B mRNA or protein, points towards a UPF3B independent branch of the NMD pathway that still remains functional. All of the missense mutations found in patients result in non conservative amino acid changes in an uncharacterized domain of UPF3B, the mechanism in which they can cause intellectual disability is unclear (Alrahbeni T, 2015).

Considering NMD functions in early cell fate specification and loss of UPF3B results in neurodevelopmental disorders. It is therefore of interest to investigate the effect of loss of UPF3B on early cell fate decision making, as perturbations in the NMD pathway are likely to lead to improper specification of the three primary germ layers with a specific focus on the impact of the ectoderm formation, as it will later give rise to the neural tissue. In this chapter, is evidence for UPF3B functioning to the maintain the pluripotent state, mediate proper cell cycle regulation, as well as progression through the early neural induction process.

RESULTS AND DISCUSSION

UPF3B Null Patient Derived iPSCs Exhibit Perturbed NMD

Patient derived fibroblasts from 3 different patients with nonsense mutations in the coding region of *UPF3B* were recieved from the Jozef Gecz lab University of Adelaide. iPSC had been previously derived by graduate student Ada Shao, these are the cells used subsequently (Nguyen LS, 2012). In order to determine if the *UPF3B* null patient iPSC had reduced NMD magnitude, the transcript levels for the UPF transcripts and 3 validated NMD targets were tested (Figure 1). As expected *UPF3B* levels were significantly reduced, however UPF transcripts expression levels were not significantly increased upon loss of *UPF3B*. NMD magnitude however appears to be somewhat perturbed as 2 out of 3 validated targets increase significantly upon loss of *UPF3B*. NMD is still functioning in these cells as the PTC bearing transcript of *UPF3B* is significantly reduced. In addition, because *UPF1* and *UPF2* are themselves substrates of NMD (Huang L, 2011)

Validation of RNA Seq Identified Differentially Expressed Genes

Previously RNASeq had been performed on patient lacking UPF3B iPSC with iPSC derived from mother serving as relative control for differential gene expression analysis. Several of the highly upregulated genes had been selected for validation by qPCR. Considering the clinical outcome of UPF3B resulting in intellectual disability the result of neurodevelopmental disorders. It was of interest to validate differentially expressed genes with a gene function associated with the nervous system, that had not already been validated as upregulated. Using a candidate approach several genes were selected for validation, those upregulated upon loss of UPF3B as those would theoretically correspond

with direct NMD targets. Shown in Figure 2 are expression levels of selected neural associated transcripts found to be upregulated upon loss of UPF3B by RNAseq analysis, which were unexpectedly not upregulated as indicated by qPCR.

From the originally identified differentially expressed genes, one of the enriched gene ontology categories was “Cell Cycle” in the upregulated genes. This in combination with previously findings that reduction of UPF3B in hES cells, results in an accumulation of cells in the G2 phase of the cell cycle (Lou CH, 2016). Suggested that upregulated transcripts identified by RNAseq associated with the cell cycle would be of interest to validate. Figure 2 also shows associated cell cycle transcripts validated by qPCR, finding CDK10 and GDF3 to be significantly upregulated. Interestingly CDK10 is an inhibitory cyclin dependent kinase found to be expressed in the G2/M phase of the cell cycle. Previously overexpression of CDK10 has been shown to reduce proliferation, while knockout results in increased proliferation, colony formation, and migration in biliary tract cancer cells (Yu JH, 2012). GDF3 has been found to act as a bimodal ligand of TGFß and Nodal signaling suggesting a dynamic role in the control of differentiation of the hES cell (Levine AJ, 2006).

UPF3B Promotes Stem Cell Self-Renewal

In order to test the impact of loss of UPF3B on cellular functions the primary focus was on proliferation, as observations of the cells had found them to grow slower when compared to control. In addition previous studies had shown that Upf1 in mice promotes proliferation of neural stem cells (Lou CH, 2014). Shown in Figure 3A-B is an EdU incorporation assay of patient lacking UPF3B iPSC. Revealing a reduced number of cells

actively proliferating as a result of loss of UPF3B. This suggests that UPF3B normally acts to promote pluripotent stem cell proliferation.

Effect of UPF3B on Pluripotency and Primary Germ Cell Markers

One potential reason for perturbations in normal cell cycle progression and specifically reduction in hES self-renewal can be attributed to improper expression of early cell fate markers, specifically known to elicit transcriptional changes during development (Perrimon N, 2012). This combined with previous studies done showing that reduction of UPF proteins in hES cells can alter the formation of the three primary germ layers (Lou CH, 2016). Pointed towards identifying the impact of loss of UPF3B on expression of pluripotency and early germ layer markers. Shown in Figure 4A is qPCR assay of expression of various pluripotency markers identified from literature (Vallier L, 2009).The iPSC patient cells had significantly upregulated expression of both *SOX2* and *NANOG* mRNA, relative to control cells (Figure 4A). This raises the possibility that UPF3B inhibits pluripotency. Another interpretation of this finding is that UPF3B acts to directly to regulate these 2 transcripts; i.e., they are NMD target transcripts. In contrast to *SOX2* and *NANOG* mRNA, *TERT* mRNA expression was significantly downregulated in the patient iPSCs. This is consistent with UPF3B promoting pluripotency, as TERT is required for pluripotency (Huang Y, 2014) However, since *SOX2* and *NANOG* mRNA had the opposite expression pattern, the role of UPF3B in pluripotency is not clear.

In order to test the hypothesis that subtle perturbation in pluripotency markers and germ layer specification were more prevalent in patient cells than in control, due to decreased NMD magnitude as a result of loss of UPF3B, I identified marker expression

variability between data sets. A list of markers of each germ layer as well as pluripotency were gathered from the literature, then expression estimates for each maker based on RNASeq FPKM were determined. This was then averaged for each marker, to calculate a distance from the mean expression of that marker between samples. Figure 5 shows a heat map of control transcripts, such as those associated with housekeeping proteins were not expected to have significant variance between biological replicates. Figures 6-9 show the same analysis for pluripotency, endoderm, mesoderm, and ectoderm respectively with the analysis summarized in Table 1. It was found that markers of pluripotency, endoderm, and mesoderm vary significantly in iPSC lacking UPF3B based on the F-Test results and significant difference in distance from mean expression values across biological replicates shown in Table 1. This lead to the conclusion that loss of UPF3B results in variable expression of pluripotency, endoderm, and mesoderm markers, while a propensity to form a particular germ layer upon loss of UPF3B is unclear. Evidence here suggests that when maintained under pluripotency condition reduction of NMD by loss of UPF3B has a greater impact on maintaining the pluripotent cell state and more directly perturbs endoderm and mesoderm lineages, while ectoderm remains undisturbed. This can be described as UPF3B functioning to confer robustness of germ layer specification.

UPF3B Governs Proper Cell Response To Neural Induction

While identifiable impacts on ectoderm formation were not clear during maintenance in pluripotent cell conditions the finding that UPF3B acts to buffer expression of pluripotency, endoderm, and mesoderm markers, would leave ectoderm marker expression unperturbed. Failure to downregulate markers of pluripotency, endoderm, and

mesoderm during differentiation process may potentially impact proper formation of ectoderm (Perrimon N, 2012). To test this patient and control iPSC were cultured in conditions promoting ectoderm formation by dual SMAD inhibition (Chambers SM, 2009). In order to test if the expression of UPF factors significant changed in response to neural induction, Figure 10 shows UPF protein expression in response to neural induction by dual SMAD inhibition with statistical test summarized in Table 2. Finding that upon induction UPF3B expression increases significantly then after two days of the induction UPF3B expression levels reduce significantly. This points towards a role in governing response to the corresponding transcriptional changes elicited by neural induction. In order to test this, expression of known markers of neural induction were assayed in patient lacking UPF3B iPSC undergoing neural induction, to determine if expected decrease in pluripotency markers occurs in patient along with upregulation of neural markers. Shown in Figure 11A-B is expression of either pluripotency markers OCT4 and NANOG or early ectoderm markers SOX1 and PAX6 in patient and control respectively during the first two days of neural induction. In control there is a characteristic downregulation of NANOG upon neural induction however in patient lacking UPF3B iPSC NANOG fails to be downregulated in response to neural induction. The differences in expression for each corresponding time point are shown in Figure 11C-D. Revealing that at day 2 of neural induction SOX1 is significantly upregulated in patient lacking UPF3B iPSC, furthermore at day 1 and 2 of neural induction NANOG is significantly upregulated in patients lacking UPF3B iPSC. As silencing of pluripotency markers is a key indicator of proper progression through the neural induction process for establishment of neuroectoderm (Chambers SM, 2009) it can be concluded that UPF3B governs proper cell response to neural induction by

downregulation of pluripotency markers. As well as maintaining proper levels of neural associated markers in response to the induction process.

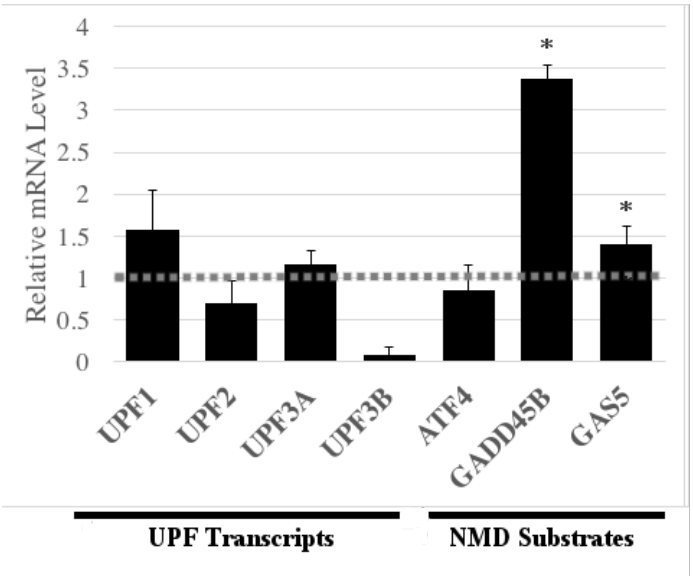

Figure 1: Analysis of patient NMD factors and substrates. Control expression level corresponds to relative mRNA level of 1. Error bars represent standard error. * indicates P<.05 t-test n=3 biological replicates.

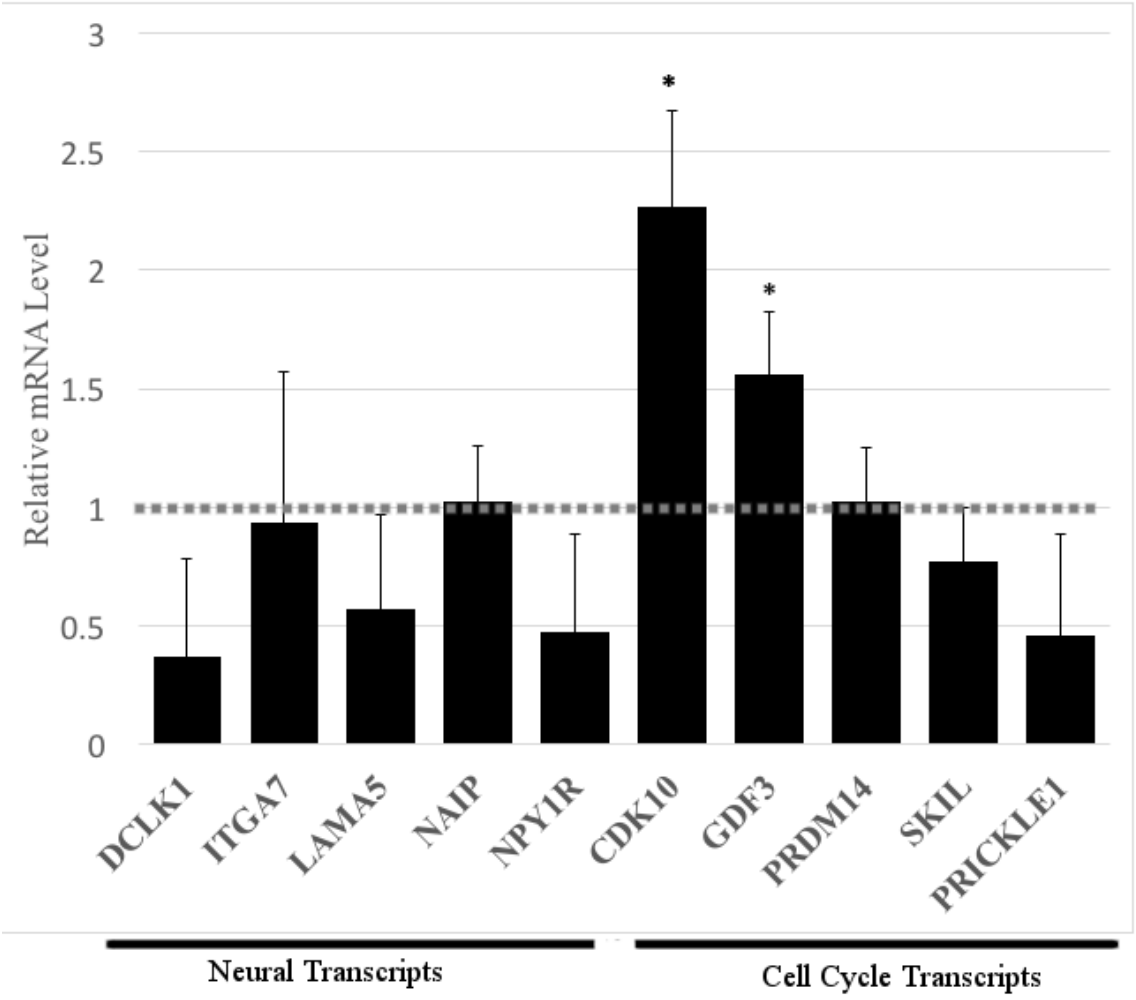

Figure 2: Analysis of patient selected transcripts from differentially expressed genes as determined by RNA Seq. Control expression level corresponds to relative mRNA level of 1. Error bars represent standard error. * indicates P<.05 t-test n=3 biological replicates.

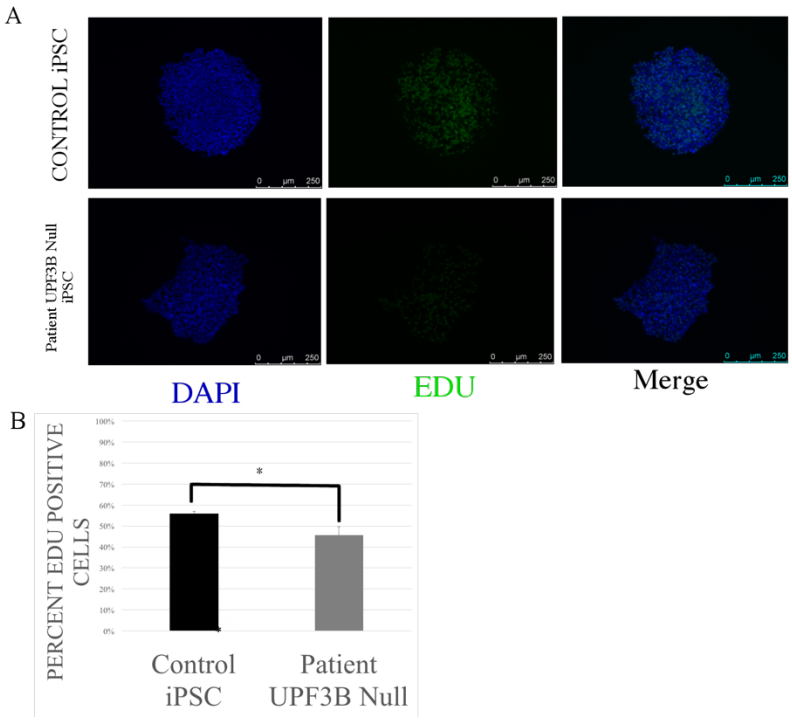

Figure 3: EdU incorporation assay of patient iPSC.(A) Total positive EdU cells were compared to DAPI positive cells present to establish as ratio of EdU positive cells. Representative images shown above, results here show a significant reduction in the percentage of EdU positive cells in patient iPSC (B) Quantification of the data shown in panel A where error bars indicate standard error. * indicates $P < .05$ two sampled t-test, n =15, five individual images across 3 biological replicates**.**

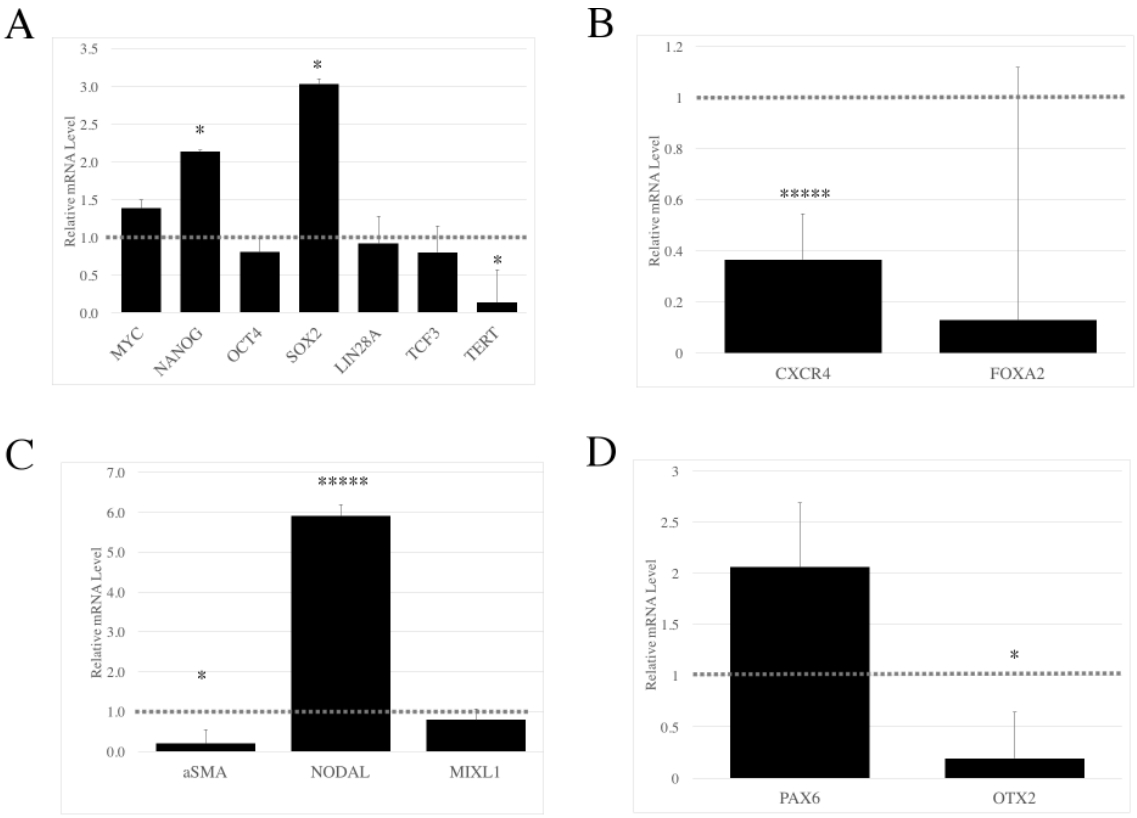

Figure 4: Analysis of patient cell fate markers (A) pluripotency markers (B) Endoderm Markers (C) Mesoderm Markers (D) Ectoderm Markers Control expression level corresponds to relative mRNA level of 1. Error bars represent standard error. * indicates P<.05 *****P<.000001
t-test n=3 biological replicates.

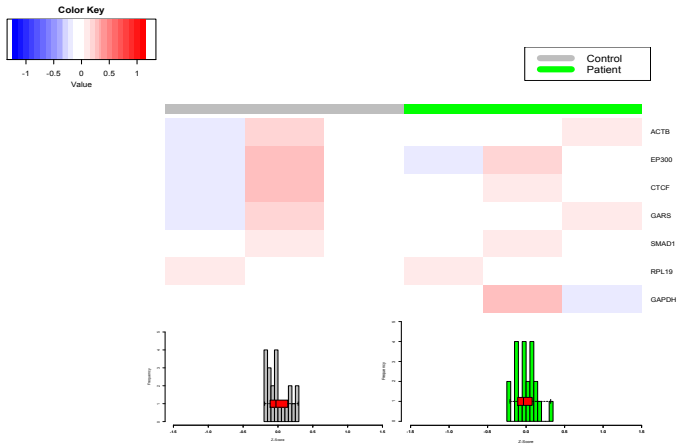

Figure 5: Control marker expression variance in patient. Heat map representing Z-score of difference from average FPKM mean expression value for a given marker. UPF3B null patient (green bar) and control (grey bar) Below corresponding histogram of Z-scores shown in heat map for each marker Z-Score with corresponding box and whisker plot represent median value (midline), upper and lower quartiles (outer edge of box), and range (whiskers).

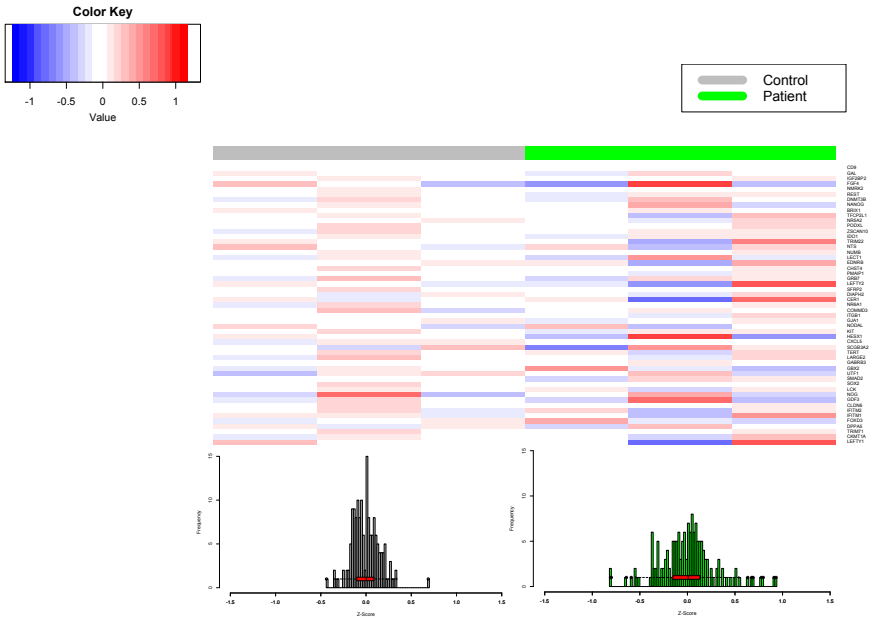

Figure 6: Pluripotency marker expression variance in patient. Heat map representing Z-score of difference from average FPKM mean expression value for a given marker. UPF3B null patient (green bar) and control (grey bar) Below corresponding histogram of Z-scores shown in heat map for each marker Z-Score with corresponding box and whisker plot represent median value (midline), upper and lower quartiles (outer edge of box), and range (whiskers).

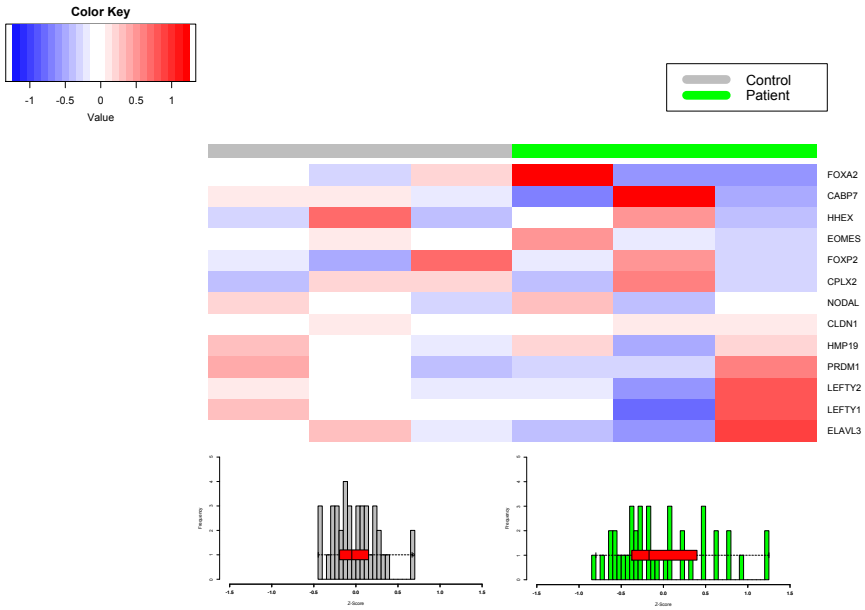

Figure 7: Endoderm marker expression variance in patient. Heat map representing Z-score of difference from average FPKM mean expression value for a given marker. UPF3B null patient (green bar) and control (grey bar) Below corresponding histogram of Z-scores shown in heat map for each marker Z-Score with corresponding box and whisker plot represent median value (midline), upper and lower quartiles (outer edge of box), and range (whiskers).

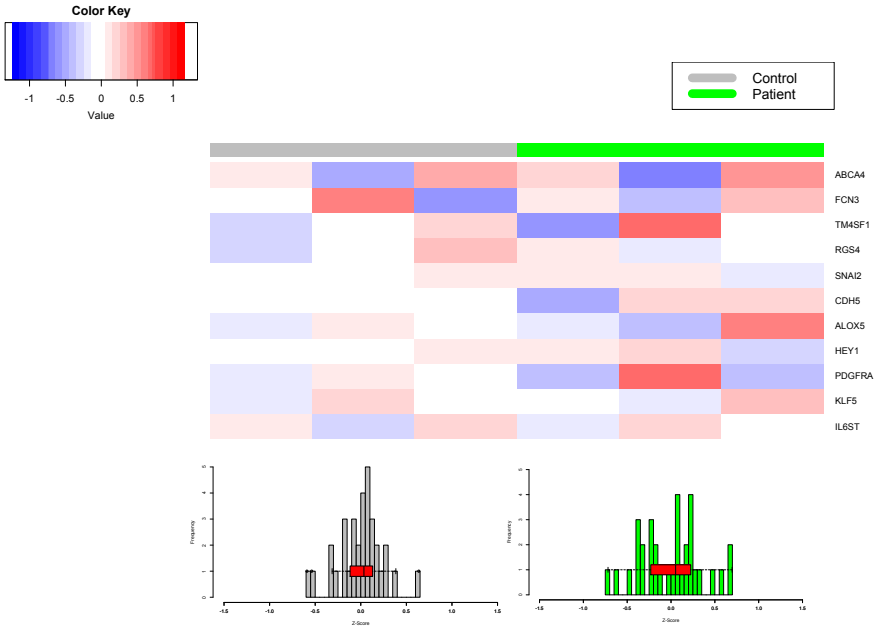

Figure 8: Mesoderm marker expression variance in patient. Heat map representing Z-score of difference from average FPKM mean expression value for a given marker. UPF3B null patient (green bar) and control (grey bar) Below corresponding histogram of Z-scores shown in heat map for each marker Z-Score with corresponding box and whisker plot represent median value (midline), upper and lower quartiles (outer edge of box), and range (whiskers).

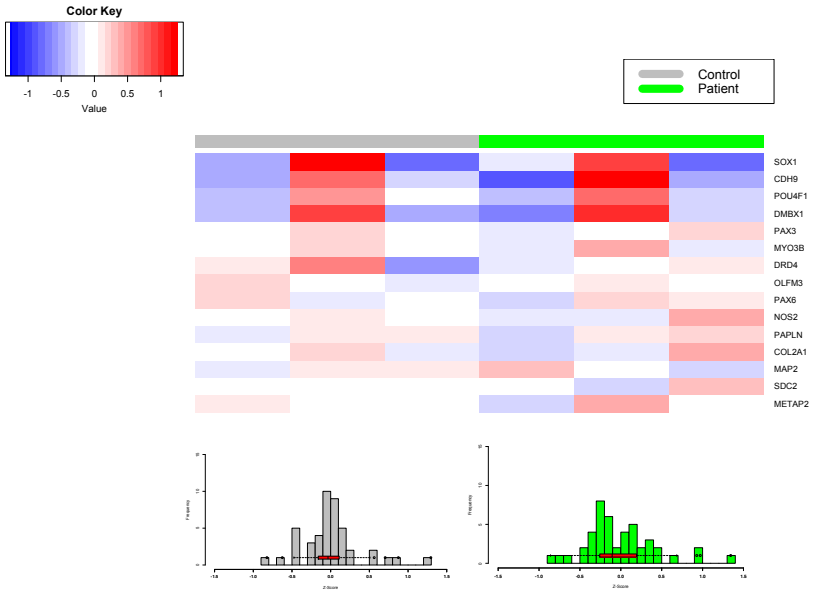

Figure 9: Ectoderm marker expression variance in patient. Heat map representing Z-score of difference from average FPKM mean expression value for a given marker. UPF3B null patient (green bar) and control (grey bar) Below corresponding histogram of Z-scores shown in heat map for each marker Z-Score with corresponding box and whisker plot represent median value (midline), upper and lower quartiles (outer edge of box), and range (whiskers).

Table 1: Statistic test Summary for variance measurements of germ layer specification markers. F test testing variability across all samples in a given population either based on FPKM values, or FPKM values normalized for difference from mean expression of a given marker. One tailed t-test of absolute value of distance from mean expression for each gene along with corresponding averages for each subset of markers.

					Average of Absolute Value of Z-Scores	
Markers	***N***	***F-Test FPKM (P-Values)***	***F-Test FPKM Z-Score (P-Values)***	***T-Test FPKM Z-Score Average (p-value)***	***Control***	***Patient***
Control	7	0.67938	0.505735804	0.2079053	0.11612	0.09524
Pluripotency	53	0.06825	4.50346E-15	0.0000001	0.11917	0.21393
Endoderm	13	0.02246	3.40182E-05	0.0007658	0.21330	0.44556
Mesoderm	11	0.75297	0.04972216	0.0123085	0.18123	0.28448
Ectoderm	15	0.13592	0.307502017	0.0903533	0.24834	0.32773

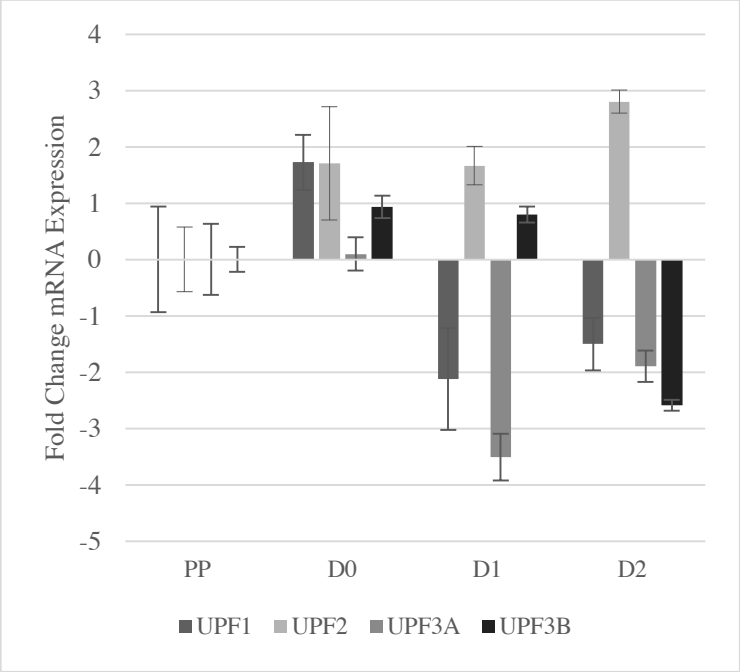

Figure 10: Control UPF factor expression in response to neural induction. Relative quantification qPCR baseline expression level corresponds to Fold Change of 0, taken from control cell expression at pluripotency. For pluripotency time point (PP), day prior to neural induction (D0), 1 day of Neural induction (D1), or 2 days neural induction (D2). Error bars represent standard error. n=3 biological replicates.

Table 2: Statistics summary of UPF factor expression in response to neural induction of paired samples t-test p-values

	D0	***D1***	***D2***
UPF1	0.02061	0.20947	0.03575
UPF2	0.63357	0.00353	0.00170
UPF3A	0.87197	0.00881	0.00098
UPF3B	0.02272	0.04147	0.00001

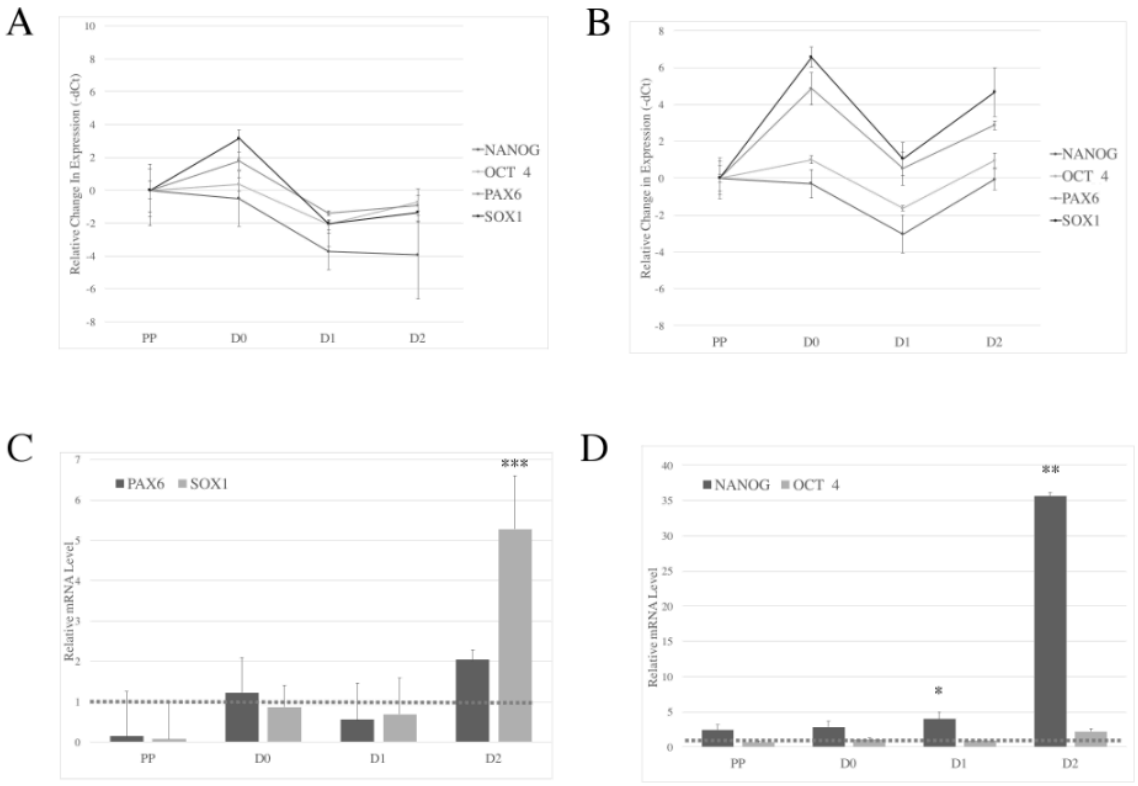

Figure 11: Effect of loss of UPF3B during early neural induction by dual SMAD inhibition in patient. Pluripotency time point (PP), day prior to neural induction (D0), 1 day of Neural induction (D1), or 2 days neural induction (D2). Control expression level corresponds to relative mRNA level of 1. Error bars represent standard error. * P<.05 ** P<.01***P<.001 t-test n=3 biological replicates

MATERIALS AND METHODS

Cell Culture Conditions

iPSC were cultured on BD Matrigel coated plates in Essential 8 Media at 37 degC in 5% CO2, media changed daily, after a single wash in 1X DPBS. Cells were passaged with manual dissociation after 3 minute incubation in 0.5mM EDTA at 37degC, 1:4-1:5 ratio every 3-5 days when cells reached 70% confluency. Cultures were monitored daily for differentiated cells which were manually removed.

RNA Isolation and qPCR

Total RNA was isolated using Trizol reagent after washing the cells in 1X DPBS as described (Rio DC, 2010). cDNA was synthesized using 1.5 microgram total RNA, 1 microliter iScript and 4 microliters iScript Buffer in a 20 microliter reaction using manufacturer reverse transcription protocol. qPCR analysis was performed using SYBR green reagents in a relative quantification method (ddCt) where RPL19 was used as endogenous control.

Neural Induction

Prior to induction iPSC were grown to 100% confluency on matrigel coated plates in Essential 8 Media. Then media was changed to neural induction media as described in (Shi Y, 2012) where dual SMAD inhibitors used were 1mM of Dorsomorphin was used in place of Noggin. RNA was isolated at pluripotency prior to 100% confluency of cell growth, Day 0 (D0) starts when cells were grown to 100% confluency prior to neural

induction, with the subsequent days 1 and 2 (D1) and (D2) the following days after switching to neural induction media.

RNASeq Analysis

Single end libraries were prepared using Illumina Hiseq2000 platform. Raw FASTQ files were quality controlled for using FASTQC program with a cut off quality score of 10 and trimmed 15 bp at each end to remove adapter content. Reads were mapped to both human genome hg38 and Ensembl transcriptome GRCh38 using STAR 2.4.2 two pass mode alignment. Fragments Per Kilobase of transcript per Million or Transcript per Million, Gene and Isoform abundance was estimated using RSEM version 1.2.23. Differentially genes and transcripts were estimated using EBSeq with a false discovery rate cut off of $P<0.05$.

CRISPR MEDIATED GENERATION OF UPF3B NULL INDUCED PLURIPOTENT STEM CELLS

INTRODUCTION

Clustered Regular Interspersed Repeats (CRISPR) were first described in 1987 as regions in the E. Coli genome containing repeat sequences interspersed by short sequences (Ishino Y, 1987). The link between genome engineering and CRISPR systems did not come until the discovery that viral DNA or RNA sequences could be found in the interspersed region of the repeats, uncovering the mechanism that CRISPR sequences along with various CRISPR associated (*Cas*) genes embodied a bacterial adaptive immunity (Brouns SJ, 2008). In brief, CRISPR sequences contain antisense RNA of previous viral invasions flanked by a protospacer adjacent motif (PAM) sequence, these are then transcribed to form CRISPR RNAs (crRNAs). Three molecular pathways converging on CRISPR were originally described all containing multiple proteins to influence the pathway; only type II system was able to have viral nucleic acid recognition and cleavage in a single protein complex (Jinek M, 2012). The crRNAs must form a complex with trans-activating crRNAs (tracrRNA) in type II CRISPR systems. Of the Cas genes *Cas9* found both in Streptococcus Thermilus and Streptococcus Pyogenes utilized crRNA:tracrRNA complex to target and introduce double-strand breaks in invading virus (Duodna J and Charpentier E, 2014). In the context of genome engineering in mammalian cells, recombinant versions of both the Streptococcus pyogenes and Streptococcus thermilus along with scaffolds for tracrRNA and crRNA coined (sgRNA) were first reported in 2013 (Cong et al, 2013) (Jinek, 2013) (Mali P, 2013). With *Cong et al.* being the first attempt to do so in hES cells, this opened a door for being able to edit the genome of a pluripotent stem cell, which has large implications on both disease modeling as well as medicine. Genome edits could be made in cells with the ability to differentiate into any cell type in the body. Patient-derived iPSC

could now have genome corrections introduced and be reincorporated into a patient's body significantly reducing the associated immune response.

When introducing genome edits with CRISPR, there are two general approaches. The first being non-homologous end joining (NHEJ) to generate simple knock outs of a gene as a result of indel mutations due to errors in the repair process commonly resulting in PTCs. The second is utilizing the homology directed repair (HDR) route found in normal DNA repair in mammalian cells (Liang F, 1998). Where double-strand DNA breaks are introduced by Cas9 and sgRNA targeting a specific sequence then a donor template also introduced into the cell is used to have directed base pair specific mutations within the cut site of interest. This can be used either to introduce specific amino acid substitutions, or repair mutations found in diseased patient derived cell lines (Ran FA, 2013).

While the applications of this technology are very promising, the largest issue facing this is effiency in generating genome edits either by NHEJ or HDR. Until the recent advent of next generation transfection reagents, nucleofection or electroporation was the standard to transfect hES cells, both of which have a significant impact on the viability of the cells (Ran FA, 2013). In addition to initial low transfection effiencies, rates of HDR still remain very low, even in sucessfully transfected cells (Chu VT, 2015). One approach taken in difficult-to-transfect cell lines and primary cell lines, was integration of Cas9 and sgRNA in a viral construct to allow either constituitive expression of one or both. This was adapted to create and test putative guides against a gene of interest and see if they were able to generate phenotype of interest (Sanjana NE, 2014). The caveat with this approach being side effects of integration of active Cas9 or stress due to viral transduction and the subseqent selection process (Hou P, 2015) (Yiu G, 2016).

Here I present two potential methods to generate CRISPR-mediated UPF3B knock out iPSC derived from a normal human male as well as establish a protocol for single cell cloning of viable undifferentiated hES cells.

RESULTS

Guide Design and Rationale

Shown in Figure 12 is a graphic summary of design of the dual approach to generate UPF3B null iPSC. By using both stable viral and transient vector methods to introduce Cas9 and sgRNA targeting UPF3B. The guide design rationale was to target an earlier exon, in this case the CDS of UPF3B exon 1, to increase that chances of a PTC being generated by frameshift mutation as a result of errors in NHEJ repair of double strand breaks in DNA. Cell lines chosen were male iPSC from a healthy control to eliminate the possibility of a heterozygote mutant as UPF3B is an X-linked gene (Tarpey PS, 2007). Multiplicity of infection was determined by relative functional titer of Cas9 expression to uninfected control (data not shown). Selection for stable integration of viral construct containing guide sequence and Cas9 was done using a puromycin resistance expression in the viral construct. After selection, cloning took place from a enriched pool of iPSC expressing Cas9 and sgRNA targeting UPF3B, with relative control cells being those undergoing transduction with a viral construct lacking sgRNA sequence.

While in the transient vector approach, iPSC were transfected with vector containing Cas9, sgRNA targeting the CDS region of UPF3B exon 1, and a GFP marker. After transfection, cells were fluorescent activated cell sorted (FACS) to obtain single cell clones. Where relative control was either those negatively sorted during flow cytometry or, those sorted as GFP positive that were later identified as wild type for UPF3B mutations.

Guide Validation

In order to validate that guides were making predicted cuts a T7 endonuclease assay was done in HEK293T cells, with sgRNA1, (as shown in Figure 12) ligated into PX458 vector. Shown in Figure 13 are the results from the T7 endonuclease assay finding that sgRNA1 was able to successfully generate cuts expected in the target region of UPF3B as indicated by the appearance of two bands on gel image, the lower of which corresponds with the expected band size given the location of the editing site relative to the ends of the primers. This directly tested the cutting efficiency in the vector approach however evidence that the guide was able to generate cuts was sufficient to assume that once properly integrated into the lentiviral construct, for the viral approach, it would retain cutting efficiency.

Generation of UPF3B Null iPSC

After guide validation, WT male iPSC were introduced to Cas9 with sgRNA targeting exon 1 of UPF3B, either by lentiviral or transient vector. Subsequently cells were single cell cloned and screened for mutations at the predicted cut site by Sanger sequencing (the complete protocol is described in the Materials and Methods). Table 3 shows a summary of the clones generated from either LCv2 (lentivirus containing Cas9, sgRNA, and puromycin resistance marker) lentiviral-mediated mutations or PX458(vector containing Cas9, sgRNA, and GFP marker)-mediated mutations. While successful edits were made at the predicted cut site, the majority were in-frame deletions (either 12 or 21 nucleotides), typically removing amino acids spanning the region from R13 to A22. Considering introduction of indel mutations is presumed to occur randomly, it is interesting

that most of the deletions were in frame and in this region. Furthermore, in the case of the constitutive Cas9 and sgRNA expression resulted in the similar amino acid deletions as were seen most often in the transient approach. This may point towards a potential gain-of-function mutation when proline 19 is deleted as this was common to all mutations found in either experiment. Considering the majority of the mutations generated were in frame one preliminary step was to use bioinformatics tools to predict the potential impact of the mutations. Shown in Table 4 are results of Provean prediction algorithm for the likelihood that the mutations give rise to misfolded protein products and would therefore be a deleterious mutation (Choi Y, 2015). According to the prediction algorithm all of the mutations generated should be deleterious. However not all were screened by western blotting as the one most likely to generate a UPF3B null cell line nucleotide 26_165del (subsequently referred to as CVB iPSC PQV2) was of foremost interest.

Isolation of a Putative UPF3B-null iPSC Cell Clone

Also shown in Table 3 is the most likely clone to result in a UPF3B null iPSC which is nucleotide 26_165del called CVB iPSC PQV2 or PQV2, which removes the normal translation start site (TSS) from the gene. This was the focus of subsequent analysis of the PX458 generated clones. Because the foremost clone of interest, CVB iPSC PQV2, destroyed the original translation start site of UPF3B, prior to screening by western blotting, bioinformatic analysis was carried to predict the potential proteins that would be made, taking into consideration the position of alternative "Start" codons with a proper Kozak context for efficient translation (Kozak M, 1987). Shown in Table 5 are the predicted proteins that could be potentially made given the removal of the translation start

site as predicted by ATGpr (Salamov AA, 1998) . Frame 3 represents the same coding frame that the original TSS used in UPF3B. While the most reliable TSS is in frame with the original start site, it remains a low probability prediction based on Kozak context. Furthermore, the predicted mutation would remove amino acids 1-73 of UPF3B which contain a portion of the complete UPF2 interaction domain that consists of amino acids 47-143 on UPF3B. However previous publications have found that amino acids 52-57 of UPF3B constitute 7 out of 9 of the polar amino acid interactions with UPF2 (Kadlec J, 2004). Experimental evidence from the same publication suggests that amino acids 52 and 57 are necessary for interaction with UPF2. However, this was binding was tested in the context of non-conservative mutations introduced to the region rather than deletion (Kadlec J, 2004).

To determine if protein was present in either the viral mediated UPF3B mutants or CVB iPSC PQV2, western blots were done probing for UPF3B full length antibody against UPF3B, which cross reacts with UPF3A, antibody. Figure 14B shows western blot analysis for clone CVB iPSC PQV1 which contained an in-frame deletion of nucleotides 145_165del. As well as CVB iPSC PQV2. As expected CVB iPSC PQV2 in three "biological replicates" does not produce a band for UPF3B. Figure 14C shows the same samples but at lower protein concentration again producing the same results. One caveat is the UPF3B reduction lysate used here in B and C as a control, H9 Tet-shUPF3B (+) is expected to have reduced UPF3B which doesn't appear to be the case in either blot. Though quantification was not carried out as it was to used identify presence or absence of UPF3B. Shown in Figure 14A are clones generated by viral mediated UPF3B mutations containing mutations NP_542199.1 P19_A22del. Both of which appear to result in no detectable

UPF3B by western blot. The HEK293T UPF3B overexpression lysate used in Figure 14A does not appear to have over expressed UPF3B. RNA levels of UPF3B were previously validated as being overexpressed in the lysate (data not shown).

As the final step to validate if the clones generated were in fact lacking UPF3B a qPCR profiling was done as none of the expected clones would have reduced UPF3B expression at the mRNA level, however as loss of UPF3B can be characterized by upregulation of known NMD factors as well as transcripts shown previously to be upregulated in patient iPSC lacking UPF3B (Huang L, 2011). Shown in Figure 15A is relative quantification qPCR for viral generated clones specifically SG1-7, where transduced control lacking guide sequence serves as a relative control, LCv-2. Noticeably Cas9 expression is significantly increased in the lines relative to control, which is unexpected as greater Cas9 expression would potentially increase the chance of off target effects. UPF3B expression does not decrease which is expected, however NMD substrates do not appear to be reduced in the cell lines suggesting that functional UPF3B may still be present even though it was not detected by western blot. Another likely possibility is that because Cas9 and sgRNA are constitutively expressed selection may have occurred from the time DNA and protein were isolated to screen the cells to favor cells with UPF3B intact as recurring DSB could result in proper repair of the cut site. Ultimately it can be concluded though these cells do not appear to have UPF3B knock out considering they do not recapitulate the “phenotype” or typical upregulation of NMD factors. They likely have functional UPF3B as selection may have occurred between time points, which does not make them a good candidate for long term culture or disease modeling.

Figure 15B shows NMD factor and substrate expression for clone PQV2. Data from PQV1, which has an in-frame deletion, is included for comparison which shows expression near that of control with the exception of *UPF3B*. Clone PQV2 exhibited characteristics of depressed NMD, including upregulation of the known NMD substrates, *GADD45B* and *GAS5* mRNA. This clone also upregulated levels of *GDF3* and *BMP4* mRNA, which were upregulated in *UPF3B* patient iPSCs, as described in Chapter 2. In contrast, clone PQV1 Though direct controls undergoing both transfection and flow sorting were not available for this experiment at this time point, averaging several earlier RNA isolation time points for untransfected cells of the same line were used. It can be concluded based on the evidence presented at the DNA, RNA, and protein level that CVB iPSC PQV2 clone is a likely UPF3B null cell line. Furthermore it can also inferred that if reduced protein product is made it would have loss of a critical region for UPF2 interaction which would yield the NMD magnitude of the cells reduced as the UPF3B dependent branch of NMD should be non-functional.

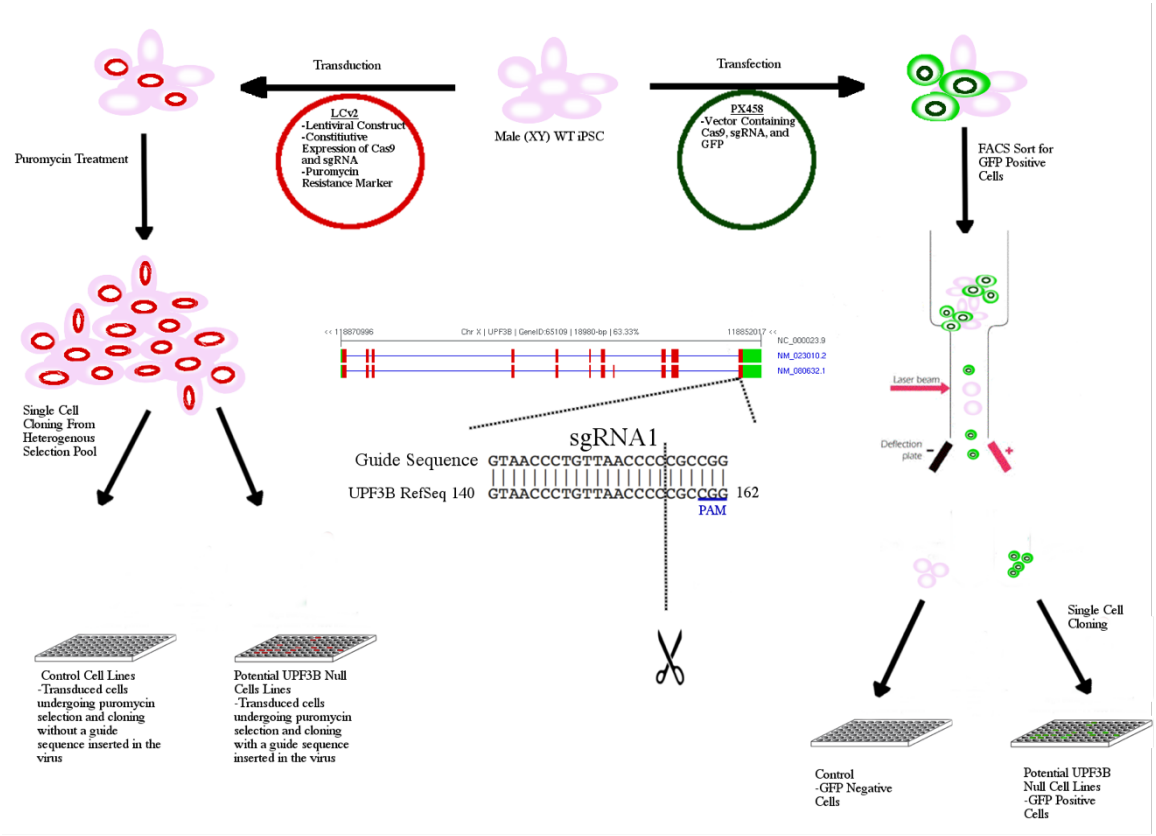

Figure 12: Graphic summary of approach taken to generate *UPF3B* null iPSC from wild type male (XY) derived iPSC. Shown on left is viral approach taking WT iPSC and transducing them with a viral construct containing Cas9, sgRNA targeting UPF3B, and puromycin resistance marker (LCv2). After puromycin selection for successfully transduced cells, an enriched pool of potential UPF3B null cells will be cloned and screened for edits resulting in null mutations of UPF3B. Shown right is transient vector approach where vector containing Cas9, sgRNA targeting UPF3B, and GFP marker (PX458), is transfected into iPSC then subsequently iPSC are FACS sorted for GFP+ positive cells and screened for loss of UPF3B. Shown middle is a graphic detailing the two isoforms of UPF3B, where NM_080632.1 is the more common isoform. Guide 1 targeting CDS of exon one of UPF3B at nucleotide 156/57.

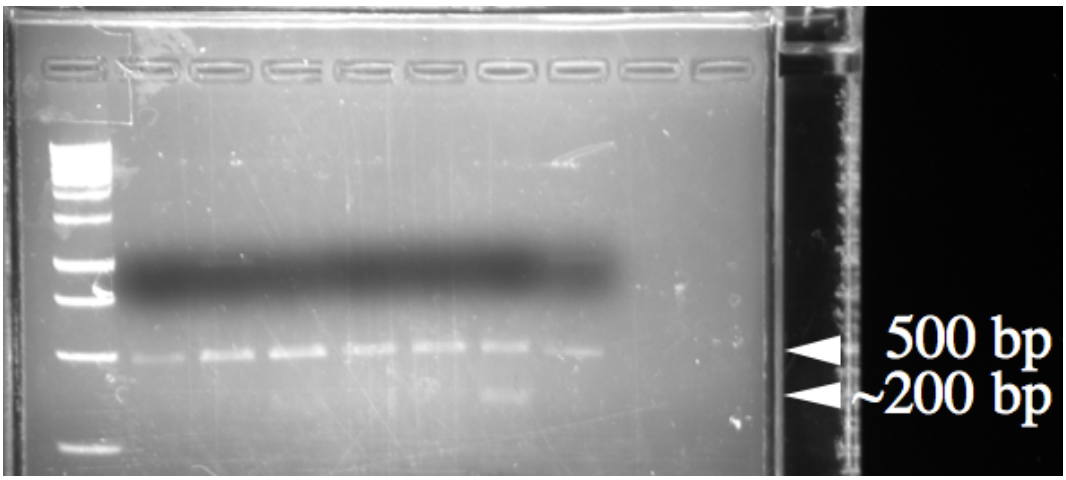

Figure 13: T7 Endonuclease digestion of transfected HEK293T cells with PX458 with guide targeting exon 1 of *UPF3B*. Wells from left to right 1. Lad 2. Untransfected control HEK293T 3. PX458 empty guide control. 4. PX458 sgRNA1-1 replicate 1 5. PX458 sgRNA1-1 replicate 2 6. PX458 sgRNA1-2 replicate 1 7. PX458 SG1-1 replicate 2 8. No T7 Endonuclease control.

Table 3: Summary of iPSC clones generated both by LCv2 and PX458

Mutation		**Clones**	
DNA	***Protein***	***LCv2 Clones****	***PX458 Clones***
NM_080632.1 152_163del	NP_542199.1 P19_A22del	2	0
NM_080632.1 137_157del	NP_542199.1 R13_P19del	1	0
NM_080632.1 140_160del	NP_542199.1 V14_A20del	0	17
NM_080632.1 154_165del	NP_542199.1 P19_A22del	0	3
NM_080632.1 145_165del	NP_542199.1 T15_G21del	0	1
NM_080632.1 26_165del	Deletes Normal Translation Start Site	0	1
NM_080632.1 42C>T	Before Translation Start Site	0	1
No mutation		0	53

Table 4: Provean prediction of impact of indel mutations on biological function of protein

Variant	*PROVEAN score*	*Result*
P19_A22del	-4.485	Deleterious
R13_P19del	-6.15	Deleterious
V14_A20del	-4.871	Deleterious
T15_G21del	-4.873	Deleterious

Table 5: ATGpr predicted proteins made from CVB iPSC PQV2 clone

No.of ATG from 5'end	*Reliability*	*Frame*	*Identity to Kozak rule A/GXXAT GG*	*Start (bp)*	*Finish (bp)*	*ORF Length (aa)*	*Stop codon found?*	*Sequence*
1	0.27	3	cXXATGc	186	1415	410	Yes	MPEHDYFEFFSNDTSLYPHMYARAYINFKNQEDIILFRD RFDGYVFLDNKGQEYPAIVEFAPFQKAAKKKTKKRDT KVGTIDDDPEYRKFLESYATDNEKMTSTPETLLEEIEAK NRELIAKKTTPLLSFLKNKQRMREEKREERRRREIERKR QREEERRKWKEEEKRKRKDIEKLKKIDRIPERDKLKDE PKIKVHRFLLQAVNQKNLLKKPEKGDEKELDKREKAK KLDKENLSDERASGQSCTLPKRSDSELKDEKPKRPEDES GRDYRERERYERDQERILRERERLKRQEEERRRQKER YEKEKTFKRKEEEMKKEKDTLRDKGKKAESTESIGSSE KTEKKEEVVKRDRIRNKDRPAMQLYQPGARSRNRLCP PDDSTKSGDSAAERKQESGISHRKEGGEE
16	0.13	1	GXXATGa	838	1062	75	Yes	MKKNWTKEKKPRNWTKRISVMKEPVGKVVHCPSVLIA NLKMKNQRDLKMR
12	0.12	3	AXXATGa	486	1415	310	Yes	MTSTPETLLEEIEAKNRELIAKKTTPLLSFLKNKQRMRE EKREERRRREIERKRQREEERRKWKEEEKRKRKDIEKL KKIDRIPERDKLKDEPKIKVHRFLLQAVNQKNLLKKPE KGDEKELDKREKAKKLDKENLSDERASGQSCTLPKRS DSELKDEKPKRPEDESGRDYRERERYERDQERILRERE RLKRQEEERRRQKERYEKEKTFKRKEEEMKKEKDTLR DKGKKAESTESIGSSEKTEKKEEVVKRDRIRNKDRPAM QLYQPGARSRNRLCPPDDSTKSGDSAAERKQESGISHR KEGGEEMYARAYINFKNQEDIILFRDRFDGYVFLDNKG QEYPAIVEFAPFQKAAKK
4	0.12	3	cXXATGt	243	1415	391	Yes	KTKKRDTKVGTIDDDPEYRKFLESYATDNEKMTSTPET LLEEIEAKNRELIAKKTTPLLSFLKNKQRMREEKREERR RREIERKRQREEERRKWKEEEKRKRKDIEKLKKIDRIPE RDKLKDEPKIKVHRFLLQAVNQKNLLKKPEKGDEKEL DKREKAKKLDKENLSDERASGQSCTLPKRSDSELKDEK PKRPEDESGRDYRERERYERDQERILRERERLKRQEEE RRRQKERYEKEKTFKRKEEEMKKEKDTLRDKGKKAES TESIGSSEKTEKKEEVVKRDRIRNKDRPAMQLYQPGAR SRNRLCPPDDSTKSGDSAAERKQESGISHRKEGGEE
13	0.11	3	AXXATGa	594	1415	274	Yes	MREEKREERRRREIERKRQREEERRKWKEEEKRKRKDI EKLKKIDRIPERDKLKDEPKIKVHRFLLQAVNQKNLLK KPEKGDEKELDKREKAKKLDKENLSDERASGQSCTLPK RSDSELKDEKPKRPEDESGRDYRERERYERDQERILRE RERLKRQEEERRQKERYEKEKTFKRKEEEMKKEKDT LRDKGKKAESTESIGSSEKTEKKEEVVKRDRIRNKDRP AMQLYQPGARSRNRLCPPDDSTKSGDSAAERKQESGIS HRKEGGEE
22	0.11	3	GXXATGa	1143	1415	91	Yes	MKKEKDTLRDKGKKAESTESIGSSEKTEKKEEVVKRDR IRNKDRPAMQLYQPGARSRNRLCPPDDSTKSGDSAAER KQESGISHRKEGGEE
2	0.07	1	AXXATGa	196	354	53	Yes	MIILSFFLMIRVCILICMPEHTSTLKTKRTLFCSGIALMV MYSLTIKVRNIPL
24	0.06	1	cXXATGa	1333	1434	34	Yes	MTAPSLEIQQQKGSRKVVLAIEKKEERSDKSRWP
26	0.06	2	GXXATGc	1673	1738	22	Yes	MHCNLQGELSSEDFITYDTREK
14	0.05	2	GXXATGG	671	757	29	Yes	MERRRETKKERYRKAKEDRQNSRKGQIKG
32	0.05	2	AXXATGt	2096	2098	1	Yes	M
3	0.04	1	cXXATGa	220	354	45	Yes	MIRVCILICMPEHTSTLKTKRTLFCSGIALMVMYSLTIK VRNIPL
5	0.04	1	tXXATGc	247	354	36	Yes	MPEHTSTLKTKRTLFCSGIALMVMYSLTIKVRNIPL
6	0.04	1	tXXATGG	310	354	15	Yes	MVMYSLTIKVRNIPL
15	0.04	1	AXXATGa	757	798	14	Yes	MNQRLRYTGFCYKL
7	0.04	1	GXXATGt	316	354	13	Yes	MYSLTIKVRNIPL
17	0.04	1	GXXATGa	898	1062	55	Yes	MKEPVGKVVHCPSVLIANLKMKNQRDLKMRAAETIGR GNGNMNEIRSAYFEKERG
18	0.04	1	AXXATGa	958	1062	35	Yes	MKNQRDLKMRAAETIGRGNGNMNEIRSAYFEKERG
19	0.04	1	AXXATGa	982	1062	27	Yes	MRAAETIGRGNGNMNEIRSAYFEKERG
20	0.04	1	AXXATGa	1021	1062	14	Yes	MNEIRSAYFEKERG

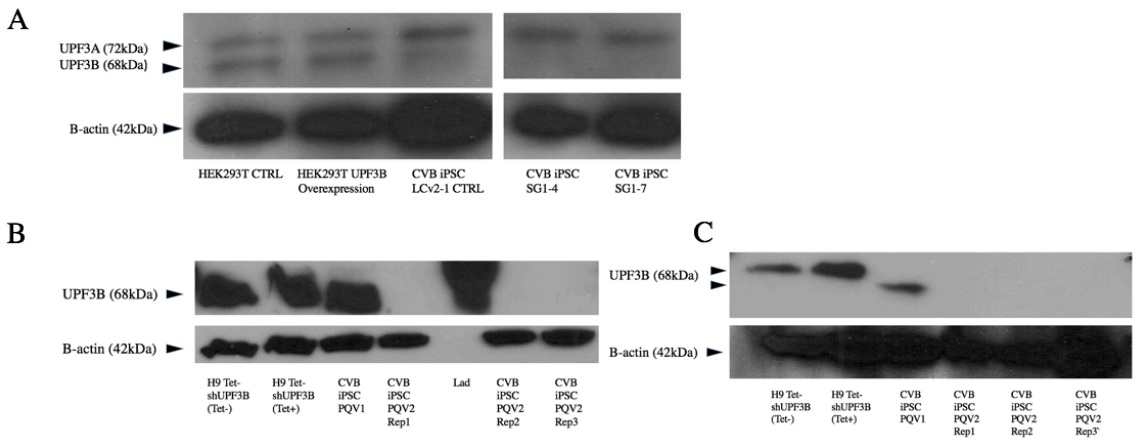

Figure 14: Western blot analysis of CRISPR mediated UPF3B null cells. (A) Shows clones generated by viral approach. HEK293T untreated control protein lysate and HEK293T viral mediated UPF3B overexpression lysate as antibody validation. Control empty virus transduced clone LCv2-1, two UPF3B null clones SG1-4 and SG1-7. (B) and (C) Show clones generated by transient vector approach. H9Tet-shUPF3B(-) control (normal UPF3B expression) and H9 Tet-shUPF3b(+) (UPF3B reduced controls) CVB iPSC PQV1 in-frame AA acid deletion clone and biological replicates of CVB iPSC PQV2, expected deleterious mutation clone. (C) shows same samples as in B at lower concentration. For A 2 replicates were done, B and C 3 replicates were done.

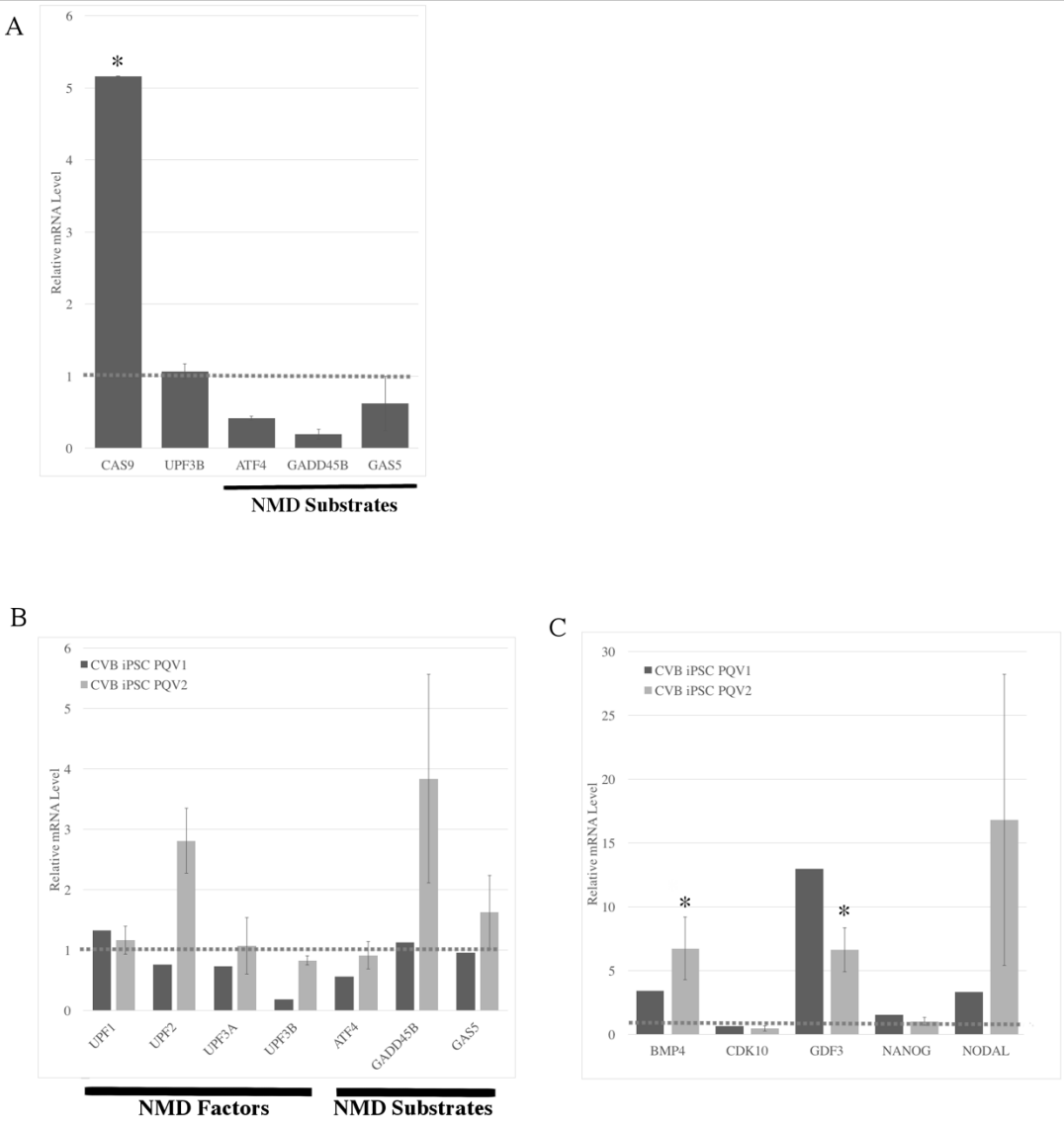

Figure 15: Analysis of CRISPR mediated UPF3B null clones characterizing NMD factors, substrates, and transcripts found to be upregulated in UPF3B null patient iPSC. (A) shows qPCR analysis of viral generated UPF3B null clone SG1-7 where LCv2-1 serves as relative control (B) and (C) show both CVB iPSC PQV1 and 2, where 3 independent biological replicates from untransfected CVB iPSC cell lines serve as relative control. For CVB iPSC PQV2 n=3 biological replicates. Control expression level corresponds to relative mRNA level of 1. * indicates P<.05.

MATERIALS AND METHODS

Description of Cell Lines and Culture Methods

CVB iPSC cell lines were derived from an apparently healthy Caucasian male as described (Gore A, 2010). Cell lines were received at passage 25 generation of clones took place within 5 passages. iPSC were cultured as described in chapter two.

sgRNA Design and Validation

sgRNA sequences were designed to target the CDS of exon 1 of UPF3B, sequence determination was done using the *crispr.mit.edu* web tool. Available from the Feng Zhang laboratory. Several potential guides were generated however only one was ultimately used to carry out the experiment. Guides were successfully ligated into either vector PX458 or Lenticrisprv2 vector available from Addgene (Catalog# 48138 and #52961) (Ran FA, 2013) (Sanjana NE, 2014) using the associated protocols outlined there. Vectors were transfected into HEK293T cells using lipofectamine 2000 for validation which was done using a T7 Endonuclease assay, manufacturers protocol available from NEB (M0302S).

Lentiviral Production and Validation

HEK293T cells were transfected with 2nd generation lentiviral vectors, 3ug pMD2G, 5ug pMDLG/pRRE, 2.5ug pRS-REV, and 10ug of the viral vector LCv2 using lipofectamine 2K. This was incubated for 48 hours upon which media was collected for "collection A" of virus. 72 hours later media was collected as "collection B" of virus. The collected media was filtered using a .45 um Steriflip filter, then virus was concentrated using a 4K Amicon Filter. The concentrated virus was aliquoted and stored at -80degC. To

perform a relative functional titer of the virus HEK293T cells were infected with 1:1000 dilutions of the virus then underwent puromycin selection for 5 days. After which RNA was isolated and screened for Cas9 expression as well as reduction of UPF3B mRNA levels. For each viral prep concentration was chosen such that it was able to produce Cas9 expression to that of endogenous control gene, either RPL19 or GAPDH.

Transduction of iPSC

For a given viral preparation optimum concentration was used to infect a pool of iPSC culture on a 6 well plate. Previous a puromycin kill curve for the CVB iPSC cell type had been done to determine the concentration of puromycin needed to kill all cells not expressing the puromycin resistance gene in 3 days. Cells were then kept in selection conditions for 5-10 days depending on confluency and time of passage. When cells kept under selection conditions reached 70% confluency they were plated sparsely on a 6 well plate to allow the development of independent clones arising from a single cell. ~100-500 cells per plate. Subsequently colonies that arose were manually disassociated using 0.5mM EDTA treatment and a P200 pipette tip and moved to a smaller cell culture dish for screening either 24 or 12 well depending on size of the colony. Afterwards cells were expanded for screening of DNA, RNA, and protein.

Comprehensive Protocol for Transient Vector CRISPR Editing and Single Cell Cloning of iPSC.

The purpose of this sub section is to provide a straightforward protocol with established culture conditions necessary for single cell cloning of iPSC/hES cells. The

reason being there is very few concisely written protocols highlighting the different culture conditions available and providing commentary on improving single cell viability of iPSC which is significantly reduced relative to other cell types.

Routine maintenance of iPSC was done as described earlier, for the purposes of single cell cloning iPSC cells must first be cultured in mTesR1 media as it is the only media suitable for single cell cloning, in addition cells need to be accustomed to growing in it prior to transfection, and last conditioned media is needed for highest post FACS viability. 3 days prior to transfection begin culturing the iPSC in mTesR1 collecting the exhausted media after each day and storing at 4degC. Two days prior to transfection place the cells in mTesR1 supplemented with 10mM Thiazovivin or RhoK inhibitor(RI). This promotes the cells to form a single layer rather than colonies, which is critical for high transfection efficiencies. Alternatively, cells can be cultured on vitronectin coated plates with the same conditions for even higher transfection efficiencies. As vitronectin further supports sparse single cells rather than colony formation (Thomson JA, 1998).

Prior to transfection optimized transfection conditions for the cell line of interest must be determined, using the DNA-In Stem transfection reagent and the associated transfection efficiency protocol that is available from the manufacturer. Taking into consideration toxicity of the cells as a function of too much DNA transfected, as the reagent itself will not kill the cells. Cells are optimum to transfect, when they are at 70% confluency in single cell layer typically after 3 days in RI. In this situation the downstream application of transfection was FACS into single cells for screening, post transfection GFP+ cell number were highest after 24 hours of transfection (~30%) though viability of positively sorted cells was reduced after subsequent culture. 48 hour post transfection GFP+ cells

have significantly decreased (~1-5%), though viability after cell sorting did increase as a result.

Several options exist for recovery of cells after FACS sorting, ideally cells are sorted directly onto a 96 well coated (either with vitronectin or matrigel) plate. Cells can also be seeded sparsely onto large cell culture dishes at densities of ~50 cells per mL of media for optimum clonality and viability. The media in which the cells are grown after FACS sorting should be "cloning media" which consists of ½ fresh mTesR1 ½ conditioned and filtered to remove cell debris mTesR1 supplemented with "SMC4" that consists of 10mM thiazovivin, a Rho Kinase Inhibitor. 1 uM of CHIR99021 a GSK3 inhibitor, 0.4 uM PD0325901 a ERK/MEK inhibitor and 2uM of SB431542 a TGFß inhibitor (Valamehr B, 2012). Initially cells are plated in ½ the volume of media they would normally be grown in for the first day, then complete volume of media is kept for the remainder, changing the media each day, by removing ½ the volume of old media and replacing it with ½ fresh media. Keeping the cells in cloning media for the first 3 days following FACS. Upon day 4 or day 5 cells should be removed from media supplemented with SMC4 and grown in normal medium at this point media changes can be the entire volume each day. Visible colonies for should appear 7-10 days after the initial plating.

The first step in screening the colonies is DNA isolation to determine if successful edits have been made. The most practical way to do this from the initial plating, is to parallel passage onto 2 coated 48 well plates. At which point one plate will be taken to freeze during the screening process while the other will be taken for DNA isolation. The plate taken for freezing is grown to 70% confluency in the majority of wells. Once that has occurred cells are disassociated from the matrix using EDTA as previously described but

are resuspended in freezing media rather than normal media. For freezing directly in the plate it is best to use a commercial iPSC/hES freezing solution. Cryo-stor in this case was used with high success and very good recovery upon thawing. Once all cells are resuspended in the freezing media the plate is wrapped heavily in parafilm then placed in Styrofoam chamber at -80degC and kept there until it is needed.

For the plate taken for DNA it can be grown to 100% confluency as this will allow for the most DNA recovered and it will not impact any results. Once the majority of these cells have reached 100% confluency, the simplest way to isolate the DNA from several samples is a simple salting out procedure carried out overnight. Here (Miller SA, 1998) was adapted to meet the volume of a 48 well plate, and initial incubation in lysis buffer was done overnight at 37degc in the incubator. The remainder of the DNA isolation was carried out as outlined (Miller SA, 1998).

Screening was done by PCR amplification with primers flanking the region of interest by 250 bp on either side, followed by gel purification of fragments, and sanger sequencing. Though direct sequencing from PCR product can also be done as well. Depending on guide design, or nature of the genome edits other screening methods may be more feasible.

When successful clones have been identified you should take time to screen all possible clones because when thawing the cells an entire plate must be thawed at a time. The whole plate is placed from the Styrofoam chamber into the 37 deg C incubator and kept there for 5-10 minutes until the majority of the media has thawed with a small amount of ice still visibly present. Cells are diluted with complete volume of DMEMF12 media, which is then transferred to a microcentrifuge tube to pellet cells for 5 minutes at 1000

rpm. Cells are resuspended in mTesrR1 supplemented with 10mM RI and plated onto either vitronectin or matrigel coated plates. Which can then be expanded for freezing stocks or subsequent screening.

Western Blotting

Whole protein lysates were isolated from cells using RIPA buffer, and protein concentration estimated using the Biorad Bradford reagent. Appropriate amounts of each protein were added to 10% SDS-PAGE gel and transferred to a PVDF membrane. UPF3B amount was determined using a full length antibody targeting UPF3B which can cross react with UPF3A, produced in rabbit. Figure 14A appears to have such a cross reaction, while Figure 14B and C do not. Blots were probed with secondary raised in mouse conjugated to HRP. WestPico Signal substrate was used to image for a 10 minute exposure in each of the blots shown.

RNA Isolation and qPCR

Was done as previously described in chapter two.

TRANSCRIPTOME ANALYSIS OF *UPF3B*-NULL PATIENT CELLS AND *UPF1*-DEPLETED EMBRYONIC STEM CELLS

INTRODUCTION

Previous genome-wide studies of the effect of *UPF3B*-null mutations have largely relied on cell lines with limited clinical relevance. In the case of Nguyen *et al*., RNAseq was done in lymphoblastoid cell lines derived from patients (Nguyen LS, 2012). This study revealed that a potential NMD target, *ARGHAP24* (associated with actin skeleton remodeling (Nguyen LS, 2012)), was dysregulated upon loss of UPF3B. This transcript had been validated as a NMD substrate by other *UPF3B* knockdown experiments done in HeLa cells (Chan WK, 2007) (Chan, 2009). *In situ* hybridization done in developing human brain tissue found that as neural differentiation occurs, *UPF3B* levels reduce while that of target gene *ARGHAP24* increase.

Considering NMD acts in a cell type-specific manner, the transcriptome profile generated from lymphoblastoid cells from the above Nguyen *et al*., study may not reveal insights into the physiological or molecular impact of loss of *UPF3B* in patients. As previously mentioned, Lou *et al*. performed transcriptome profiling of *UPF1*-depleted hES cells and found that reduction of NMD perturbed the levels of several transcripts involved in pathways governing early cell germ layer specification and cell fate decision making (Lou CH, 2016). This study highlighted the importance in studying transcriptional profiles generated in relevant cell types to reveal insights on affected tissues.

Another limitation of early studies using "transcriptome profiling" to examine the effect of loss or reduction of NMD factors, is previous studies have identified differentially expressed genes rather than differentially expressed transcripts (Nguyen LS, 2012) (Lou CH, 2016). This is a result of how the data is analyzed, where reads from the data set are mapped to the genome, rather than the transcriptome. Given that degradation by NMD is

sensitive to both aberrant transcripts and alternative splicing (Karousis ED, 2016), differentially expressed genes likely represent differentially expressed transcripts as a result of reduction or loss of NMD but the reciprocal is not true. There are likely several transcripts associated with a given gene locus that are differentially expressed, but would be missed when mapping to a genome rather than a transcriptome. Though genomic-based approaches may provide a more simplistic view on the overall physiological implication of loss of NMD factors, it does not fully capture the impact on the transcriptome. Considering that NMD is a post transcriptional mRNA regulatory pathway, bioinformatic approaches profiling NMD deficient cells may provide better insights through direct analysis of the transcriptome. More recent NMD studies have taken this approach, utilizing transcriptome-based analysis to determine degree of overlap between potential branches of the NMD pathway and to gain insights into the nature of the transcripts degraded by NMD (Colombo M, 2016).

Another potential way to provide mechanistic insights into the potential functions of NMD in the developing human brain is not only to look at the transcripts dysregulated as a result of loss of any of the factors, but also to look at the way in which these factors and potential NMD targets are expressed in a temporal fashion during development (Chou SJ, 2016). To date it has been found in mouse and rat that NMD activity is highest in the pluripotent stem cell and decreases as neural differentiation occurs, however this has yet to be determined in humans (Alrahbeni T, 2015) (Lou CH, 2014). As mentioned previously, it has been found that the expression of several NMD factors are the highest in pluripotent cells relative to other cell types in humans (Lou CH, 2016). However, this previous study did not look at expression levels at different cell stages. For example, this

information is important to provide insight as to what stage(s) in neurodevelopment NMD, are most needed.

In this chapter, I provide evidence that, in humans, NMD and EJC core factors expression are the highest in pluripotent cells and are downregulated during neurodevelopment. In addition, I identify transcripts dysregulated as a result of loss of UPF3B in patient iPSC that may be responsible for generating neurodevelopmental disorders. I propose that transcriptomic analysis of NMD-deficient hES cells has the potential to provide mechanistic insights to the impact of loss of UPF3B in neurodevelopment.

RESULTS AND DISCUSSION

NMD and EJC Core Factor Expression Levels Decrease During Neural Differentiation

In order to determine the expression of NMD and EJC core factors during hES differentiation, I made use of publically available RNASeq data sets. In particular, I analyzed pluripotent stem cell, neural progenitor cells, differentiated neurons, and fully differentiated neurons for relative mRNA expression values, as determined by Reads Per Kilobase of transcript per Million mapped reads (RPKM). NMD and EJC core component RNAs were analyzed. Table 6 summarizes the RNAseq data sets used and the associated Gene Expression Omnibus (GEO) accession numbers (Marchetto MC, 2016) (Mariani J, 2015) (van de Leemput J, 2014). Studies shown were chosen for sufficient size at each time point for statistical analysis (3 technical replicates for all 2 to 3 biological replicates).

Table 6: Summary of public data sets used for analysis

Study	***Gene Expression Omnibus (GEO) Accession***	***Time Scale***	***Size***
Cortecon	GSE56796	hES to Cortical Neuron	n=2,3 each timepoint
SRR19508	GSE67528	iPSC to Neuron	n=3 each timepoint
SRR15764	GSE61476	NPC to Organoid	n=3 each timepoint

Shown in Figure 16-18 are heat maps detailing the relative expression levels of NMD and EJC core transcripts available from the Cortecon, SRR19508, and SRR15764 data sets, respectively. In the case of the Cortecon data shown in Figure 16, NMD and EJC transcripts have higher expression at D0 and D7 (pluripotent cells and neural induction) relative to D77 (differentiated neurons). An exception is *SMG6*, which has an unusual expression pattern – first increasing and then decreasing (Figure 16). Interestingly, Figure

16 shows 3 NMD transcripts with expression highest upon neural induction – *UPF3B*, *SMG6*, and *SMG7*. Figure 16B largely replicates the trend seen in Figure 17; fewer time points for Figure 17 make direct comparison difficult, however the trend with *SMG6* being lowest at pluripotency upholds, as well as other transcripts having highest expression during pluripotency and decrease as neurons form. Figure 18 replicates the same trend, however, note that D0 corresponding to the neural progenitors, not pluripotent cells. In all 3 studies, the expression of NMD and EJC core factors decreases as differentiation into neurons occurs.

Temporal Expression of Potential NMD Targets

Considering most NMD factors and EJC core transcripts follow roughly the same expression pattern during neurodevelopment, this permitted further analysis of normal temporal expression patterns of potential NMD targets. In order to infer a potential role of *UPF3B* during neurodevelopment several subsets of genes found to be differentially expressed upon loss of *UPF3B* in patient iPSC were analyzed for their normal temporal expression patterns in control data sets. In the situation where a subset of transcripts that are potentially regulated by *UPF3B* show the same temporal expression pattern it could suggest what context *UPF3B* is acting during neurodevelopment. For example, in the situation where all potential *UPF3B* targets have the highest expression during pluripotent stem cell stage, this would suggest *UPF3B* acts to regulate the pluripotent stem cell stage by preventing overexpression of these transcripts. In the opposite case, where the expression of potential *UPF3B* regulated targets are lowest in the pluripotent stem cell

stage this would suggest that *UPF3B* actively degrades a subset of transcripts that promote neurodevelopment.

Potential *UPF3B* regulated targets were broken down into groups by their NMD-inducing features, with intron in the 3'UTR being the best predictor to determine if a transcript will be degraded by NMD (Colombo M, 2016). However, there was no clearly identifiable expression pattern found in potential direct UPF3B target transcripts (data not shown). The closest to a clear pattern is shown in Figure 19, which are overlapping transcripts that were found to be upregulated in patient cell lines and also bound by phosphorylated UPF1 (Kurosaki T, 2014). Transcripts bound by phosphorylated UPF1 are considered strong candidates to be direct NMD targets, as UPF1 is phosphorylated by SMG1 prior to a UPF1 bound transcript being degraded (Karousis ED, 2016). The clustering has been removed to aid visual analysis; there would appear to be two distinct subsets of expression patterns. One where expression is the highest in the pluripotent state and decreases along with NMD and EJC core factors. The other has the opposite pattern, where expression is the lowest in the pluripotent stem cell state and is upregulated upon reduction of NMD and EJC factor expression.

Transcriptome Sequencing of NMD Reduced Pluripotent Stem Cells

The transcripts shown in Figure 19 were derived from gene level differential expression analysis. However, transcript level differential expression analysis was needed to directly determine all dysregulated transcripts associated with a given gene locus. As NMD is a post-transcriptional RNA regulatory pathway transcript level differential expression analysis is most appropriate to describing the impact of reduction of NMD.

To address this issue, I reanalyzed existing RNAseq data sets for the *UPF3B*-null patients iPSC as well as si*UPF1*-mediated *UPF1*-depleted hES cells from Lou *et al.* (Lou CH, 2016). This allowed me to identify differentially expressed transcripts (not only differentially expressed genes). As a preliminary step, the different types of transcripts dysregulated were analyzed. I found that the majority upregulated transcripts in both *UPF3B*-null patient iPSC and *UPF1*-depleted hES cells were "protein coding" (Tables 7 and 9). In the case of *UPF3B*-null iPSC, upregulated retained intron transcripts were found to be more prevalent than in down regulated transcripts. In the case of upregulated transcripts in *UPF1* knockdown hES, lincRNAs appear to more abundant relative to the downregulated transcripts.

Shown in Figure 20 is a summary of the differentially expressed transcripts identified in patient iPSCs relative to control iPSCs. The functional implications of dysregulation were inferred with Gene Ontology (GO) analysis (Figure 20 A and B). Interestingly, significantly enriched in the upregulated group, which contains transcripts likely to be directly regulated by *UPF3B*, are "actin skeleton organization" and several categories associated with "RNA processing." This raises the possibility that *UPF3B* branch of the NMD pathway directly regulates other RNA processing pathways, perhaps to buffer them, in accordance with my data in Chapter 2. Enriched in mRNAs downregulated in patient iPSCs are those encoding proteins associated with "neural development" and "axon formation" (Figure 20B).

To interpret results obtained from NMD-deficient patient iPSCs, I examined functional categories enriched among transcripts dysregulated in NMD-deficient (*UPF1*–depleted) hES cells. Figure 20C and D shows the GO categories found for both upregulated

and downregulated transcripts upon reduction of *UPF1* in hES. Interestingly, upregulated transcripts encode proteins involved in "muscle organ development" and "blood vessel morphogenesis;" both terms corresponding with mesoderm formation. Transcripts downregulated by *UPF1* depletion were significantly enriched in another mesoderm-associated term – "embryonic cranial skeleton morphogenesis." The presence of enriched GO terms for mesoderm in *both* upregulated and downregulated transcripts upon reduction of *UPF1* in hES is not consistent with the previous finding that NMD promotes the formation of mesoderm (Lou CH, 2016).

Enriched GO terms associated with transcripts both upregulated in patient iPSC (i.e., likely triggered by loss of *UPF3B*) and upregulated in *UPF1*-depleted hES cells (i.e., high-confidence NMD targets) are "RNA metabolism" and "alternative splicing." This finding is consistent with the possibility that NMD regulates other RNA regulatory pathways; it will be interesting in the future to define the post-transcriptional networks that NMD participates in. This is interesting considering that transcripts dysregulated by reduction in NMD magnitude either by loss of *UPF3B* or reduction of *UPF1* were able to produce an enriched GO term associated with germ layer specification individually, while overlapping dysregulated transcripts between the two groups did not upregulated by both loss of *UPF3B* or reduction of *UPF1* did no*t*.

My results have several potential implications. First, they raise the possibility that the indirect effects of UPF3B impact early formation of the ectoderm, with several neural-associated GO term being enriched. Second, putative direct targets of the UPF3B-dependent branch of NMD encode proteins involved in RNA metabolism. Third, reduction of *UPF1* in hES dysregulates transcripts whose encoded proteins are associated with

mesoderm formation. This is evidence to suggest that direct effects of NMD in early germ layer specification are directly responsible mesoderm formation as *UPF1* is needed for all branches of NMD (Karousis ED, 2016).

Temporal Expression of Neural-Associated UPF3B-Regulated Transcripts

Next I examined the temporal expression patterns of UPF3B-regulated transcripts associated with the neural-associated GO terms. I did this as a means to elucidate how UPF3B may impact neurodevelopment. Shown in Figure 21-23 are heat maps for the temporal expression pattern of *UPF3B* patient-downregulated neural-associated transcripts from the Cortecon, SRR19508, and SRR15764, data sets, respectively. Figure 21 shows that the majority of those transcripts have their lowest expression during the pluripotent stem cell stage and their expression increases during neuronal differentiation (with the exception of *ALS2, ILK,* and *FGFR1*). This trend upholds (with the exception of *UNC5A)* in Figure 22 which, depicts expression of *UPF3B* patient-downregulated neural-associated transcripts from pluripotent stem cell to fully differentiated neuron in control cell lines from the SRR19508 study. The trend is also supported (with expectation of *METRN*) by Figure 23 which, shows *UPF3B* patient-downregulated neural-associated transcripts expression pattern from neural progenitor cells differentiated into neural tissue organoids in the SRR15764 study.

In order to define significant changes in expression between time points and infer temporal expression trends, appropriated statistical analysis, such as EBSeqHMM (Leng N, 2015), is required. Unfortunately, I was unable to complete this within the time frame

of my Master's degree. Another future goal will be to examine the temporal expression of datasets from individuals with neurodevelopmental disorders. The analysis I did was from control cells from normal individuals; the studies I used for my analysis also had datasets from individuals with neurodevelopmental disorders, including autism spectrum disorders. This would be particularly exciting to do given that many patients with *UPF3B*-null mutations have autism spectrum disorder (Addington AM 2011). Thus, loss of *UPF3B* may contribute to this neuro-developmental disorder in particular genetic contexts. Analysis of the expression pattern of UPF3B-regulated transcripts in autistic patients may reveal candidate transcripts whose misregulation in *UPF3B*-mutant individuals causes one more symptoms in autism spectrum disorder.

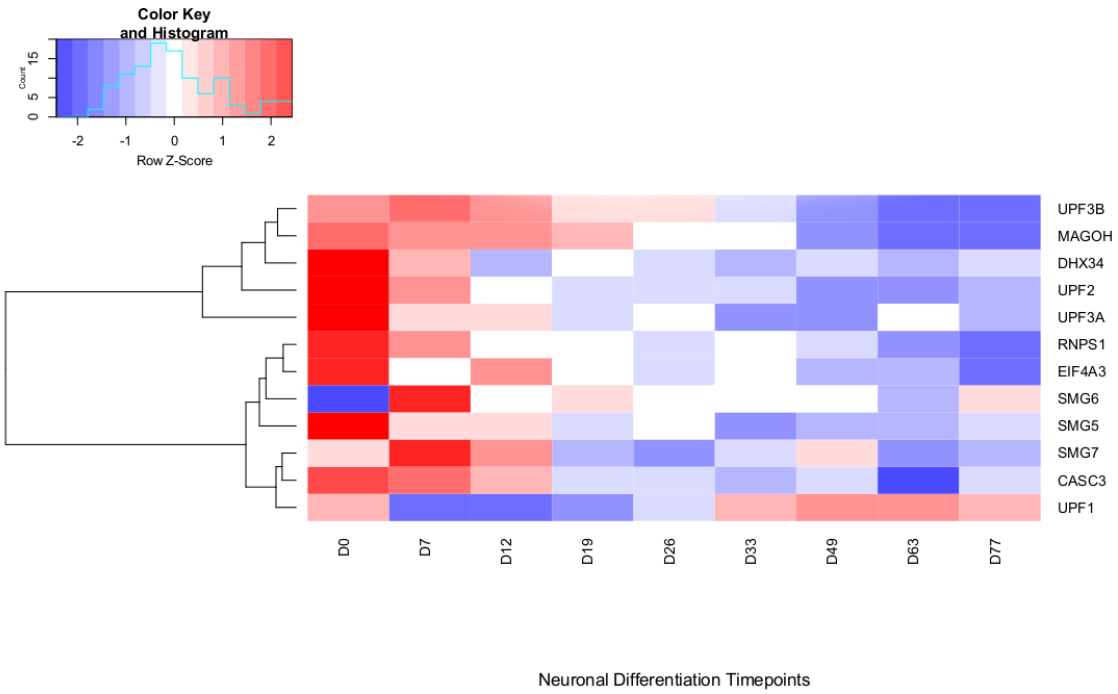

Figure 16: Temporal expression of NMD and EJC core factors during neurodevelopment in Cortecon data set. Log2 transformed RPKM values set at time points starting at pluripotency (D0) extending to fully differentiated neuron (D77). Of note, D7 corresponds with the day following neural induction by dual SMAD inhibition and D19 corresponds with established neural progenitor cells. For each time point, n=2,3. Ward method hierarchical clustering of Euclidean distance shown in dendrogram on left, within chosen subset of genes.

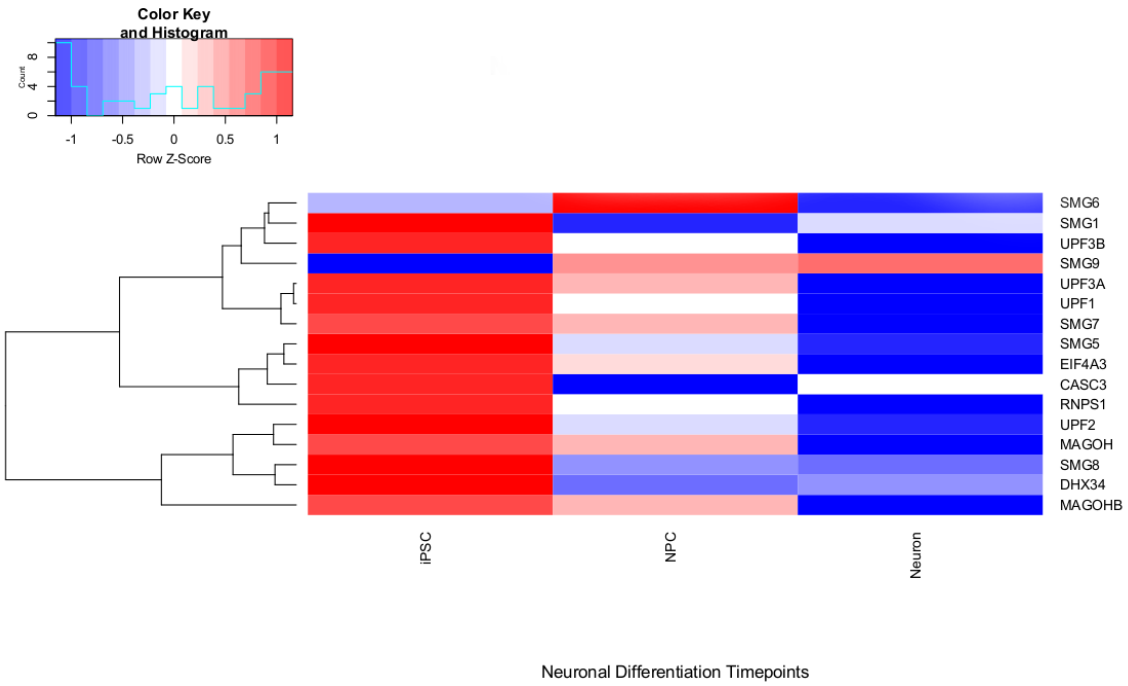

Figure 17: Temporal expression of NMD and EJC core factors during neurodevelopment in SRR19508 data set. Shows Log2 transformed RPKM values set at time points starting at iPSC cell state extending to Neuron) for each time point n=3 biological replicates. Ward method hierarchical clustering of Euclidean distance shown in dendrogram on left, within chosen subset of genes.

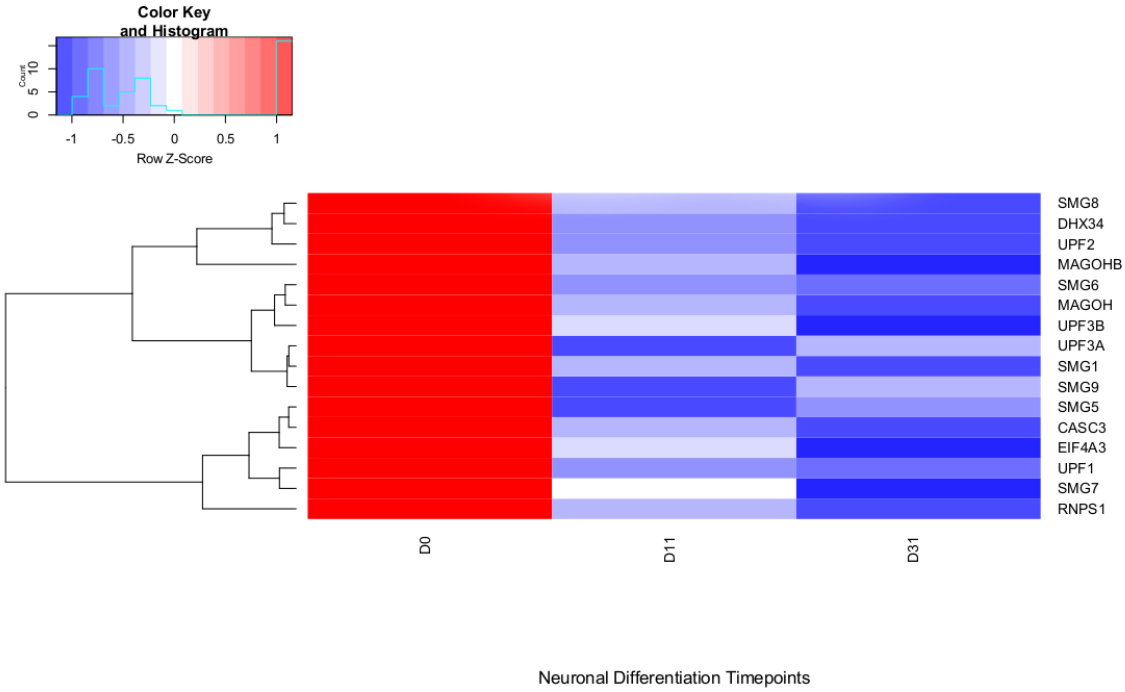

Figure 18: Temporal expression of NMD and EJC core factors during neurodevelopment in SRR15764 data set. Shows Log2 transformed RPKM values set at time points starting at NPC cell state extending to formed organoid culture. For each time point n=3 biological replicates. Ward method hierarchical clustering of Euclidean distance shown in dendrogram on left, within chosen subset of genes.

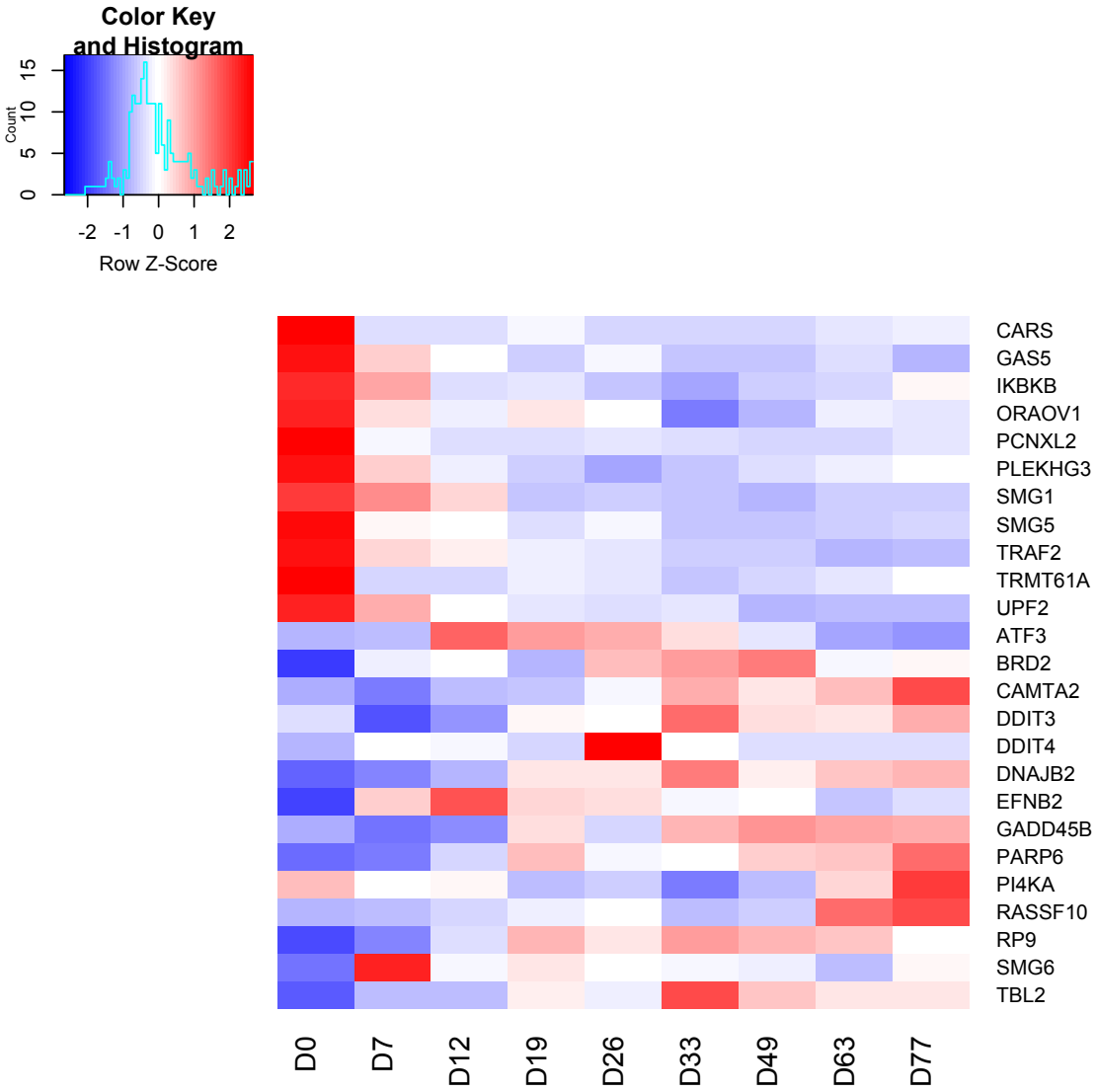

Figure 19: pUPF1 targets as a measure for NMD activity during neuronal development. Log-fold 2 RPKM values of pUPF1-bound and patient upregulated overlapping transcripts. From the Cortecon data set RNAseq of the developing human cortex.

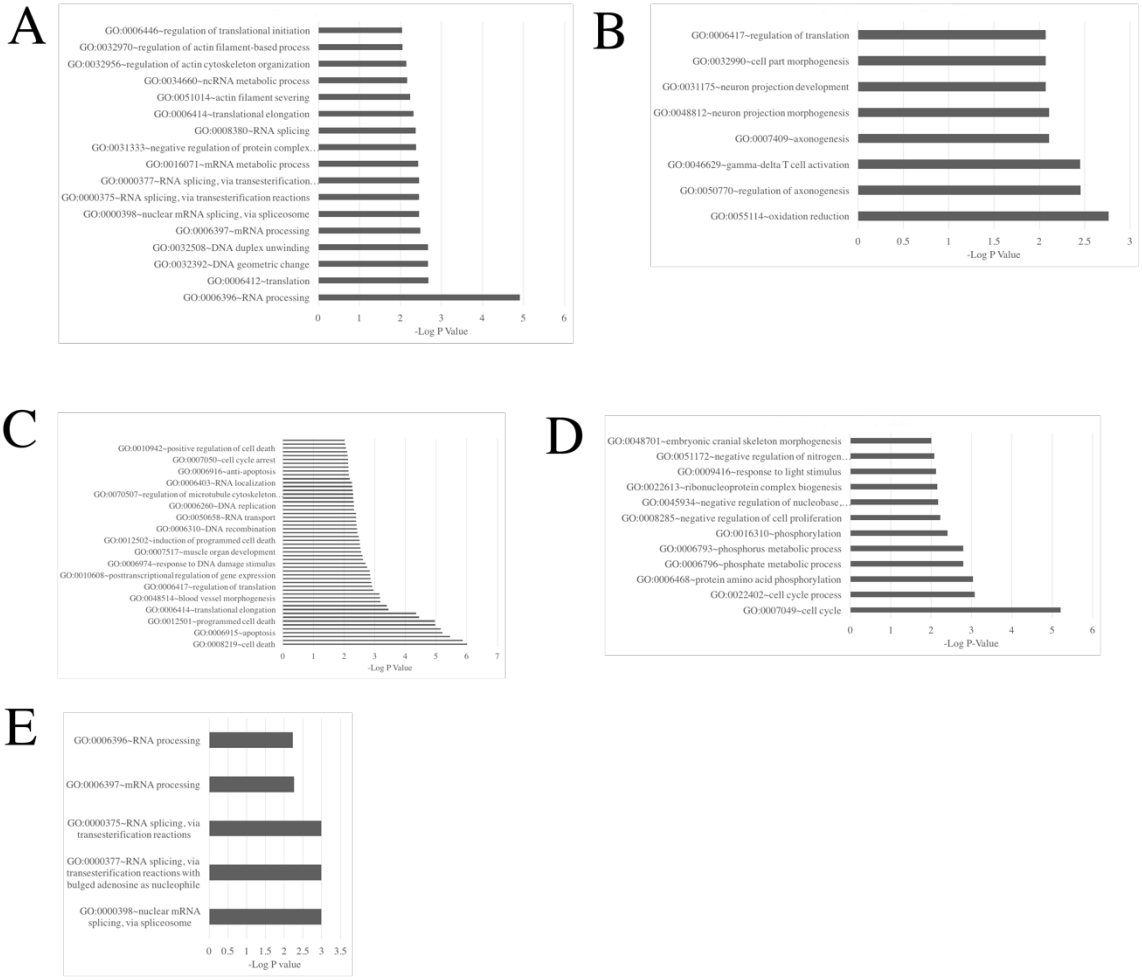

Figure 20: GO analysis of transcripts found to be significantly upregulated or down regulated in patient iPSC and *UPF1* depleted hES. (A) upregulated in patient, (B) downregulated in patient. (C) *UPF1* depleted upregulated (D) *UPF1* depleted downregulated. (E) Overlapping transcripts upregulated in both patient iPSC and *UPF1* depleted cells.

Table 7: Summary of transcript types upregulated by loss of *UPF3B* in patient iPSC

Transcript Type	***Transcript Percentage***
protein_coding	46.52%
retained_intron	24.21%
processed_transcript	11.55%
nonsense_mediated_decay	8.23%
lincRNA	3.48%
antisense	3.16%
processed_pseudogene	0.71%
TEC	0.71%
transcribed_unprocessed_pseudogene	0.47%
unprocessed_pseudogene	0.40%
transcribed_processed_pseudogene	0.24%
Mt_rRNA	0.08%
sense_overlapping	0.08%
non_stop_decay	0.08%
sense_intronic	0.08%
Total Transcript Count	1264

Table 8: Summary of transcript types downregulated by loss of *UPF3B* in patient iPSC

Transcript type	***Transcript Percentage***
protein_coding	66.87%
processed_transcript	10.83%
retained_intron	9.41%
nonsense_mediated_decay	7.64%
lincRNA	2.13%
antisense	1.24%
processed_pseudogene	0.62%
sense_overlapping	0.27%
non_stop_decay	0.27%
unprocessed_pseudogene	0.18%
transcribed_processed_pseudogene	0.18%
transcribed_unprocessed_pseudogene	0.09%
sense_intronic	0.09%
misc_RNA	0.09%
TEC	0.09%
Total Transcript Count	1126

Table 9: Summary of transcript types found to be upregulated upon reduction of *UPF1* in hES

Transcript type	***Transcript Percentage***
protein_coding	55.35%
nonsense_mediated_decay	17.45%
processed_transcript	12.11%
retained_intron	8.09%
antisense	1.97%
lincRNA	3.43%
transcribed_unprocessed_pseudogene	0.50%
snoRNA	0.06%
unprocessed_pseudogene	0.16%
processed_pseudogene	0.25%
transcribed_processed_pseudogene	0.09%
sense_overlapping	0.09%
non_stop_decay	0.09%
TR_C_gene	0.06%
sense_intronic	0.03%
misc_RNA	0.09%
TEC	0.12%
Total Transcript Count	3202

Table 10: Summary of transcript types found to down regulated upon reduction of *UPF1* in hES

Transcript type	*Transcript Percentage*
protein_coding	70.02%
nonsense_mediated_decay	6.01%
retained_intron	12.15%
processed_transcript	8.49%
antisense	2.16%
lincRNA	0.78%
processed_pseudogene	0.07%
sense_overlapping	0.07%
transcribed_unprocessed_pseudogene	0.07%
non_stop_decay	0.07%
TEC	0.07%
sense_intronic	0.07%
Total Transcript Count	1531

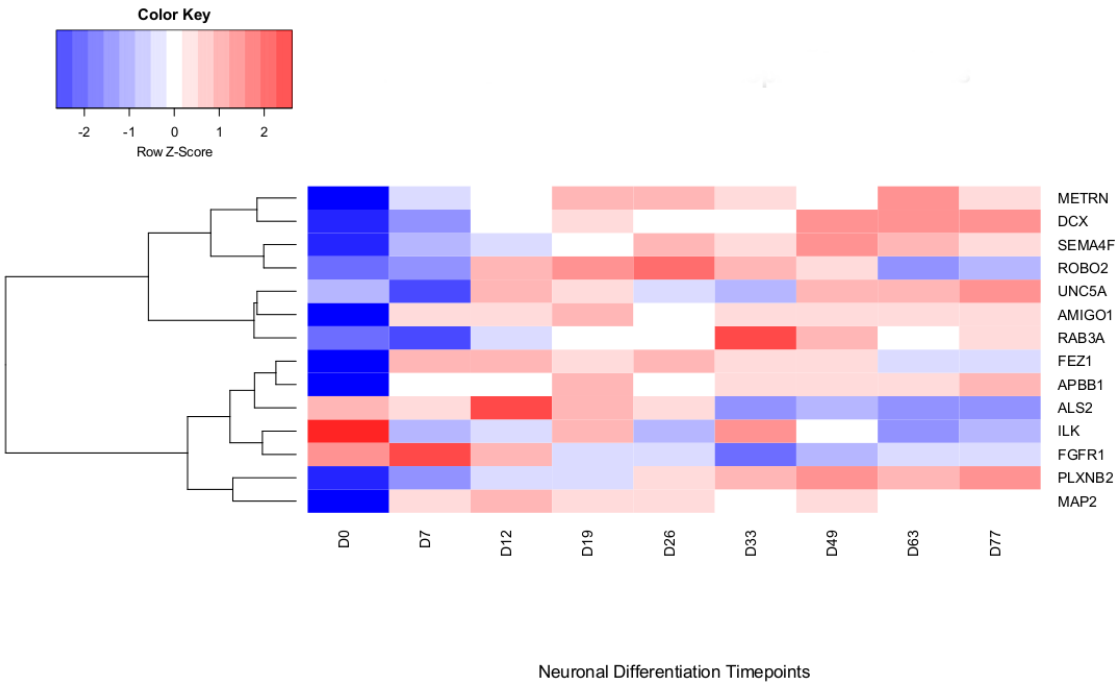

Figure 21: Temporal expression of neural associated dysregulated transcripts in patient iPSC analyzed in control Cortecon data set. Shows Log2 transformed RPKM values. For each time point n=3 biological replicates. Ward method hierarchical clustering of Euclidean distance shown in dendrogram on left, within chosen subset of genes.

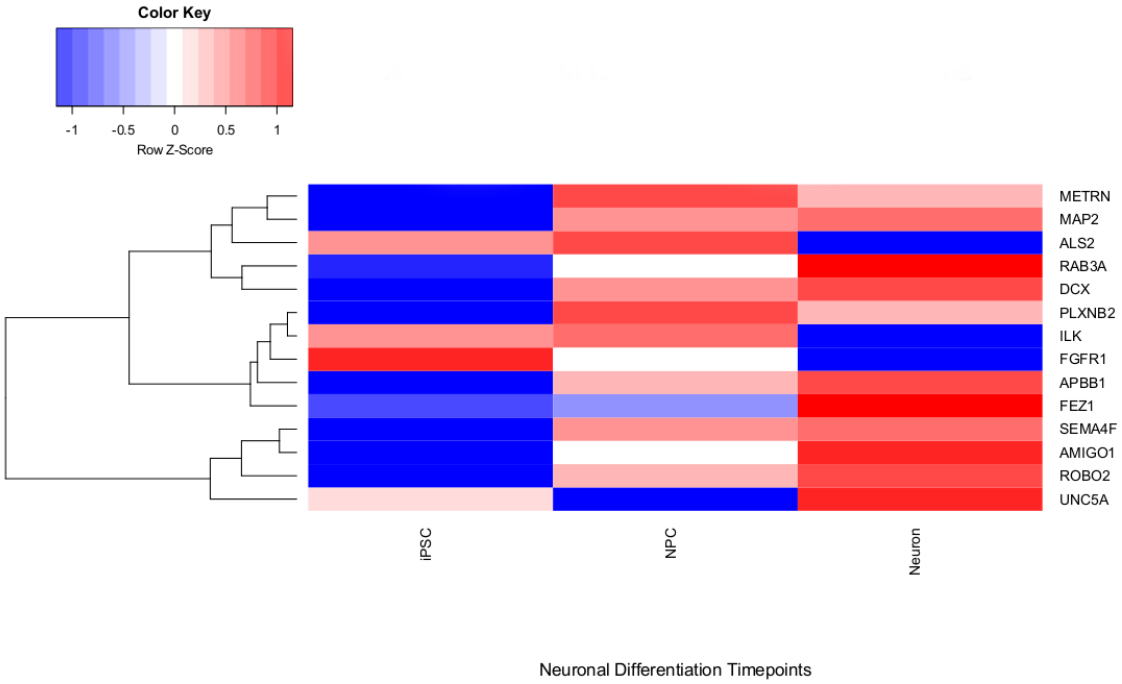

Figure 22: Temporal expression of neural associated dysregulated transcripts in patient iPSC analyzed in control SRR19508 data set. Shows Log2 transformed RPKM values. For each time point n=3 biological replicates. Ward method hierarchical clustering of Euclidean distance shown in dendrogram on left, within chosen subset of genes.

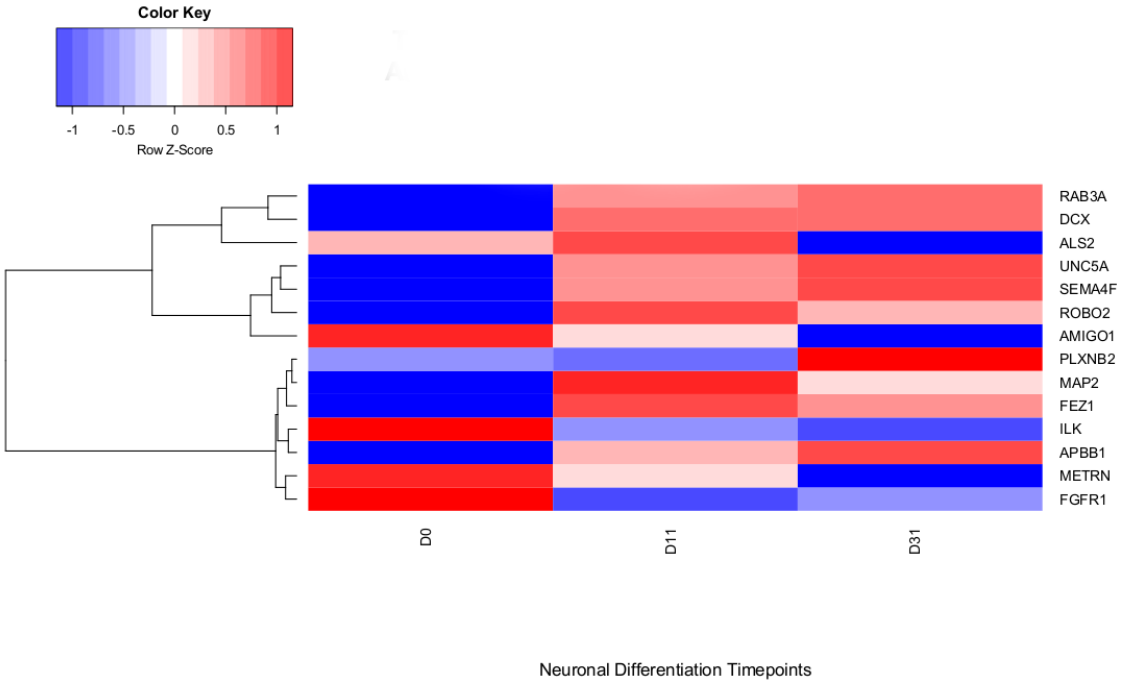

Figure 23: Temporal expression of neural associated dysregulated transcripts in patient iPSC analyzed in control SRR15764 data set. Shows Log2 transformed RPKM values. For each timepoint n=3 biological replicates. Ward method hierarchical clustering of Euclidean distance shown in dendrogram on left, within chosen subset of genes.

Closing Remarks and Model

A previous study suggested that the NMD pathway promotes the formation of mesoderm and inhibits endoderm differentiation via regulation of well-established signaling pathways (Lou CH, 2016). However, the impact of NMD on ectoderm differentiation and the specific roles of UPF3B in germ layer formation have been unclear. My data raises the possibility of another role for NMD – maintenance of robustness during early germ layer formation. In other words, NMD functions to promote canalization, which is the tendency to progress through development to create normal adult status when faced with perturbing outside influences. As defined by Conrad Waddington in his ground-breaking work of the 1950s (Waddington CH, 1956) (Waddington CH, 1959). My evidence for this was the finding that patient iPSCs that lack UPF3B exhibit greater variance in expression of pluripotency, endodermal, and mesodermal markers, as compared to control iPSCs that express UPF3B. In contrast, I found that ectoderm-associated markers do not exhibit more variable expression in patient iPSCs relative to control iPSCs. This suggests that UPF3B functions to maintain cell identity. For example, in pluripotent cells, loss of UPF3B leads to variable expression of pluripotency markers, suggesting that these cells are not stably maintained in the undifferentiated cell state. Indeed, I observed variable expression endoderm and mesoderm differentiation markers, indicative of failure to maintain the pluripotent state. I postulate that another consequence of this loss of cell identity is a reduced capacity to form ectoderm. Evidence in support of this is my finding that patient iPSCs exhibited reduced proliferation relative to control iPSCs and they exhibited latent expression of pluripotency markers when pushed down the ectodermal lineage. Further evidence that UPF3B is critical for ectoderm differentiation is my finding

that patient iPSC have reduced expression of transcripts associated with neuronal development. In control iPSCs, these transcripts are expressed lowest at the pluripotent stem cell stage and increase as differentiation into neurons occurs. This highlights that the indirect effects of loss of *UPF3B* contribute to early capacity to differentiate into ectoderm.

NMD is known to maintain the pluripotent cell state by proper regulation of endoderm and mesoderm differentiation (Lou CH, 2016). My results lead me to posit a model in which the UPF3B-dependent branch of NMD acts to buffer expression levels of pluripotency, endoderm, and mesoderm fate specifiers. In this process UPF3B is conferring full strength NMD to allow proper formation of the three primary germ layers (Figure 24). In the absence of UPF3B, weakened NMD results in greater variance in expression of factors controlling pluripotency, as well as endoderm and mesoderm fate, leading to defects in cell identity. Thus, pluripotent cells are not stably maintained in this state under non-differentiation signals, as shown by variability in pluripotency marker genes. Likewise, *UPF3B*-deficient cells fail to respond normally to differentiation signals and only inefficiently undergo differentiation, as evidenced by variability in expression in differentiation markers.

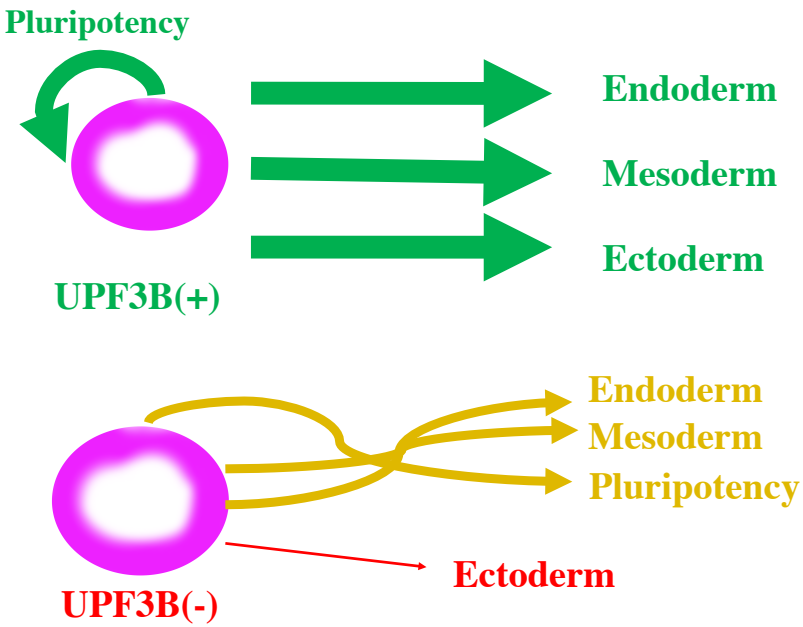

Figure 24: Role of UPF3B in neurodevelopment: At the pluripotent stem cell stage UPF3B acts to confer proper NMD magnitude the net result of which is proper governing of pluripotency, endoderm, and mesoderm cell fate and full formation of all primary germ layers. Shown top. In absence of UPF3B NMD magnitude is decreased the net result of which is improper governance of germ layer specification and a decreased magnitude to generate ectoderm.

THE *SMG6* GENE HARBORS HUMAN ACCELERATED REGIONS

INTRODUCTION

Human Accelerated Regions (HARs) are regions of the genome that have undergone rapid evolution in humans, relative to chimpanzees, but remain highly conserved amongst other vertebrates (Pollard KS, 2006). Many HARs correspond with enhancer regions associated with development (Capra JA, 2013). First identified in 2006, recent studies have found a significant association between HARs and neurodevelopmental disorders, including autism spectrum disorders (Doan RN, 2016). It is thus tempting to speculate that such HARs harbor important mutations that give rise to higher cognitive function and larger brain size relative to early hominids (Doan RN, 2016).

Given the growing evidence that NMD has roles in neural development (Nguyen LS, 2012) (Lou CH., 2014) (Alrahbeni T, 2015). I hypothesized that one or more NMD factor genes contain HARs. Such HARs could alter the regulation or function of such NMD factors in a manner that elicits human-specific neural development. Here, I provide promising evidence that one NMD factor gene, *SMG6*, harbors two HARs; a third HAR maps to a nearby locus. Given that *SMG6* encodes an endonuclease critical for degrading NMD substrates (Karousis ED, 2016), this suggests that selective forces have caused endonucleolytic mRNA decay (and hence NMD) to evolve rapidly between early hominids and modern-day humans.

RESULTS AND DISCUSSION

Three SMG6 Associated HARs

Doan *et al.* contains a concise summary of all HARs found to date (Doan RN, 2016). Using this publication and a HAR database provided by Dr. Pollard (U.C.S.F.) (Pollard KS, 2006), a list was generated to screen NMD and EJC core factor genes for HARs. I considered any HAR within 150 kb of a NMD/EJC gene as a positive hit. Using this approach, I found three HARs associated with SMG6, two of which map directly within the *SMG6* gene (Figure 26). HAR123 is intronic; HAR53 is in the first exon; and HAR471 is ~150 kb of the *SMG6* gene locus. I also screened *UPF1, UPF2, UPF3A, UPF3B, SMG1, SMG5, SMG7, SMG8, SMG9, EIF4A3, Y14, MLN51, MAGOH, MAGOHB, DHX34, BTZ, CASC3, RNPS1, PYM,* and *NAG* (Karousis ED, 2016). I did not identify any HARs associated with any other NMD or EJC genes.

HAR123 and HAR471 Have Characteristics of Astrocyte Enhancers

As most HAR regions correspond with developmental enhancers it was of interest to determine whether HAR123 might have characteristics of an enhancer, given its intronic location. Figure 27 shows that HAR123 lines up with a putative enhancer region harboring DNase I and H3K27 acetylation sites present in astrocytes. This raises the possibility that HAR123 is an astrocyte enhancer. HAR471 also maps to a putative astrocyte enhancer region (Figure 28).

Non-conservative codon mutations in HAR53

HAR53 is in the coding region of SMG6 and thus it has the potential to alter the SMG6 protein. Indeed, the predicted amino-acid sequence of human *SMG6* has a region with multiple amino acid differences from other vertebrate species (Figure 29). This HAR encompasses amino acids 307-440 of human SMG6; it corresponds to a region of unknown function downstream of the 2 EJC interaction domains (AA 39-59 and 133-153) and before the 14-3-3 domain (AA 576-816) (Karousis ED, 2016). This region has 4 amino-acid differences between humans and chimpanzees; 2 out of 4 are non-conservative amino-acid changes. One amino-acid substitution—K-420 (S)—while different between human and chimpanzees (*Pan. Troglodytes*), is not conserved through all vertebrates (Table 11). Thus, the relevance of this particular amino-acid substitution to human-specific evolution is unclear.

Implications and Future Studies

The identification of two HARs mapping directly to the *SMG6* gene locus, as well as one within ~150 kb of the *SMG6* gene raises the possibility that *SMG6* has been subject to strong selective forces in the hominid lineage. Given that HARs appear to most often correspond to transcriptional regulatory elements that act in development (Pollard KS, 2006), the 2 HARs in the non-coding region of *SMG6* are good candidates to also act in this manner. In particular, many HARs are thought to guide neural development (Doan RN, 2016), which leads to the intriguing possibility that the putative astrocyte HAR enhancer region near *SMG6* serves to drive and regulate *SMG6* expression in either glial progenitor cells or mature astrocytes (Molofsky AV, 2012).

Because SMG6 is a key component of a post-transcriptional regulatory pathway and other HARs most often result in new developmental enhancer regions of the associated gene, the rapid evolution of *SMG6* may be to regulate the enhanced expression of those genes. SOX2-binding sites were previously found to be enriched in HARs that mapped to genes associated with autism spectrum disorders (Doan RN, 2016). Given that SOX2 is a critical transcription factor in maintaining pluripotency (Chambers I, 2009) and neural cell lineage decisions (López-Juárez A, 2012), this raises the possibility that a subset of HARs serve to alter SOX2 responsiveness of select pluripotency and neural genes as a means to drive normal human brain development and behavior. In the future, it will be interesting to see if there is a significant overlap between SMG6/NMD target RNAs and genes whose SOX2-binding sites correspond with a HAR. If so, these genes may have coevolved to be regulated by NMD, and have known implications in neurodevelopment. Future analysis of HARs and links to NMD targets is likely to reveal novel insights about the role of NMD in neurodevelopment.

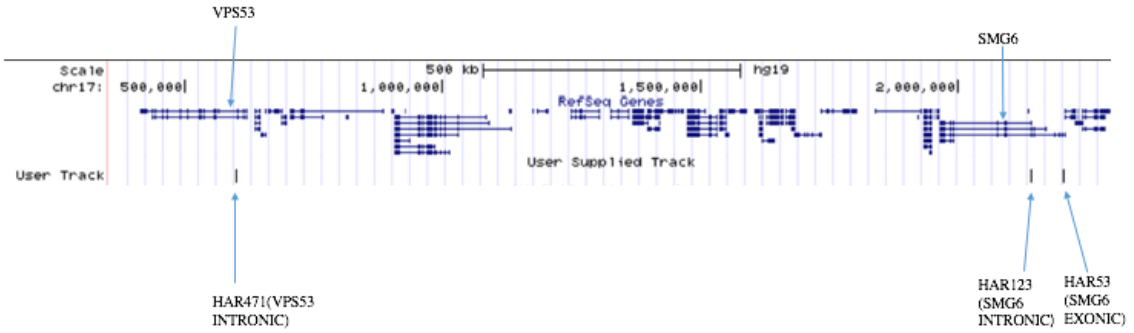

Figure 25: HARs associated with *SMG6*. Two HARs map directly within *SMG6* gene locus. HAR123 is intronic, while HAR53 is in the first exon. HAR471 is within ~150 kb downstream of the *SMG6* gene locus.

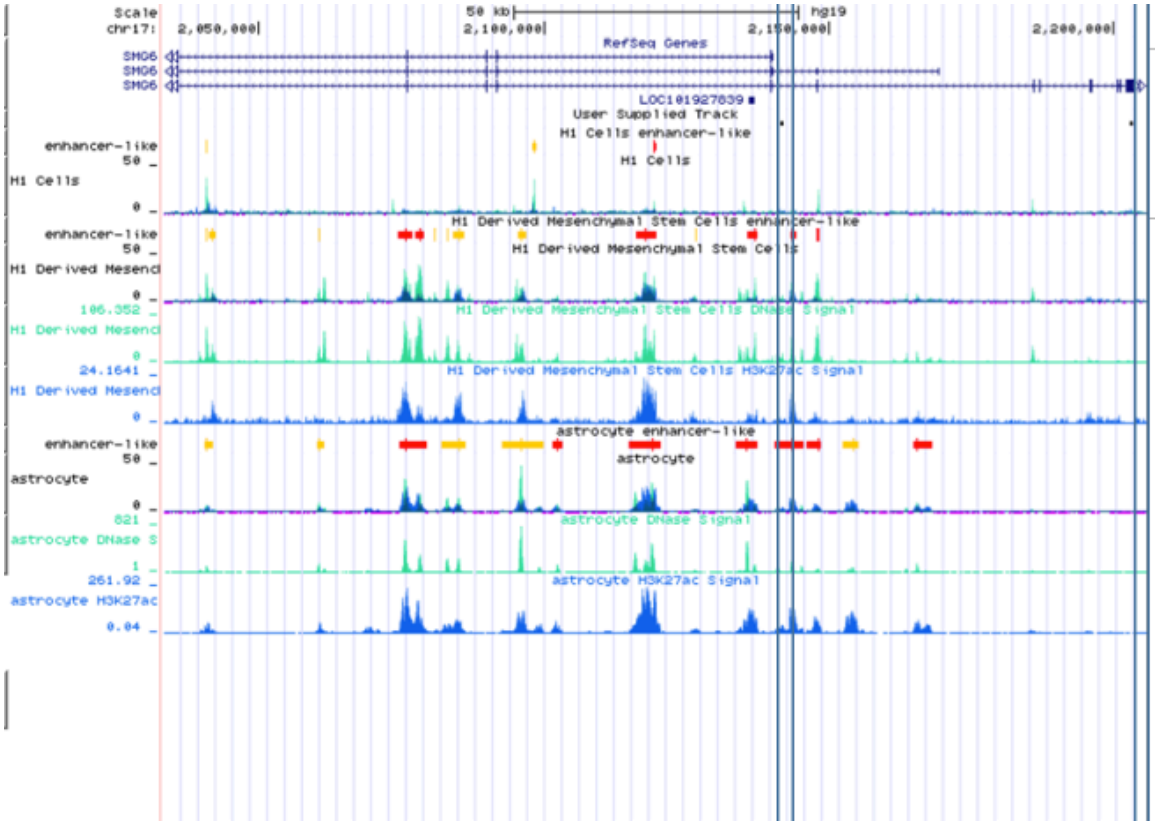

Figure 26: HAR123 corresponds with an astrocyte enhancer. Shown is an alignment of HAR123 and HAR53 (outlined in vertical blue bars) with select cell type-specific DNase I hypersensitivity sites and H3K27 acetylation sites. Predicted enhancers are denoted in red (Vista database) or yellow (Encode database).

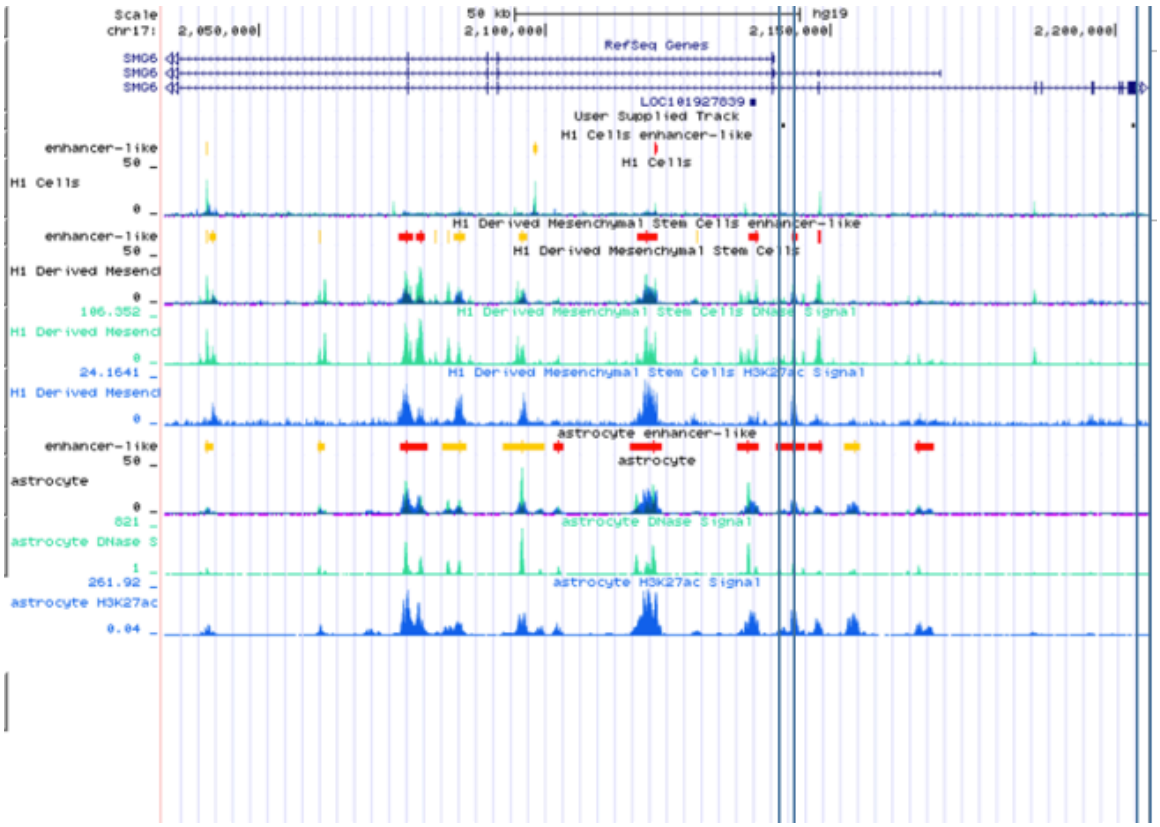

Figure 27: HAR471 corresponds with an astrocyte enhancer. Shown is an alignment of HAR471 (outlined in vertical blue bars) with select cell type-specific DNAseI digestion sites and H3K27 acetylation sites. Predicted enhancers are denoted in red (Vista database) or yellow (Encode database)

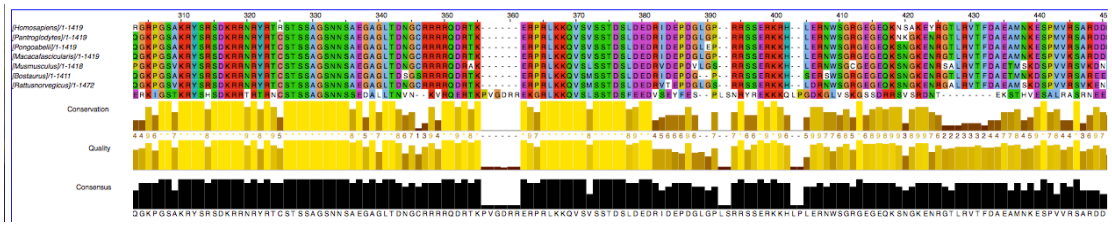

Figure 28: Impact of HAR53 on the *SMG6* coding region. HAR53 spans amino acids 239-362 in human SMG6 (indicated as 307 to 440 in this figure). The numbering in this figure is based on protein sequence alignment of the multiple species shown.

Table 11: Amino-acid substitutions in the HAR53 region. The left column shows differences between *Pan. troglodytes* and human, where the left residue is human and the right residue is *Pan. troglodytes* (numbering follows Figure 29). * indicates an amino acid in *Pan. Troglodytes* that differs is some other vertebrates. The right column has amino acid residues numbered as in human SMG6.

Amino-Acid Substitution	**Type**	**Human SMG6 Residue**
R-325 (C)	Non-Conservative	R257
K-420 (S)*	Non-Conservative	K352
A-421 (G)	Conservative	A353
Y-424 (N)	Conservative	Y356

REFERENCES

Addington AM, Gauthier J, Piton A, Hamdan FF, Raymond A, Gogtay N, Miller R, Tossell J, Bakalar J, Inoff-Germain G, Gochman P, Long R, Rapoport JL, Rouleau GA. 2011. "A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders." *Mol. Psychiatry,* 238–239.

Alrahbeni T., Sartor F., Anderson J., Miedzybrodzka Z., McCaig,C., & Mülle. 2015. "Full UPF3B Function Is Critical for Neuronal Differentiation of Neural Stem Cells." *Mol Brain* 8 (33).

Anastasaki C., Longman D., Capper A., Patton EE., & Cáceres J. F. 2011. "Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish." *Nucleic Acids Res.*

Avery P, Vicente-Crespo M, Francis D, Nashchekina O, Alonso CR, Palacios IM. 2011. "Drosophila Upf1 and Upf2 loss of function inhibits cell growth and causes animal death in a Upf3-independent manner." *RNA.*

Boehm V, Haberman N, Ottens F, Ule J, Gehring NH. 2014. "3′ UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons." *Cell Reports* 9 (2): 555–568.

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman J, Makarova KS, Koonin EV, van der Oost J,. 2008. "Small CRISPR RNAs guide antiviral defense in prokaryotes." *Science* 321: 960–964.

Capra JA, Erwin GD, McKinsey G, Rubenstein JLR, Pollard KS,. 2013. "Many human accelerated regions are developmental enhancers." *Phil. Trans. R. Soc. B* 368.1632.

Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, and Studer L. 2009. "Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling." *Nat Biotechnol* 27 (3).

Chan WK, Huang L, Gudikote JP, Chang YF, Imam JS, MacLean JA,. 2007. "An alternative branch of the nonsense-mediated decay pathway." *EMBO J.* 26: 1820–1830.

Chan, WK., Bhalla AD, Le Hir H, Nguyen LS, Huang L, Gécz J, Wilkinson MF. 2009. "A UPF3-mediated regulatory switch that maintains RNA surveillance." *Nat Struct Mol Biol.* 16: 747–753.

Choi Y, Chan AP. 2015. "In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine." *Bioinformatics* 31 (16): 2745-2747.

Chou SJ, Wang C, Sintupisut N, Niou ZX, Lin CH, Li KC, Yeang CH. 2016. "Analysis of spatial-temporal gene expression patterns reveals dynamics and regionalization in developing mouse brain." *Scientific Reports* 6: 19274.

Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, Kühn R. 2015. "Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells." *Nature Biotechnology* 33: 543–548.

Colombo M, Karousis ED, Bourquin J, R Bruggmann R and Mühlemann M,. 2016. "Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways." *RNA*.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F,. 2013. "Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013)." *Science* 339: 819–823.

Doan RN, Bae BI, Cubelos B, Chang C, Hossain AA, Al-Saad S, Mukaddes NM, Oner O, Al-Saffar M, Balkhy S, Gascon GG, Nieto M, Walsh CA. 2016. "Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior." *Cell* 167 (2): 341-354.e12.

Duodna J and Charpentier E. 2014. "The new frontier of genome engineering with CRISPR-Cas9." *Science* 346 (6213).

Gore A, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J, Canto I, Giorgetti A, Israel MA, Kiskinis E, Lee JH, Loh YH, Manos PD, Montserrat N, Panopoulos AD, Ruiz S, Wilbert ML, Yu J, Kirkness EF, Izpisua Belmonte JC, Rossi DJ, Thomson JA, Eggan K, Daley GQ, Goldstein LS, Zhang K,. 2010. "Somatic coding mutations in human induced pluripotent stem cells." *Nature* 471: 63-70.

He F and Jacobson A. 1995. "Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen." *Genes & Dev* 9: 437-454.

Hou P, Chen S, Wang S, Yu X, Chen Y, Jiang M, Zhuang K, Ho W, Hou W, Huang J, Guo D. 2015. "Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection." *Scientific Reports* 5: 15577.

Huang L, Lou CH, Chan W, Shum EY, Shao A, Stone E, Karam R, Song HW, Wilkinson MF. 2011. "RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD." *Mol. Cell*. 950-61.

Huang Y, Liang P, Liu D, Huang J,. 2014. "Telomere regulation in pluripotent stem cells." *Protein Cell*. 5 (3): 194–202.

Ishino Y, Shinagawa H, Makino K, Amemura M, and Nakata A,. 1987. "Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product." *J. Bacteriol* 169 (12): 5429-5433.

Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JW, Maquat LE. 2008. "Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay." *Cell* 133 (2): 314-27.

Jensen, Lykke-Andersen S and. 2015. "Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes." *Nature Reviews Molecular Cell Biology* 16: 665–677.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E,. 2012. "A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity." *Science* 337: 816–821.

Jinek, M., East A, Cheng A, Lin S, Ma E, Doudna J,. 2013. "RNA-programmed genome editing in human cells." *eLife* 2.

Kadlec J, Izaurralde E, Cusack S,. 2004. "The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3, The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3." *Nature Structural & Molecular Biology* 11 (30): 330 - 337.

Karousis ED, Nasif S, Mühlemann O. 2016. "Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact." *WIREs RNA*.

Kebaara and Atkin. 2009 . "Long 3′-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae." *Nucleic Acids Res*. 37 (9): 2771–2778.

Kervestin S and Jacobson A. 2012. "NMD: a multifaceted response to premature translational termination. ." *Nat. Rev. Mol. Cell Biol*. 13: 700-712.

Kozak M. 1987. "At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells." *J Mol Biol*. 196 (4): 947-50.

Laumonnier F, Shoubridge C, Antar C, Nguyen LS, Van Esch H, Kleefstra T, Briault S, Fryns JP, Hamel B, Chelly J, Ropers HH, Ronce N, Blesson S, Moraine C, Gécz J, Raynaud M. 2010. "Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism." *Mol. Psychiatry* 767–776.

Lee and Culbertson. 1995. "Identification of an additional gene required for eukaryotic nonsense mRNA turnover." *Proc Natl Acad Sci U S A* 92 (22): 10354–10358.

Leeds P, Peltz SW, Jacobson A, Culbertson MR. 1991. "The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon." *Genes & Dev* 5: 2303-2314.

Leng N, Li Y, McIntosh BE, Nguyen BK, Duffin B, Tian S, Thomson JA, Dewey CN, Stewart R, Kendziorski C. 2015. "EBSeq-HMM: a Bayesian approach for identifying gene-expression changes in ordered RNA-seq experiments." 31 (16): 2614-22.

Levine AJ, Brivanlou AH. 2006. "GDF3 at the crossroads of TGF-beta signaling." *Cell Cycle* 5 (10): 1069-73.

Li T, Shi Y, Wang P, Guachalla LM, Sun B, Joerss T, Chen YS, Groth M, Krueger A, Platzer M, Yang YG, Rudolph KL, Wang ZQ. 2015. "Smg6/Est1 licenses embryonic stem cell differen- tiation via nonsense-mediated mRNA decay." *EMBO J* 34: 1630–1647.

Liang F, Han M, Romanienko P, Jasin M,. 1998. "Homology-directed repair is a major double-strand break repair pathway in mammalian cells." *Proc Natl Acad Sci U S A*. 95 (9): 5172–5177.

Longman D, Plasterk RH, Johnstone IL, Cáceres JF. 2007. "Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway." *Genes Dev*. 21: 1075–1085.

Losson and Lacroute. 1979. "Interference of nonsense mutations with eukaryotic messenger RNA stability." *Proc Natl Acad Sci U S A*. 10: 5134-5137.

Lou CH, Dumdie J, Goetz A, Shum EY, Brafman D, Liao X, Mora-Castilla S, Ramaiah M, Cook-Andersen H, Laurent L, Wilkinson MF. 2016. "Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate." *Stem Cell Reports* 6 (6): 844-57.

Lou CH, Shao A, Shum EY, Espinoza JL, Huang L, Karam R, Wilkinson MF. 2014. "Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway." *Cell Rep* 6 (4): 748-64.

Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, and Jensen TH. 2014. "Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes." *Genes & Dev*. 28: 2498-2517.

Lyncha SA, Nguyen LS, Nga LY, Waldrond M, McDonalde D, Gecz J,. 2012. "Broadening the phenotype associated with mutations in UPF3B: two further cases with renal dysplasia and variable developmental delay." *Eur. J. Med. Genet.* 55: 476–479.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM,. 2013. "RNA-guided human genome engineering via Cas9." *Science* 339: 823–826.

Mammoto T & Ingber DE. 2010. "Mechanical control of tissue and organ development." *Development* 137 (9): 1407-1420.

Maquat L. 1994. "Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics." *Nat Rev MolCell Biol* 5: 89-99.

Marchetto MC, Belinson H, Tian Y, Beltrao-Braga P, Trujillo C, Mendes A, Nunez Y, Ou J, Gosh H, Brennand K, Pierce K, Pramparo T, Eyler L, Barnes CC, Courchesne E, Gage FH, Geschwind DH, Wynshaw-Boris A, Muotri AR. 2016. "Evidence for proliferation and synaptogenesis impairments in neural cells derived from idiopathic autistic patients." *Mol Psychiatry*.

Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L, Amenduni M, Szekely A, Palejev D, Wilson M, Gerstein M, Grigorenko EL, Chawarska K, Pelphrey KA, Howe JR, Vaccarino FM. 2015. "FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders." *Cell* 162 (2): 375-90.

McMahon JJ, Shi L, Silver DL,. 2014. "Generation of a Magoh conditional allele in mice." *Genesis* 52 (8): 752–758.

Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, Lawler AM, Dietz HC,. 2001. "Rent1, a trans- effector of nonsense-mediated mRNA decay, is essen- tial for mammalian embryonic viability." *Hum Mol Genet* 10: 99-105.

Miller SA, Dykes DD, and Polesky HF,. 1998. "A simple salting out procedure for extracting DNA from human nucleated cells." *Nucleic Acids Res* 16 (3).

Nguyen LS, Jolly L, Shoubridge C, Chan WK, Huang L, Laumonnier F, Raynaud M, Hackett A, Field M, Rodriguez J, Srivastava AK, Lee Y, Long R, Addington AM, Rapoport JL, Suren S, Hahn CN, Gamble J, Wilkinson MF, Corbett MA, Gecz J. 2012. "Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability." *Mol Psychiatry* 17 (11): 1103-15.

Nguyen, LS., Kim HG, Rosenfeld JA, Shen Y, Gusella JF, Lacassie Y, Layman LC, Shaffer LG, Gécz J,. 2013. "Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders." *Hum. Mol. Genet.* 22: 1816–1825.

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M Jr. 2007. "Ultraconserved elements are associated with homeostatic control of splicing regulators by alternativesplicing and nonsense-mediated decay." *Genes Dev* 21: 708–718.

Niakan, K. K., Han, J., Pedersen, R. A., Simon, C., & Pera, R. A. R. 2012. "Human pre-implantation embryo development." *Development* 139 (5): 829–841.

Oliveira and McCarthy. 1995. "The relationship betweeneukaryotic translation and mRNA stability. A shortupstream open reading frame strongly inhibits transla-tional initiation and greatly accelerates mRNA degra-dation in the yeast Saccharomyces cerevisiae." *J Biol Chem* 270: 8936-8943.

Perrimon N, Pitsouli C, Shilo BZ,. 2012. "Signaling Mechanisms Controlling Cell Fate and Embryonic Patterning." *Cold Spring Harb Perspect Biol.*

Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, Rosenbloom KR, Kent J, Haussler D,. 2006. "Forces Shaping the Fastest Evolving Regions in the Human Genome." *Plos Genetics.*

Pulak and Anderson. 1993. "mRNA surveillance by the Caenorhabditis elegans stag genes." *Genes & Dev* 7: 1885-1897.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. "Genome engineering using the CRISPR-Cas9 system. ." *Nat Protoc.* 8 (11): 2281-308.

Rio DC, Ares M, Hannon GJ, Nilsen TW,. 2010. "Purification of RNA Using TRIzol." *Cold Spring Harbor Press Protocols.*

Rungarunlert, S., Techakumphu, M., Pirity, M. K., & Dinnyes, A. 2009. "Embryoid body formation from embryonic and induced pluripotent stem cells: Benefits of bioreactors." *World Journal of Stem Cells,* 1: 1-21.

Salamov AA, Nishikawa T, Swindells MB. 1998. "Assessing protein coding region integrity in cDNA sequencing projects." *Bioinformatics* (http://atgpr.dbcls.jp/manuscript.html) 14 (5): 384-90.

Sanjana NE, Shalem O, Zhang F. 2014. "Improved vectors and genome-wide libraries for CRISPR screening. ." *Nat Methods* 11 (8): 783-4.

Schweingruber C, Rufenera SC, Zünda D, Yamashitac A, Mühlemann O,. 2013. ". Nonsense-mediated mRNA decay — mechanisms of substrate mRNA recognition and degradation in mammalian cells." *Biochim. Biophys. Acta* 1829: 612-623.

Shi Y, Kirwan P, Livesey FJ. 2012. "Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks." *Nature Protocols* 7: 1836–1846.

Solnica-Krezel and Sepich. 2012. "Gastrulation: Making and Shaping Germ Layers." *Annual Review of Cell and Developmental Biology* 28: 687-717.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." *Cell* 131 (5): 861-72.

Tarpey PS, Raymond FL, Nguyen LS, Rodriguez J, Hackett A, Vandeleur L, Smith R, Shoubridge C, Edkins S, Stevens C, O'Meara S, Tofts C, Barthorpe S, Buck G, Cole J, Halliday K, Hills K, Jones D, Mironenko T, Perry J, Varian J, West S, Widaa S, Teague J, Dicks E, Butler A, Menzies A, Richardson D, Jenkinson A, Shepherd R, Raine K, Moon J, Luo Y, Parnau J, Bhat SS, Gardner A, Corbett M, Brooks D, Thomas P, Parkinson-Lawrence E, Porteous ME, Warner JP, Sanderson T, Pearson P, Simensen RJ, Skinner C, Hoganson G, Superneau D, Wooster R, Bobrow M, Turner G, Stevenson RE, Schwartz CE, Futreal PA, Srivastava AK, Stratton MR, Gécz J. 2007. "Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation." *Nature* 1127–1133.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM,. 1998. "Embryonic stem cell lines derived from human blastocysts." *Science* 1145–1147.

Toma KG, Rebbapragada I, Durand S, Lykke-Andersen J. 2015. "Identification of elements in human long 3′ UTRs that inhibit nonsense-mediated decay." *RNA*. 21 (5): 887–897.

Valamehr B, Abujarour R, Robinson M, Le T, Robbins D, Shoemaker D, Flynna P,. 2012. "A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs." *Sci Rep* 2 (213).

Vallier L, Touboul T, Chng Z, Brimpari M, Hannan N, Millan E, Smithers LE, Trotter M, Rugg-Gunn P, Weber A, Pedersen RA. 2009. "Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways." *PLoS One*. 4 (6).

van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V, Lee C, Martinez RA, Levi BP, Thompson CL, Yao S, Kaykas A, Temple S, Fasano CA. 2014. "CORTECON: A Temporal Transcriptome Analysis of In Vitro Human Cerebral Cortex Development from Human Embryonic Stem Cells ." 83 (1): 51-68.

Wang L & Chen YE. 2016. "Signaling Control of Differentiation of Embryonic Stem Cells toward Mesendoderm." *Journal of Molecular Biology* 1409–1422.

Weinberger, Ayyash, Novershtern, Hanna. 2016. "Dynamic stem cell states: naive to primed pluripotency in rodents and humans." *Nature Reviews Molecular Cell Biology* 17: 155–169.

Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Mönch K, Thoren LA, Nielsen FC, Jacobsen SE, Nerlov C, Porse BT. 2008. "NMD is essential for hematopoi- etic stem and progenitor cells and for eliminating by- products of programmed DNA rearrangements." *Genes Dev* 22: 1381–1396.

Wittkopp N, Huntzinger E, Weiler C, Saulière J, Schmidt S, Sonawane M, Izaurralde E. 2009. "Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival." *Mol Cell Biol*. 29: 3517–3528.

Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Mühlemann O. 2011. "Autoregulation of the nonsense-mediated mRNA decay pathway in human cells." *RNA* 17 (12): 2108–2118.

Yiu G, Tieu E, Nguyen A, Wong B, Smit-McBride Z,. 2016. "Genomic Disruption of VEGF-A Expression in Human Retinal Pigment Epithelial Cells Using CRISPR-Cas9 Endonuclease." 57 (13): 5490–5497.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA. 2007. "Induced pluripotent stem cell lines derived from human somatic cells." *Science* 318 (5858): 1917-20.

Yu JH, Zhong XY, Zhang WG, Wang ZD, Dong Q, Tai S, Li H, Cui YF. 2012. "CDK10 functions as a tumor suppressor gene and regulates survivability of biliary tract cancer cells." *Oncol Rep* 4: 1266-76.

Zhang WY, de Almeida PE, Wu JC. 2012. "Teratoma formation: A tool for monitoring pluripotency in stem cell research." *StemBook*.