bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Gemin4 is an essential gene in mice, and its overexpression in human cells causes relocalization of the SMN complex to the nucleoplasm

Ingo D. Meier[1,3,#], Michael P. Walker[1,2,4,#] and A. Gregory Matera[1,*]

[1]Integrative Program for Biological and Genome Sciences, Departments of Biology and Genetics, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599-3280, USA

[2]Department of Genetics, Case Western Reserve University, Cleveland, OH 44106-4955, USA

[#]These authors contributed equally to this work

[3]Current address: Horizon Discovery Group PLC, Cologne, Germany

[4]Current address: BrainBits LLC, Springfield, IL, USA

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Key words: Survival motor neuron; Spinal muscular atrophy; nuclear import; ribonucleoproteins; spliceosome snRNPs; Cajal bodies

Running Title: Effects of Gemin4 knockout and overexpression

*__Address correspondence to__: Dr. Greg Matera
Department of Biology
University of North Carolina
3352 Genome Sciences
Chapel Hill, NC 27599-3280
Tel: (919) 962-4567
email: matera@unc.edu

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Gemin4 is a member of the Survival Motor Neuron (SMN) protein complex, which is responsible for the assembly and maturation of Sm-class small nuclear ribonucleoproteins (snRNPs). In metazoa, Sm snRNPs are assembled in the cytoplasm and subsequently imported into the nucleus. We previously showed that the SMN complex is required for snRNP import *in vitro*, although it remains unclear which specific components direct this process. Here, we report that Gemin4 overexpression drives SMN and the other Gemin proteins from the cytoplasm into the nucleus. Moreover, it disrupts the subnuclear localization of the Cajal body marker protein, coilin, in a dose-dependent manner. We identified three putative nuclear localization signal (NLS) motifs within Gemin4, one of which is necessary and sufficient to direct nuclear import. Overexpression of Gemin4 constructs lacking this NLS sequestered Gemin3 and, to a lesser extent Gemin2, in the cytoplasm but had little effect on the nuclear accumulation of SMN. We also investigated the effects of Gemin4 depletion in the laboratory mouse, *mus musculus*. *Gemin4* null mice die early in embryonic development, demonstrating that Gemin4 is an essential mammalian protein. When crossed onto a severe SMA mutant background, heterozygous loss of *Gemin4* failed to modify the early postnatal mortality phenotype of SMA type I ($Smn^{-/-}$;$SMN2^{+/+}$) mice. We conclude that Gemin4 plays an essential role in mammalian snRNP biogenesis, and may facilitate import of the SMN complex (or subunits thereof) into the nucleus.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Pre-mRNA splicing is a central feature of the eukaryotic gene expression programme. The removal of intronic sequences from pre-mRNAs is catalyzed by a macromolecular machine called the spliceosome. Key components of spliceosomes include the small nuclear ribonucleoproteins (snRNPs). Each of these snRNPs contains a common set of seven RNA binding factors, called Sm proteins, that forms a heptameric ring around the snRNA, known as the Sm core. Biogenesis of the Sm core is carried out by another macromolecular assemblage called the Survival Motor Neuron (SMN) complex, consisting of at least nine proteins (Gemins 2-8, unrip and SMN), reviewed in (BATTLE *et al.* 2006a; MATERA *et al.* 2007; MATERA and WANG 2014).

Following RNA polymerase II-mediated transcription in the nucleus, pre-snRNAs are exported to the cytoplasm for assembly into stable RNP particles (JARMOLOWSKI *et al.* 1994; OHNO *et al.* 2000). The SMN complex is thought to bind both the Sm proteins (B, D1, D2, D3, E, F, G) and the uridine-rich snRNAs (U1, U2, U11, U12, etc.), bringing them together and forming the Sm core RNP (CHARI *et al.* 2008; PELLIZZONI *et al.* 2002). Following 3′ processing and hypermethylation of the snRNA 5′ cap, the Sm core RNP is transported from the cytoplasm back into the nucleus. This process requires one of two known nuclear localization signals (NLSs), the 5′ trimethylguanosine (TMG) cap and the Sm core (FISCHER *et al.* 1993; MARSHALLSAY and LUHRMANN 1994). The adaptor that recognizes the cap is a protein called Snurportin (HUBER *et al.* 1998), whereas the Sm core adaptor is recognized by factor(s) within the SMN complex itself (NARAYANAN *et al.* 2004). Both adaptors use a common import receptor protein, Importin-β (HUBER *et al.* 2002; PALACIOS *et al.* 1997).

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The cytoplasmic SMN complex is thought to chaperone RNP biogenesis by conferring stringent specificity toward snRNAs and preventing illicit Sm core assembly on nontarget RNAs (CHARI *et al.* 2008; PELLIZZONI *et al.* 2002). The function of the nuclear SMN complex is completely unclear, and Gemin4 is the only member of the complex that contains a classical NLS motif (see below). A number of subcomplexes containing SMN and/or various Gemins have also been described (BATTLE *et al.* 2007; HAO *et al.* 2007; YONG *et al.* 2010), the functions of which are also unknown. Biochemical and cell biological analyses of the SMN complex have begun to elucidate functions of some of its individual components. SMN and Gemin2 form the core of the complex and are thought to be the primordial proteins in the evolution of the complex (KROISS *et al.* 2008). Gemin5 is thought to deliver snRNAs to the SMN complex by recognizing specific RNA structural and sequence features including the Sm protein binding site (BATTLE *et al.* 2006b; LAU *et al.* 2009; YONG *et al.* 2010) and the 7-methylguanosine cap (BRADRICK and GROMEIER 2009). Gemins 6 and 7 are hypothesized to act as scaffolding intermediates during assembly of the Sm core (BATTLE *et al.* 2006a; MA *et al.* 2005). Gemin8 interacts directly with SMN and forms a bridge for the Gemin6/7 heterodimer complex with unrip, bringing this trimeric module into the rest of the SMN complex (CARISSIMI *et al.* 2005; CARISSIMI *et al.* 2006a; CARISSIMI *et al.* 2006b). Gemin3/dp103/Ddx20 is a DEAD box RNA helicase (CHARROUX *et al.* 1999; YAN *et al.* 2003) that is essential for Sm core formation (SHPARGEL and MATERA 2005) as well as for metazoan development (MOUILLET *et al.* 2008; SHPARGEL *et al.* 2009). Gemin4 does not bind SMN directly, but is brought to the complex by virtue of its interaction with Gemin3 (CHARROUX *et al.* 2000; OTTER *et al.* 2007). Gemin4 is necessary for Sm core

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

formation *in vitro* (SHPARGEL and MATERA 2005), but little else is known regarding the function of this protein.

Here, we show that *Gemin4* is an essential gene in the mouse and that the protein displays a dominant effect on the localization of SMN, Gemin3 and other members of the SMN complex when overexpressed in cultured human cells. This relocalization effect was completely dependent on the presence of an eight amino acid NLS within the N-terminal region of Gemin4 that is necessary and sufficient for targeting heterologous GFP constructs to the nucleus.

Results

Gemin4 contains a functional nuclear localization signal

Like most of the Gemin proteins, Gemin4 has no other known paralogs in the vertebrate proteome, and no known orthologs among non-vertebrate species (KROISS *et al.* 2008). In humans and mice, Gemin4 is a 1058 aa protein with few known sequence features. Bioinformatic analysis using several web-based prediction tools suggested the existence of three classical (pat7 subtype) NLS sequences, one located within a slightly larger leucine zipper motif in the carboxy (C) terminal half and two others in the amino (N) terminal half of the protein (Fig. 1A).

The sequences of human and mouse Gemin4 are greater than 84% identical and more than 90% similar. In order to distinguish endogenous from exogenous Gemin4 we created expression vectors that encode N- or C-terminal GFP-tagged versions of mouse Gemin4 (GFP-Gemin4 and Gemin4-GFP, respectively). As with SMN, endogenous

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Gemin4 typically localizes diffusely throughout the cytoplasm and in distinct nuclear foci called Cajal bodies (CHARROUX *et al*. 2000; Fig. 1B). Curiously, both myc and GFP N-terminally tagged mouse or human Gemin4 proteins predominantly localized to the nucleoplasm (Fig. 1B, and Fig. 4A,C). This accumulation in the nucleus was neither due to the placement of the tag nor due to the tag itself, as C-terminally tagged Gemin4-GFP (or -myc) showed the same pattern and GFP-SMN mirrored SMN's endogenous distribution (Fig. 1B and data not shown). Thus, overexpression of Gemin4 results in a dominant gain-of-function phenotype, perhaps exposing a *cis*-acting NLS that is normally masked in the cytoplasm or titrating a *trans*-acting factor that normally facilitates Gemin4 nuclear export.

We mapped the NLS activity to the N-terminal half of Gemin4 using N- and C-terminal truncations (Fig. 2). Precise deletions of the predicted NLS motifs (Fig. 2A) identified an eight amino acid sequence that regulates GFP-Gemin4 nuclear import. Deletion of NLS1 (aa residues 62-69) or the leucine zipper (aa 714-735) had little effect, however, deletion of NLS2 (aa 199-206) relocalized the construct to the cytoplasm (Fig. 2B). Consistent with these findings, Lorson et al. (2008) reported that a ~50 aa sequence (aa 194-243) overlapping this region was required for Gemin4 import. The NLS2 motif is therefore necessary for nuclear import of Gemin4. To address whether NLS2 is sufficient for this process, we fused it to the carboxy terminus of GFP. Two constructs were expressed, one consisting of only the NLS2 sequence (NLS2min), and one with five additional residues flanking the NLS in both directions (NLS2ext). As shown in Fig 2B, native GFP displayed an overall pan-cellular localization pattern, whereas GFP-NLS2min was much more concentrated in the nucleus, albeit with a weak signal in the cytoplasm. In contrast, the GFP-NLS2ext construct was almost entirely nuclear (Fig. 2B). These

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

observations demonstrate that the Gemin4 NLS2 motif is both necessary and sufficient for nuclear import.

Gemin4 overexpression redistributes the SMN complex to the nucleoplasm

Although the tudor domain of SMN can interact directly with Importin-β (NARAYANAN *et al.* 2004), Gemin4 is the only SMN complex member that contains an NLS. We therefore examined the ability of Gemin4 to influence the subcellular localization of other SMN complex proteins. HeLa cells were transiently transfected with GFP-Gemin4 and immunostained for SMN, Gemin2, Gemin3 and Unrip (Fig. 3). In each case, overexpression of GFP-Gemin4 caused relocalization of the endogenous protein to the nucleus. There was a clear dosage effect to the redistribution. In weakly transfected cells, SMN was qualitatively reduced (but still visible) in the cytoplasm (Fig. 3A,B). In cells strongly expressing GFP-Gemin4, endogenous SMN was no longer detectable in the cytoplasm and the nuclear SMN foci (Cajal bodies) were much less frequent (Fig. 3A,B). Similar effects on the cellular distributions of SMN and Gemin2 were observed upon overexpression of myc-Gemin4 (Fig. 4A,C). Quantitative analysis showed that expression of myc- or GFP-Gemin4 significantly relocalized endogenous SMN and Gemin2 to the nucleoplasm (Fig. 4B,F).

High levels of Gemin4 overexpression disrupts Cajal bodies

Gemin4 overexpression drives cytoplasmic SMN complexes into the nucleoplasm, but these proteins frequently fail to concentrate in Cajal bodies (Fig. 3). We quantified the loss of SMN, Gemin2, and Gemin3 nuclear foci in myc- or GFP-Gemin4 expressing cells and found that the ability of these SMN complex components to accumulate within

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

nuclear foci was significantly inhibited (Fig. 4 and data not shown). To determine if the loss of these nuclear foci in myc- or GFP-Gemin4 expressing cells was due to Cajal body disassembly, we examined the distribution of known Cajal body marker proteins: coilin, WDR79 and NPAT (Fig. 5A-C). Nuclear foci with these Cajal body markers were largely unaffected in cells expressing lower levels of GFP-Gemin4, but we observed an appreciable reduction of nuclear foci in cells with high GFP-Gemin4 expression (Fig. 5C). We then quantified the loss of coilin-positive nuclear foci and found a significant increase in cells that contained no distinguishable nuclear foci (Fig. 5D). Thus, the observed Gemin4-dependent relocalization of SMN and its binding partners to the nucleoplasm perturbs Cajal body integrity in a dose dependent manner. Because coilin interacts with SMN and mediates recruitment of the SMN complex to Cajal bodies (HEBERT *et al.* 2002; HEBERT *et al.* 2001), it is likely that the SMN binding sites on coilin are simply swamped by the preponderance of nuclear SMN.

The Gemin4 NLS is required for SMN and Cajal body disruption

To further examine the critical Gemin4 regions involved in SMN relocalization, we overexpressed Gemin4 lacking the NLS (myc-G4ΔNLS2). We predicted that NLS2 was responsible for the disruption of SMN nuclear foci upon overexpression of Gemin4. We found that SMN nuclear foci were relatively unaffected by myc- or GFP-Gemin4ΔNLS2 expression (Fig. 6A,B), demonstrating that the loss of SMN nuclear foci upon Gemin4 overexpression requires the presence of Gemin4 NLS2. In contrast, both Gemin2 and Gemin3 nuclear foci were significantly reduced upon myc- or GFP-Gemin4ΔNLS2 expression (Fig. 6D,F). This observation indicates that Gemin2 and Gemin3 are sequestered by Gemin4ΔNLS2 in the cytoplasm. Notably, the reduction in Gemin3 nuclear foci was much more robust than that of Gemin2 (Fig. 6D,F). This is perhaps due

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

to the fact that Gemin4 is a direct binding partner of Gemin3, whereas Gemin2 is not (OTTER *et al.* 2007). Additionally, we note that Coilin nuclear foci were observed in nearly all myc- or GFP-Gemin4ΔNLS2 transfected and control cells (Fig. 6G,H). Taken together, these data suggest that the disruption of coilin and SMN nuclear foci depends upon a functional Gemin4 NLS.

Nucleoplasmic relocalization of Gemin3 requires a C-terminal motif in Gemin4

As mentioned above, Gemin4 does not interact directly with SMN; it is thought to be tethered to the SMN complex via its strong and direct interaction with Gemin3 (CHARROUX *et al.* 2000; OTTER *et al.* 2007). As shown in Figs. 3C and 7A, overexpression of GFP-Gemin4 mislocalizes endogenous Gemin3 to the nucleoplasm. We used this effect as a way to map the interaction between Gemin4 and Gemin3. Interestingly, the GFP-Gemin4ΔCT construct, which contains the NLS and localizes to the nucleus, did not perturb localization of endogenous Gemin3 (Fig. 7B). This finding suggests that the truncated Gemin4 no longer interacts with Gemin3 and that the domain responsible for this interaction is located in the C-terminal half of Gemin4. Consistent with this interpretation, we found that relocalization of Gemin3 depended on the presence of the putative leucine zipper motif in Gemin4. Cells transfected with GFP-Gemin4Δzip displayed a Gemin3 localization pattern comparable to that of non-transfected cells (Fig. 7C). These data demonstrate that the Gemin4-dependent translocation of the SMN complex to the nucleus depends on a leucine-rich domain in the Gemin4 C-terminus and further suggest that Gemin4 may bind to Gemin3 via this motif.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Characterization of a Gemin4 gene trap allele in the mouse

Experiments in cultured cells can contribute a great deal of information for studying protein function. However, it is difficult to extrapolate the relevance of a given protein from single cell studies to the importance that it may have in the context of organismal development. Therefore, we created and characterized a loss-of-function *Gemin4* mouse model. We obtained mouse embryonic stem (ES) cells (Lexicon Genetics) that contain a retroviral insertion (ZAMBROWICZ *et al.* 1998) in the single *Gemin4* intron. This "gene trap" insertion is derived from an engineered retroviral cassette that preferentially inserts itself into upstream introns within the mouse genome. The genomic organization of *Gemin4* is ideal for this kind of gene-trapping scheme because it is a single-intron gene and the upstream exon (exon 1) is particularly short, encoding only the first three aa residues (see Fig. 8). Exon 2 thus contains essentially the entire protein coding sequence. The gene-trap contains an upstream element consisting of a 5′ splice acceptor site fused to a β-galactosidase/neomycin-resistance (β-geo) cassette that also contains a transcription termination sequence. The β-geo cassette is transcribed from the endogenous promoter of the *Gemin4* gene, located on mouse chromosome 11. The result is a fusion transcript in which the exon upstream of the insertion site is spliced in-frame to the β-geo reporter. The fusion transcript encodes three residues of the Gemin4 polypeptide fused to the β-geo cassette. The downstream element contains a puromycin-resistance cassette along with its own PGK promoter that 'hijacks' expression of exon 2. The gene trap has stop codons introduced in all three downstream reading frames. There is a splice donor site that fuses this transcript with exon 2 (Fig. 8A).

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Blastocysts harboring the mutant ES cells were injected into pseudopregnant females, resulting in four chimeric males. Two of these mice displayed germline transmission of the gene-trap, which was ascertained by PCR genotyping of the founder progeny (Fig. 8B). Founder mice were then backcrossed for three generations onto the C57BL/6J genetic background.

Gemin4 is an essential gene in the mouse

Heterozygous animals were intercrossed and the resulting offspring were PCR genotyped at various developmental stages. Whereas heterozygous and wild-type progeny were readily detected at postnatal day 1 (P1), embryonic day 13.5 (E13.5), E7.5 and E5.5, no *Gemin4* homozygotes were detected at any of these time points (Table 1 and data not shown). These findings demonstrate that although *Gemin4* is thought to be an evolutionary newcomer to the SMN complex (it has been identified only in vertebrate genomes; KROISS *et al.* 2008), it is an essential mammalian gene. The data are also consistent with gene targeting experiments showing that mouse *Gemin2* and *Gemin3* knockouts are also early embryonic lethal mutations (JABLONKA *et al.* 2002; MOUILLET *et al.* 2008).

Transgenic suppression of Smn lethality by SMN2 is background-dependent

The human genome contains two copies of the SMN gene, *SMN1* and *SMN2*. Homozygous loss of *SMN2* is asymptomatic, but loss of SMN1 results in a neuromuscular disease, Spinal Muscular Atrophy (SMA) (LEFEBVRE *et al.* 1995). SMA is caused by hypomorphic reduction of SMN protein levels, whereas complete loss of

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

gene function is embryonic lethal (reviewed in BURGHES and BEATTIE 2009; CAUCHI 2010). The mouse genome contains only a single copy of the *Smn* gene, null mutation of which is early embryonic lethal (SCHRANK *et al.* 1997). Transgenic expression of human *SMN2* in the *Smn* knockout background rescues embryonic lethality, and recapitulates the SMA type I phenotype (HSIEH-LI *et al.* 2000; MONANI *et al.* 2000).

In order to analyze the effect of *Gemin4* copy number on the SMA phenotype, we generated a colony of $Smn^{-/-};SMN2^{+/+}$ mice and backcrossed them onto the C57BL/6J inbred background for >6 generations. We then intercrossed the animals and were surprised to find that the $Smn^{-/-};SMN2^{+/+}$ genotype was never detected postnatally (Table 2). Similar results were observed on a mixed C57BL/6J;129X1/SvJ background (Table 2). In contrast, and consistent with previous findings (MONANI *et al.* 2000), $Smn^{-/-};SMN2^{+/+}$ animals were detected in their expected numbers on the FVB/NJ hybrid background. The expected 1:2 ratios of wild-type to heterozygous mice in the non-FVB strains (Table 2) suggest that the $Smn^{-/-};SMN2^{+/+}$ embryos in those strains died *in utero*. We conclude that *SMN2's* function as a genetic suppressor of the *Smn* embryonic lethal phenotype is background dependent.

Gemin4 does not function as a genetic modifier of mouse Smn

In order to assay the effects of *Gemin4* copy number on the SMA phenotype, we crossed the *Gemin4* gene trap onto the SMA background to obtain $Gemin4^{+/-};Smn^{+/-};SMN2^{+/+}$ animals. We intercrossed these mice and analyzed their progeny. As shown in Table 3, SMA mice were present in expected numbers irrespective of the *Gemin4* copy number. This failure to exacerbate (or ameliorate) the early lethality phenotype was puzzling, considering the fact that both *Gemin2* and *Gemin3* heterozygous mice produce roughly

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

half the protein levels compared to wild-type and have associated phenotypes as a result (JABLONKA *et al.* 2002; MOUILLET *et al.* 2008). Notably, *Gemin4* heterozygotes are phenotypically indistinguishable from their wild-type littermates.

We therefore examined the expression profile of Gemin4 in various tissues. Consistent with microarray expression profiles (NCBI gene ontology database profile GDS3142), Gemin4 was readily detectable in all the tissues we examined (data not shown). Although wild-type adult and neonatal mice expressed approximately two-fold greater amounts *Gemin4* mRNA than their heterozygous littermates, the protein levels from these animals were consistently equivalent (Fig. S1). It is possible that post-translational mechanisms exist to regulate Gemin4 protein levels in heterozygotes. We screened numerous commercial and non-commercial antibodies against Gemin4 for their ability to recognize the mouse protein (not shown). Although a few worked nominally well for westerns, unfortunately none of them were suitable for immunofluorescence experiments. Thus we were unable to analyze *Gemin4* preimplantation embryos to carry out additional phenotypic analyses. We also note that *Gemin4* heterozygous mice are phenotypically identical to wild-type mice and, as a result, they are unable to modify the SMA phenotype.

Discussion

Gemin4 contains a functional NLS in its N-terminal domain that is necessary and sufficient for nuclear import of exogenous cargoes. Curiously, the localization pattern of epitope-tagged versions of Gemin4 does not mirror that of the endogenous protein. Ectopically expressed Gemin4 is primarily nucleoplasmic, whereas the endogenous

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

protein is localized throughout the cytoplasm, and in nuclear Cajal bodies. Mislocalization of the tagged constructs was not due to the presence of the tag itself, as a construct containing a deletion of the NLS (GFP-Gemin4ΔNLS2; Fig. 2) was completely excluded from the nucleus, with the exception of a very small number of cells that had distinct myc-Gemin4ΔNLS2 nuclear foci that colocalized with SMN (Fig. 6A). This rare targeting to Cajal bodies suggests that a small fraction of myc-Gemin4ΔNLS2 proteins may gain access to the nucleus by piggybacking onto Gemin3 and the rest of the SMN complex. We also discovered that relocalization of Gemin3 depends on the presence of the leucine zipper motif in Gemin4 (Fig. 7). Because Gemin4ΔNLS2 retains this motif, it is somewhat surprising to find that Gemin4ΔNLS2 is apparently limited in its ability to piggyback onto Gemin3 for import into the nucleus. The fact that SMN, but not other SMN complex components like Gemin2 and Gemin3, retains the ability to localize to Cajal bodies upon overexpression of Gemin4ΔNLS2 suggests that SMN can be imported independently from the other Gemins. Consistent with this observation is the fact that SMN was shown to directly interact with Importin-β (NARAYANAN *et al.* 2004; NARAYANAN *et al.* 2002).

Gemin4 is thought to be tethered to the SMN complex via its interaction with Gemin3, and it was shown to form a subcomplex together with Gemin5 to create a Gemin3/4/5 heterotrimer (BATTLE *et al.* 2007; CARISSIMI *et al.* 2005; CARISSIMI *et al.* 2006b). Overexpression of Gemin4 likely alters the stoichiometry of the various members of the SMN complex (or subcomplexes thereof). Thus it is also a bit surprising to find that Gemin4 overexpression results in the nucleoplasmic accumulation of SMN and all other tested members of the complex. These findings indicate that Gemin4 may play a

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

regulatory role involved in the nuclear import of the SMN complex. SMN enters the nucleus during import of newly-formed snRNPs and is part of a two-component nuclear import signal (FISCHER *et al.* 1993; NARAYANAN *et al.* 2004; NARAYANAN *et al.* 2002). Depending on the combination of binding partners in a given complex or subcomplex, it is possible that the Gemin4 NLS is masked. Under certain conditions, the Gemin4 NLS might serve to tip the balance of import in one direction or the other.

Characterization of Gemin4 loss-of-function mice

Our data clearly demonstrate that *Gemin4* is an essential mammalian gene. Importantly, null mutations in *Smn, Gemin2* and *Gemin3* are all embryonic lethal, demonstrating that these genes are also essential (JABLONKA *et al.* 2002; MOUILLET *et al.* 2008; SCHRANK *et al.* 1997). U snRNP biogenesis is an essential cellular process that ensures the availability of splicing factors required for gene expression. Previous knockdown experiments revealed that depletion of Gemin4 (or Gemin3) causes an intermediate, yet significant, loss of U snRNP assembly activity in a HeLa cells (SHPARGEL and MATERA 2005). This intermediate effect was not overly detrimental, as cell death was not nearly as pronounced as it is upon SMN knockdown (LEMM *et al.* 2006, and our unpublished observations). These observations suggest that Gemin4 and Gemin3 may be dispensable when it comes to the basal levels of U snRNPs needed to maintain cells in culture. However, during mammalian development U snRNA and snRNP levels increase dramatically from the 2-16 cell stage to the blastocyst stage (DEAN *et al.* 1989; LOBO *et al.* 1988). The SMN complex would most likely need to operate at peak efficiency to account for the large increase in U snRNP production during these critical stages of development and the full complement of Gemins may be required to stabilize the complex or directly aid in assembly of snRNPs (STRZELECKA *et al.* 2010). This idea is

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

bolstered by the fact that *Smn*, *Gemin2*, *Gemin3* and *Gemin4* null mice all die during early embryogenesis (JABLONKA *et al.* 2002; MOUILLET *et al.* 2008; SCHRANK *et al.* 1997; this work).

Gemin4 was also reported to reside in certain mammalian miRNP complexes (DOSTIE *et al.* 2003; HUTVAGNER and ZAMORE 2002; MOURELATOS *et al.* 2002). Given their wide range of gene regulatory roles, including post-transcriptional mRNA cleavage or translational repression (HE and HANNON 2004), miRNPs are required for mammalian development (BERNSTEIN *et al.* 2003; CHEN *et al.* 2004; HOUBAVIY *et al.* 2003). Gemins 3 and 4 were reported to bind in a separate complex along with the Argonaute protein, eIF2C2/Ago2 (MOURELATOS *et al.* 2002; NELSON *et al.* 2004), although the significance of these findings has not been explored. Because relatively few proteins make up this miRNP subclass, it is likely that removal of any of its members would render it non-functional. Gemin4 is not known to have a paralog or redundant equivalent; it is reasonable to assume that if Gemin4 is required for miRNP assembly, loss of its function would be detrimental to this pathway as well as for snRNP biogenesis.

In summary, we have shown that Gemin4 has the potential to redirect the localization of other SMN complex members from the cytoplasm to the nucleoplasm in cultured mammalian cells. Gene disruption in mice demonstrated that *Gemin4* is required for embryonic viability. We also found that *SMN2* failed to rescue the embryonic lethality phenotype of *Smn* knockouts, when bred on the C57BL/6J inbred background. These data demonstrate the existence of additional genetic modifiers of SMA. We were unable to conclude whether *Gemin4* is such a modifier, as we found that heterozygous *Gemin4* mice express wild-type levels of protein.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Materials and Methods

Plasmid construction

Mouse total brain cDNA was used to PCR amplify mouse Gemin4. The amplicon was cloned into the pEGFP (Clontech) or pcDNA-myc (Invitrogen) vectors to express both N- and C-terminally tagged proteins. The QuickChange site-directed mutagenesis kit (Stratagene) was used to create the deletion constructs following the manufacturer's protocol (primer sequences available upon request).

Immunofluorescence microscopy

HeLa cells were seeded on two-well glass slides and grown in an incubator at 37 °C with 5% CO_2 in DMEM (Mediatech) supplement with 10% BSA and 1% Penicillin/Streptomycin. HeLa cells were grown until they reached the 60% confluency, and transiently transfected using Effectene transfection kit (Qiagen, manufacturer protocol). Cells were harvested 24 h later, fixed in 4 % PFA in 1x PBS solution for 20 min at RT, and permeabilized in Triton X-100 followed by 3 washes in 1x PBS for 5 min at RT. Immunofluorescence experiments were performed by incubations of primary antibody diluted in PBS containing 3% BSA followed by incubation with secondary Ab. Antibodies used were as follows: anti-SMN mAb (clone 8, BD biosciences, 1:200), anti-dp103/Gemin3 mAb (4G7, 1:10), anti-Gemin4 mAb (clone 3E1, Abnova, 1:10), anti-WDR79 mAb (ab77333, abcam, 1:100), anti-NPAT mAb (DH4, gift from J. Zhao, 1:50), anti-coilin pAb (R124, 1:400), anti-Unrip mAb (3G6, 1:10), anti-Gemin2 mAb (6084-100 abcam 1:200), Alexa 594 goat anti-mouse and goat anti-rabbit secondary antibody (Invitrogen) were used. The incubations were carried out at 37 °C for an hour, stained 3 min with DAPI, washed with 1x PBS, and covered with antifade solution to avoid

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

bleaching. Laser confocal fluorescence microscopy was performed with Leica TCS SP5 high speed and high-resolution spectral confocal microscope. Images from each channel were taken within a single plane with an objective of 63x with a 3x zoom factor and recorded separately and the files were merged as needed.

Mouse lines and crosses

Various inbred mouse strains used in this study (FVB/NJ, C57BL/6J and 129Sv/J) were obtained from the Jackson Laboratory. Wild-type animals thus obtained were used for colony maintenance and to outcross *Gemin4* and *Smn* mutants onto the various genetic backgrounds described in the results section. Mice carrying a human *SMN2* transgene in the background of a null mutation in the endogenous *Smn* gene (SMA type I mice: FVB.Cg-Tg(*SMN2*)89^{Ahmb};*Smn*^{tm1Msd}/J) were obtained from the Jackson Laboratory. *Gemin4* mice were created by purchasing ES cells with the *Gemin4* gene-trap cassette from Lexicon Genetics and these cells were injected into donor blastocysts and subsequently injected into a pseudopregnant female mouse by the Case Western Reserve University transgenic mouse facility. All strains were maintained on a standard diet of 50/10 food pellets and sterile water. These mice were housed in micro-isolation chambers. Breeding pairs for SMA type I mice consisted of mice that were homozygous for the transgene and heterozygous for the knockout allele, which resulted in pups that display the SMA phenotype and control littermates. Breeding pairs for *Gemin4* mice were heterozygous for the gene trap cassette. All mice were humanely euthanized according to protocols and standards set forth by the appropriate Institutional Animal Care and Use Committees (IACUC): the Case Western Reserve University Animal Resource Center (CWRU ARC) and the University of North Carolina Division of Laboratory Animal Medicine (UNC DLAM).

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Reverse transcription and quantitative real-time PCR

RNA was extracted from homogenized liver using the RNeasy kit (Qiagen; manufacturer's protocol), including an RNase-free DNase (Qiagen) on-column digestion step to remove genomic DNA. The SuperScript First Strand Synthesis System for RT-PCR (Invitrogen) was used to synthesize cDNA in a 20 μl reaction containing 5 μg RNA, 179 ng random hexamer primer and 40 U RNaseOut RNase inhibitor (Invitrogen). The reverse transcription reaction was performed (Invitrogen SuperScript; manufacture's protocol). In addition to the PCR mastermix buffer (gotaq, Promega), each PCR reaction mixture (20 μl) contained 1 μl of cDNA 1:10 dilution. The Primers used are listed below. PCR conditions consisted of one step at 95 °C with 5 min hold and two-segment cycles (95 °C with 15 s hold and 60 °C with 1 min hold) followed by a terminal step 72 °C.

Genotyping and RT-PCR

Tail clippings from the tip of approximately 3 mm were collected from mice and used for DNA extraction (Roche, High Pure PCR Template Preparation Kit; manufacture's protocol). For genotyping of Gemin4 genetrap (G4GT) the following primers were used. LTR-Forward: AAATGGCGTTACTTAAG-CTAGCTTGC, G4GT-Forward: GGAGCGAATATAGCCTTGATTCTCTGGAAATG, G4GT-Reverse: CTTCCCAGGACGGCCTCCTAGTCTTACCCTCTA.

The genotyping primers for the murine Smn Neo cassette are as follows. neostop-F: TCGCCTTCTTGACGAGTTCTTCTG, Smn-Forward: AGGATCTCTGTGTTCGTGCG-TGGTG, Smn-Reverse: CCTTAAAGGAAGCCACAGCTTTATC. Dr. Cathleen Lutz of

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the Jackson Laboratory graciously supplied primer sequences for the *SMN2* transgene. PCR amplification was performed using standard protocols. For RT-PCR, total RNA was isolated from mouse tissues using Trizol reagent (Invitrogen; manufacture's protocol). For RT-PCR analysis, the following primers were used: Gemin4 exon1-Forward – CAGACTACAGCACGGAAGCGGAG, Gemin4 exon2-Reverse – CTAAGCAGTTGGTGGTGCAGGATG.

Western blotting

Protein lysates were prepared by flash freezing mouse tissues in liquid nitrogen then crushing the tissue into a fine powder. This powder was then transferred into RIPA-buffer with protease inhibitor (Thermo Scientific), homogenated with a 27G½ syringe and centrifuged at max speed at 4 °C for 10 min. Supernatants were quantified using a standard Bradford assay protocol. Equal amounts of samples (60 μg) were subjected to 4 - 12 % MOPs/NuPage gradient gel system (Invitrogen) and transferred onto a nitrocellulose membrane (Schleicher & Schuell) following standard protocols. Lanes were transferred to nitrocellulose membranes and incubated with appropriate anti-bodies. Gemin4 immunoreactive bands were detected using the polyclonal Gemin4 Ab and horseradish peroxidase-conjugated goat antibodies to rabbit IgG (Gemin4 Ab 1:200 Santa Cruz Biotechnologies), monoclonal anti α tubulin Ab 1:5000 (Sigma-Aldrich) (secondary mouse Ab 1:10000, secondary rabbit Ab 1:5000, both from Thermo Scientific). Detection of protein bands was achieved using standard chemiluminescence substrates (SuperSignal West Femto, Thermo Scientific).

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Acknowledgments

We thank K.B. Shpargel for valuable help during early stages of the project and A.H. Natalizio for assistance with reformatting the figures. This work was supported by grants to A.G.M. from the NIH (R01-GM118636) and the Muscular Dystrophy Association (RG-4070). M.P.W. was supported in part by an NIH predoctoral traineeship (T32-GM08613).

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figures and Legends

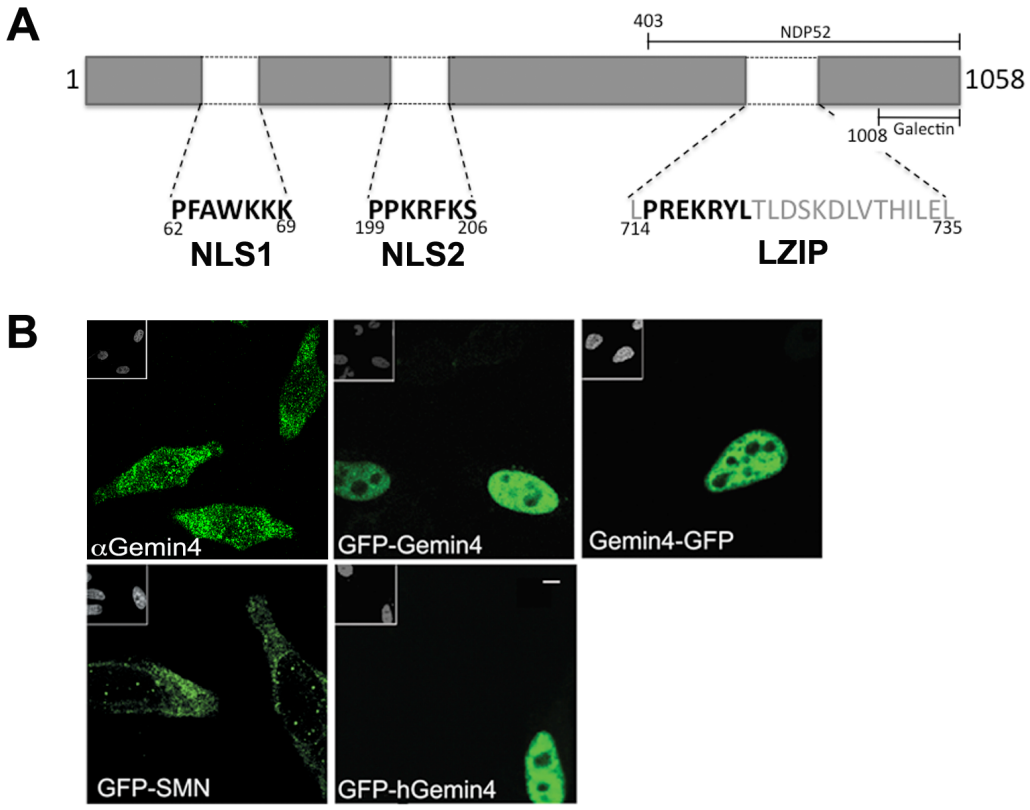

Figure 1. Gemin4 constructs and localization. (A) Schematic of Gemin4, illustrating structural features along with Galectin and NDP52 binding domains. Three predicted nuclear localization signal (NLS) sequences are shown (in bold type), one of which resides within a larger C-terminal leucine-rich motif (gray typeface). (B) Endogenous Gemin4 protein (α-Gemin4) localizes throughout the cytoplasm as well as in distinct nuclear foci called Cajal bodies. Full-length GFP-tagged mouse constructs (GFP-Gemin4 and Gemin4-GFP), as well as human Gemin4 (GFP-hGemin4) localize to the nucleoplasm when transiently transfected into HeLa cells. GFP-Smn localizes primarily to the cytoplasm, accumulating in nuclear foci, called Cajal bodies. Inserts in upper left corners are the DAPI stained nuclei. Scale bar, 5 μm.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

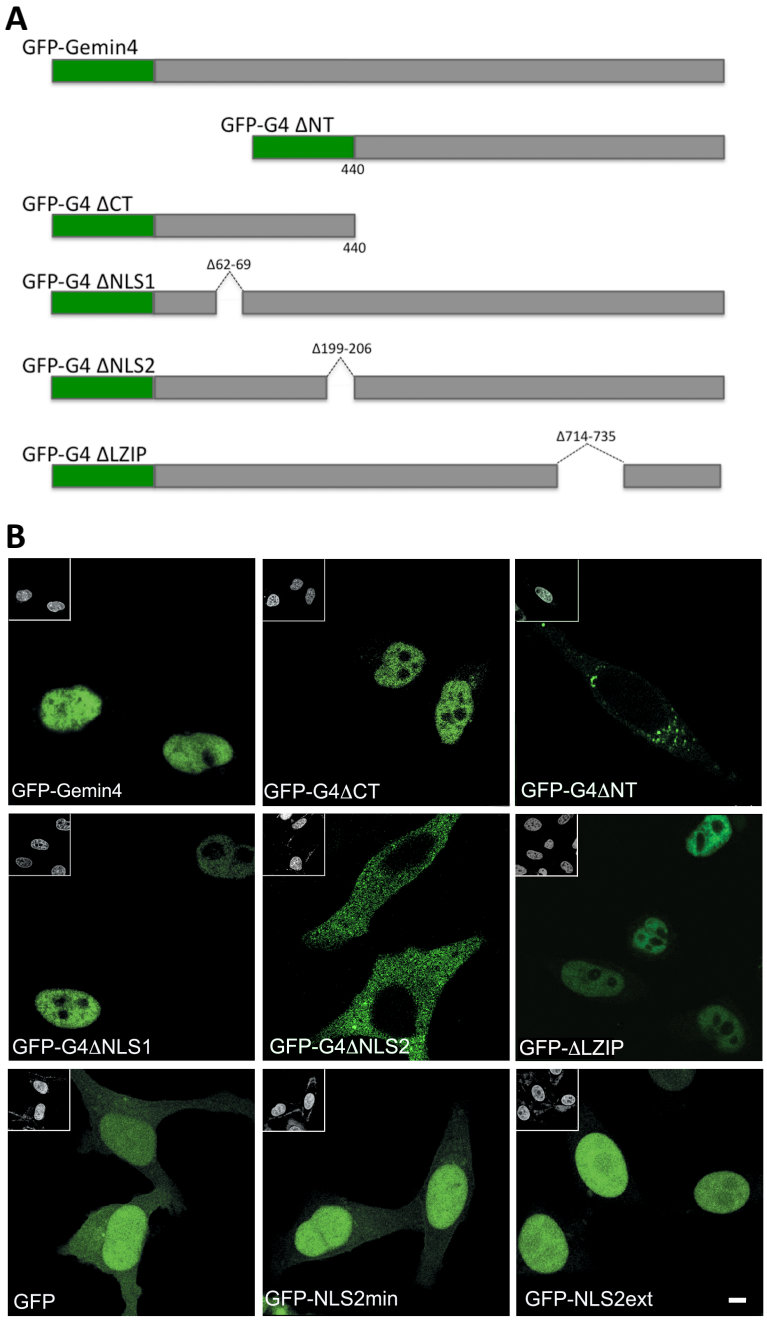

Figure 2. Identification of a functional NLS within mouse Gemin4. (A) Schematic of mouse Gemin4 protein. (B) (Top row) Full-length (GFP-Gemin4) and an N-terminal (GFP-G4ΔCT) fragment of Gemin4 localize to the nucleoplasm. A C-terminal (GFP-G4ΔNT) fragment accumulates in the cytoplasm. (Middle row) Expression of internal deletion constructs shows that only NLS2 is required for nuclear accumulation of Gemin4. (Bottom row) Although GFP-alone displays pan-cellular localization, the minimal NLS2 motif (GFP-NLS2min) is primarily nuclear. The extended NLS sequence (GFP-NLS2ext) is exclusively nuclear. Inserts in upper left corners of panels show the DAPI stained nuclei. Scale bar, 5 μm.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

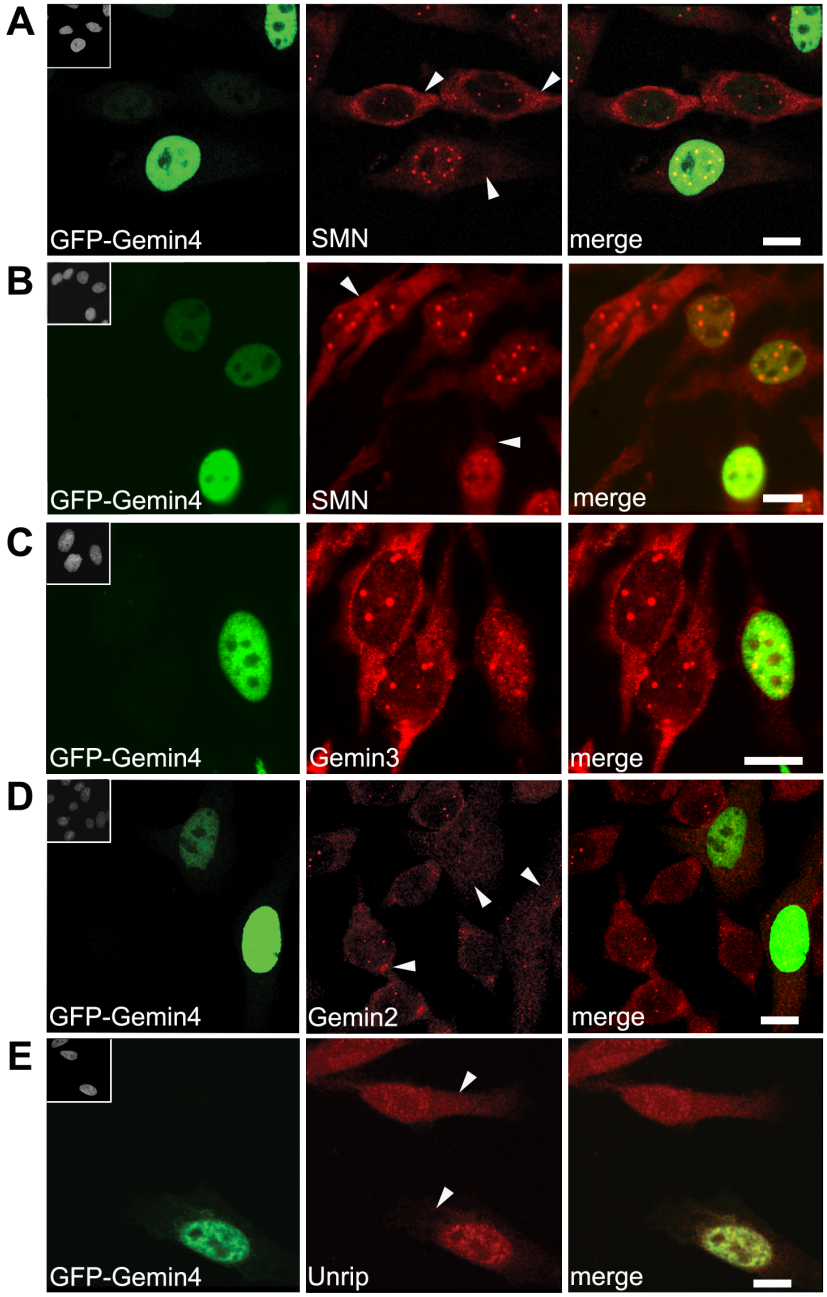

Figure 3. Overexpression of GFP-Gemin4 causes mislocalization of SMN and other members of the SMN complex to the nucleoplasm. HeLa cells were transfected with GFP-Gemin4 (shown in green) and then co-stained (in red) with antibodies targeting SMN (A and B), Gemin3 (C), Gemin2 (D), or Unrip (E). Arrowheads are used to illustrate the reduced cytoplasmic staining of endogenous proteins in the transfected cells compared to the untransfected cells. Inserts in upper left corners of the GFP images show the DAPI stained nuclei. Scale bars, 10 μm.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

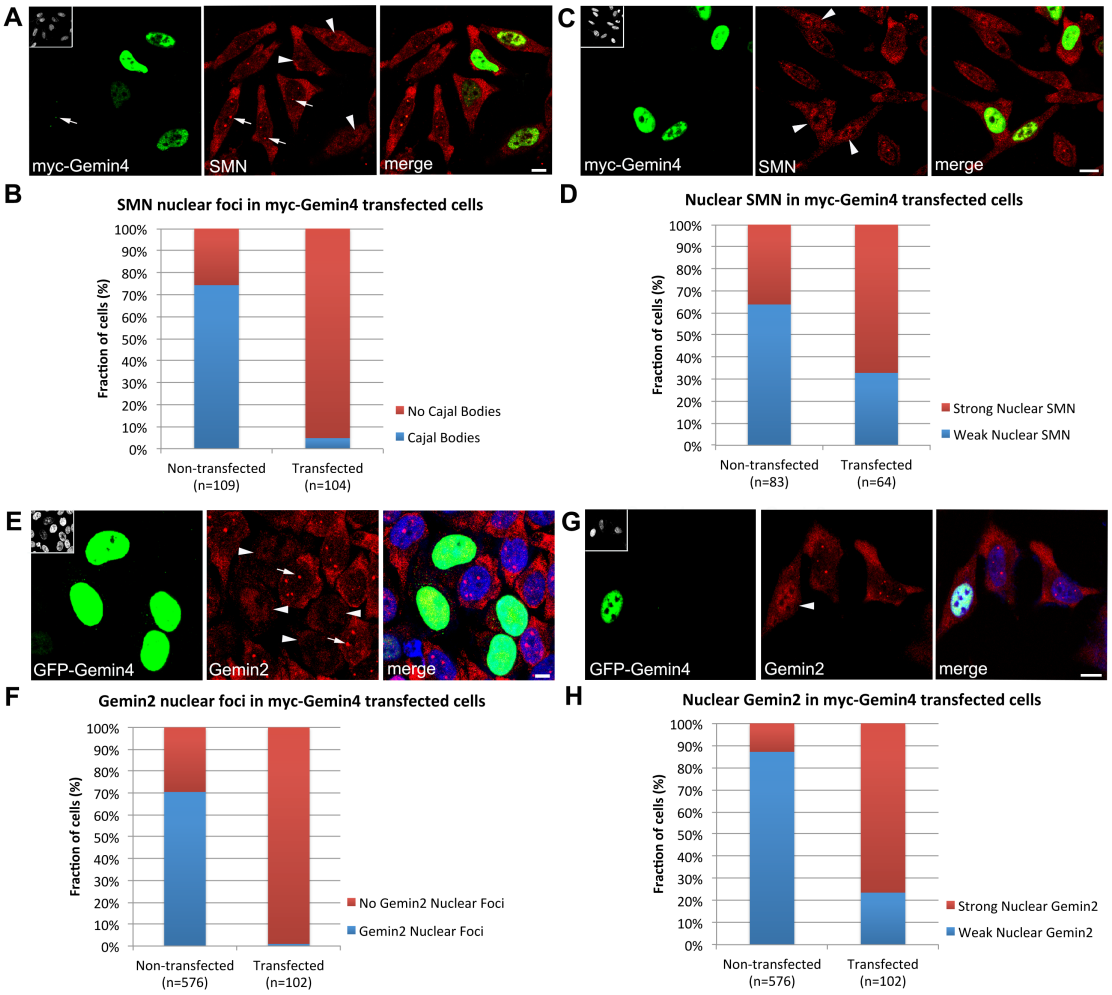

Figure 4. Quantification of SMN and Gemin2 phenotypes in HeLa cells overexpressing Gemin4. HeLa cells were transfected with either myc- (panels A-D) or GFP-Gemin4 (E-H) and then co-stained with antibodies targeting SMN or Gemin2, as shown. Each nucleus was scored for the presence or absence of nuclear foci (B and F) as well as for nucleoplasmic accumulation (D and H). Scale bars, 5 μm.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

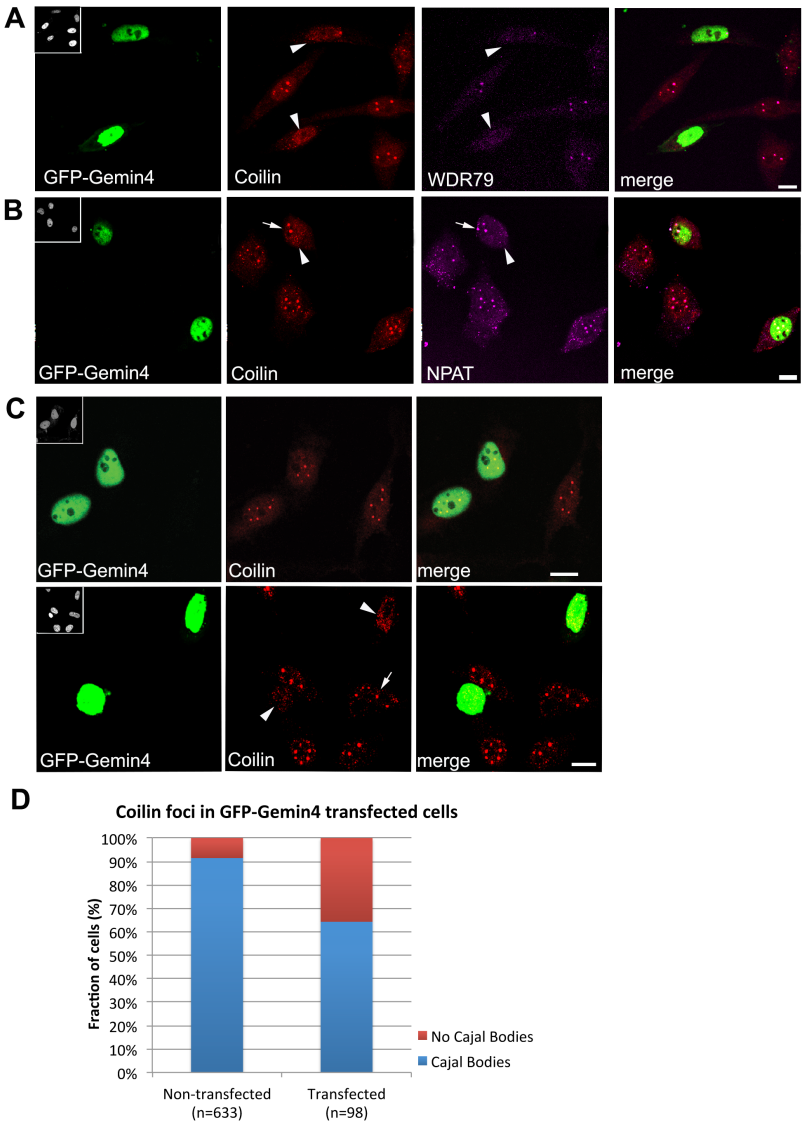

Figure 5. Cajal bodies are disrupted by high levels of GFP-Gemin4 expression. HeLa cells were transfected with GFP-Gemin4 (shown in green) and then co-stained (in red) with antibodies targeting Coilin (A-C), WDR79 (A) and NPAT (B). Inserts in upper left corners of GFP images show the DAPI stained nuclei. Arrowheads are used to identify transfected cells. Arrows indicate coilin foci corresponding to Cajal bodies. Cells strongly expressing GFP-Gemin4 show major disruptions to Cajal bodies, whereas lower expression levels had little effect (B). Nuclei were scored for the presence or absence of coilin foci (D). Chi squared analysis reveals a significant difference between the two cell populations, $p << 0.001$. Scale bars, 10 μm.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

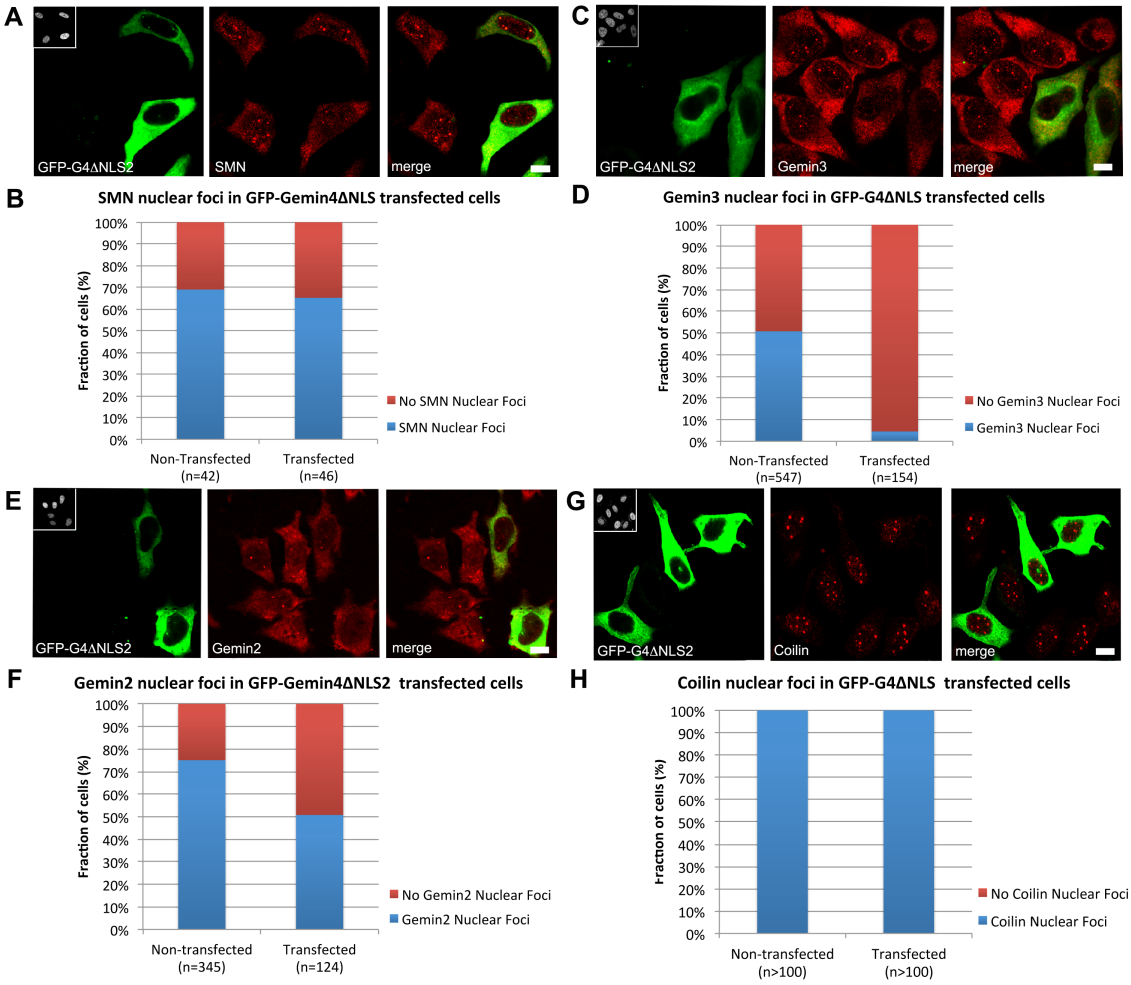

Figure 6. Immunofluorescence of endogenous SMN complex and coilin proteins in cells transfected with GFP-Gemin4ΔNLS2. HeLa cells were transfected with GFP-Gemin4ΔNLS2 (shown in green) and then co-stained (in red) with antibodies targeting SMN (A), Gemin2 (C), Gemin3 (E) or Coilin (G). As quantified above, GFP-Gemin4ΔNLS2 expression disrupts Gemin2 (F) and Gemin3 (D), but not SMN (B) or Coilin (H) nuclear foci. Chi squared analysis shows p=1.0 for coilin, p>0.7 for SMN and p << 0.001 for Gemin3, and Gemin2. Inserts in upper left corners of GFP images are the DAPI stained nuclei. Arrowheads are used to identify transfected cells with reduced cytoplasmic staining of endogenous proteins. Arrows indicate nuclear foci in non-transfected cells. Scale bars, 5 μm.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

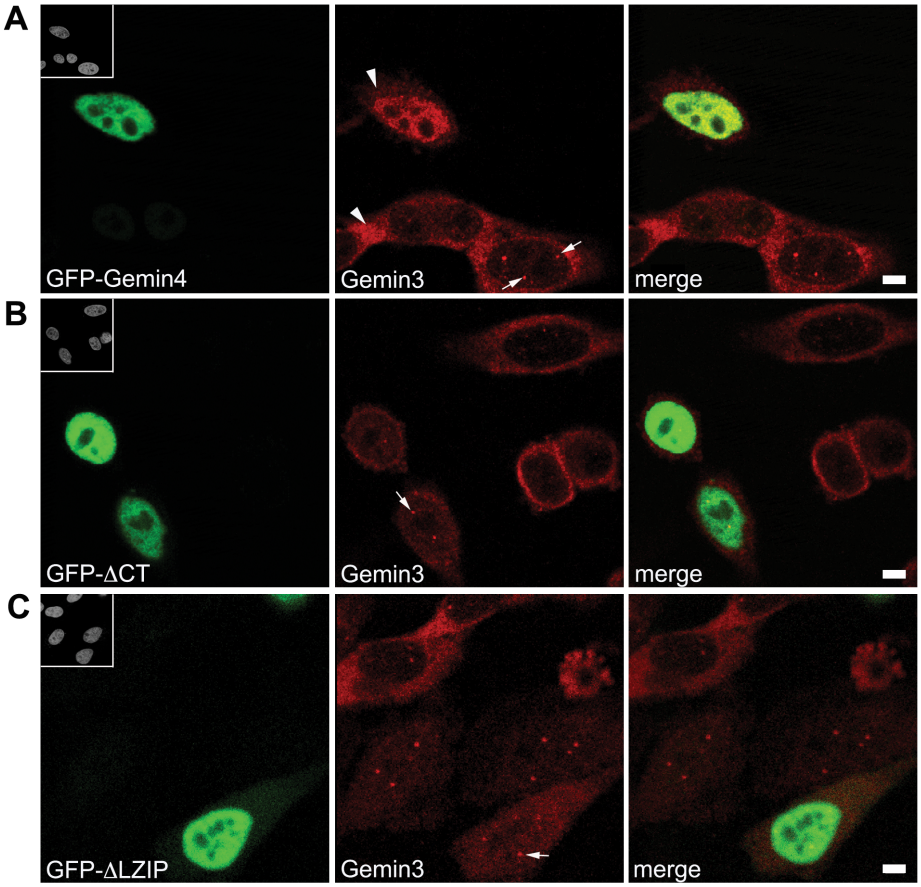

Figure 7. Gemin4-mediated nuclear relocalization of Gemin3 requires a C-terminal region within Gemin4. HeLa cells were transfected with various GFP-Gemin4 constructs (green) and then co-stained using anti-Gemin3 (red). Expression of full-length GFP-Gemin4 (A) relocalized Gemin3 to the nucleoplasm, whereas expression of an N-terminal truncation, GFP-ΔCT (B), or an internal deletion, GFP-ΔLZIP (C), did not. Arrowheads in panel (A) are used to illustrate the reduced cytoplasmic staining of endogenous Gemin3 in transfected cells, as compared to untransfected cells. Arrows mark Cajal bodies. Scale bars, 5 μm.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

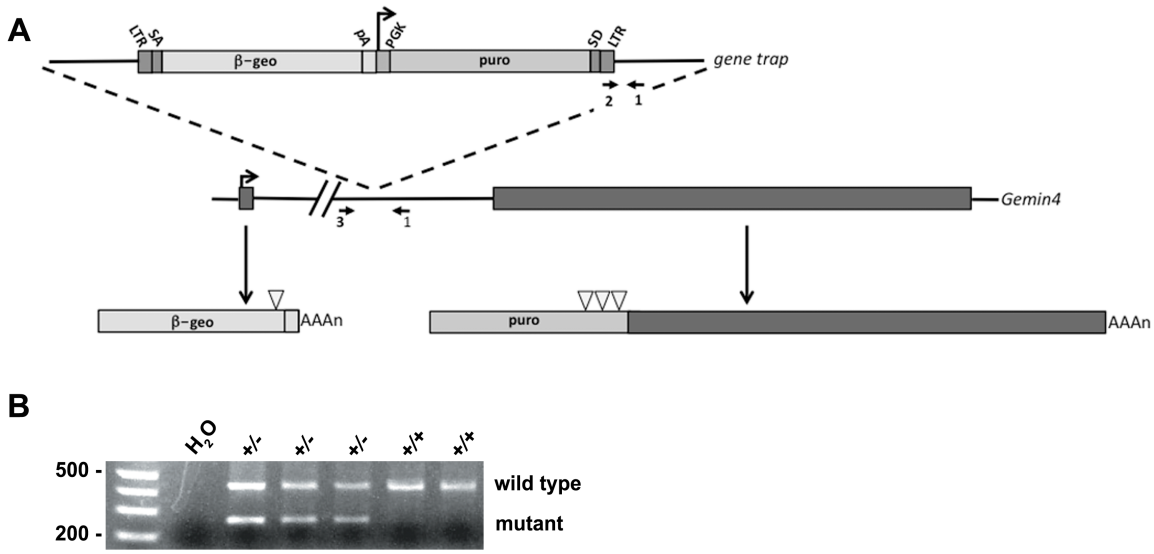

Figure 8. Schematic of Gemin4 gene trap. (A) Retroviral insertion of the gene trap vector into the single intron of murine *Gemin4* creates an upstream transcript that incorporates exon 1 with a β-geo cassette. Exon 2 is thus trapped into expressing a puromycin cassette that contains stop codons (open triangles) introduced in all three reading frames just upstream of the exonic sequence. The locations of the three PCR primers used in the genotyping assay are shown (arrows). LTR, retroviral long terminal repeat; SA, splice acceptor; SD, splice donor; PGK, phosphoglycerate kinase promoter region. (B) Germline transmission was observed in two of four chimeric males when crossed with a wild-type female. Primers 1 and 2 (shown in panel A) generate a 250 bp band derived from the mutant chromosome, whereas primers 1 and 3 generate a 450 bp wild-type band.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Tables and Legends

	Gemin4$^{+/+}$	*Gemin4*$^{+/-}$	*Gemin4*$^{-/-}$	Total
P1	30	69	0	99
E13.5	16	36	0	52
E7.5	18	33	0	51

Table 1. *Gemin4* is an essential gene in the mouse. *Gemin4* heterozygotes were intercrossed and the genotypes of F1 progeny were analyzed by PCR at P1, E13.5 and E7.5.

	Smn$^{+/+}$; *SMN2*$^{+/+}$	*Smn*$^{+/-}$; *SMN2*$^{+/+}$	*Smn*$^{-/-}$; *SMN2*$^{+/+}$	Total
C57BL/6J;129X1SvJ	19	42	0	61
C57BL/6J	20	33	0	53
FVB/NJ	20	40	28	88

Table 2. Analysis of genetic background contributions to the SMA phenotype. C57BL/6J and FVB/NJ mice that had been bred on pure (backcrossed more than 10 generations) inbred strains or those from a mixed background (C57BL/6J;129X1/SvJ).

	Observed	Expected	
Gemin4$^{+/+}$; *Smn*$^{-/-}$; *SMN2*$^{+/+}$	9	10	
Gemin4$^{+/-}$; *Smn*$^{-/-}$; *SMN2*$^{+/+}$	22	20	p-value = 0.859

Table 3. *Gemin4* heterozygotes do not modify the early lethality phenotype of type I SMA animals. Parental mice with the genotype of *Gemin4*$^{+/-}$;*Smn*$^{+/-}$;*SMN2*$^{+/+}$ were intercrossed and the resulting progeny were PCR genotyped to look for SMA type I-like mice that were either wild type or heterozygous for *Gemin4*.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplemental Figure

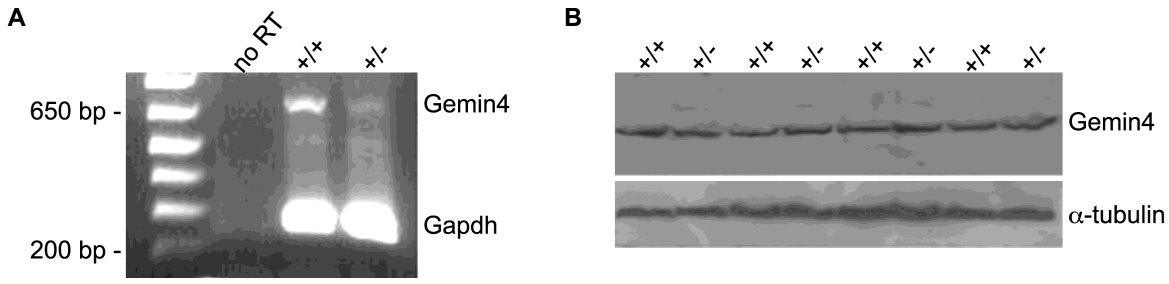

Figure S1. Expression of Gemin4 mRNA and protein. (A) RNA was isolated from livers of P4 *Gemin4*$^{+/-}$ and *Gemin4*$^{+/+}$ mice and semi-quantitative RT-PCR analysis was performed. Pooled *Gemin4*$^{+/-}$ and *Gemin4*$^{+/+}$ RNA for no RT control. (B) Western analysis of liver lysates was performed to determine protein levels of Gemin4 in wild-type and heterozygous animals; α-tubulin was used as a loading control.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

BATTLE, D. J., M. KASIM, J. WANG and G. DREYFUSS, 2007 SMN-independent subunits of the SMN complex. Identification of a small nuclear ribonucleoprotein assembly intermediate. J Biol Chem **282:** 27953-27959.

BATTLE, D. J., M. KASIM, J. YONG, F. LOTTI, C. K. LAU *et al.*, 2006a The SMN complex: an assembly machine for RNPs. Cold Spring Harb Symp Quant Biol **71:** 313-320.

BATTLE, D. J., C. K. LAU, L. WAN, H. DENG, F. LOTTI *et al.*, 2006b The Gemin5 protein of the SMN complex identifies snRNAs. Mol Cell **23:** 273-279.

BERNSTEIN, E., S. Y. KIM, M. A. CARMELL, E. P. MURCHISON, H. ALCORN *et al.*, 2003 Dicer is essential for mouse development. Nat Genet **35:** 215-217.

BRADRICK, S. S., and M. GROMEIER, 2009 Identification of gemin5 as a novel 7-methylguanosine cap-binding protein. PLoS One **4:** e7030.

BURGHES, A. H., and C. E. BEATTIE, 2009 Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? Nat Rev Neurosci **10:** 597-609.

CARISSIMI, C., J. BACCON, M. STRACCIA, P. CHIARELLA, A. MAIOLICA *et al.*, 2005 Unrip is a component of SMN complexes active in snRNP assembly. FEBS Lett **579:** 2348-2354.

CARISSIMI, C., L. SAIEVA, J. BACCON, P. CHIARELLA, A. MAIOLICA *et al.*, 2006a Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly. J Biol Chem **281:** 8126-8134.

CARISSIMI, C., L. SAIEVA, F. GABANELLA and L. PELLIZZONI, 2006b Gemin8 is required for the architecture and function of the survival motor neuron complex. J Biol Chem **281:** 37009-37016.

CAUCHI, R. J., 2010 SMN and Gemins: 'we are family' ... or are we?: insights into the partnership between Gemins and the spinal muscular atrophy disease protein SMN. Bioessays **32:** 1077-1089.

CHARI, A., M. M. GOLAS, M. KLINGENHAGER, N. NEUENKIRCHEN, B. SANDER *et al.*, 2008 An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. Cell **135:** 497-509.

CHARROUX, B., L. PELLIZZONI, R. A. PERKINSON, A. SHEVCHENKO, M. MANN *et al.*, 1999 Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. J Cell Biol **147:** 1181-1194.

CHARROUX, B., L. PELLIZZONI, R. A. PERKINSON, J. YONG, A. SHEVCHENKO *et al.*, 2000 Gemin4: A novel component of the SMN complex that is found in both gems and nucleoli. J Cell Biol **148:** 1177-1186.

CHEN, C. Z., L. LI, H. F. LODISH and D. P. BARTEL, 2004 MicroRNAs modulate hematopoietic lineage differentiation. Science **303:** 83-86.

DEAN, W. L., A. C. SEUFERT, G. A. SCHULTZ, R. S. PRATHER, C. SIMERLY *et al.*, 1989 The small nuclear RNAs for pre-mRNA splicing are coordinately regulated during oocyte maturation and early embryogenesis in the mouse. Development **106:** 325-334.

DI, Y., J. LI, Y. ZHANG, X. HE, H. LU *et al.*, 2003 HCC-associated protein HCAP1, a variant of GEMIN4, interacts with zinc-finger proteins. J Biochem **133:** 713-718.

DOSTIE, J., Z. MOURELATOS, M. YANG, A. SHARMA and G. DREYFUSS, 2003 Numerous microRNPs in neuronal cells containing novel microRNAs. Rna **9:** 180-186.

FISCHER, U., V. SUMPTER, M. SEKINE, T. SATOH and R. LUHRMANN, 1993 Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. EMBO J **12:** 573-583.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

HAO, L.-T., H. R. FULLER, L. T. LAM, T. T. LE, A. H. BURGHES *et al.*, 2007 Absence of gemin5 from SMN complexes in nuclear Cajal bodies. BMC Cell Biol **8:** 28.

HE, L., and G. J. HANNON, 2004 MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet **5:** 522-531.

HEBERT, M. D., K. B. SHPARGEL, J. K. OSPINA, K. E. TUCKER and A. G. MATERA, 2002 Coilin methylation regulates nuclear body formation. Dev Cell **3:** 329-337.

HEBERT, M. D., P. W. SZYMCZYK, K. B. SHPARGEL and A. G. MATERA, 2001 Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. Genes Dev **15:** 2720-2729.

HOUBAVIY, H. B., M. F. MURRAY and P. A. SHARP, 2003 Embryonic stem cell-specific MicroRNAs. Dev Cell **5:** 351-358.

HSIEH-LI, H. M., J. G. CHANG, Y. J. JONG, M. H. WU, N. M. WANG *et al.*, 2000 A mouse model for spinal muscular atrophy. Nat Genet **24:** 66-70.

HUBER, J., U. CRONSHAGEN, M. KADOKURA, C. MARSHALLSAY, T. WADA *et al.*, 1998 Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. EMBO J **17:** 4114-4126.

HUBER, J., A. DICKMANNS and R. LUHRMANN, 2002 The importin-beta binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs in vitro. J Cell Biol **156:** 467-479.

HUTVAGNER, G., and P. D. ZAMORE, 2002 A microRNA in a multiple-turnover RNAi enzyme complex. Science **297:** 2056-2060.

JABLONKA, S., B. HOLTMANN, G. MEISTER, M. BANDILLA, W. ROSSOLL *et al.*, 2002 Gene targeting of Gemin2 in mice reveals a correlation between defects in the biogenesis of U snRNPs and motoneuron cell death. Proc Natl Acad Sci U S A **99:** 10126-10131.

JARMOLOWSKI, A., W. C. BOELENS, E. IZAURRALDE and I. W. MATTAJ, 1994 Nuclear export of different classes of RNA is mediated by specific factors. J Cell Biol **124:** 627-635.

KROISS, M., J. SCHULTZ, J. WIESNER, A. CHARI, A. SICKMANN *et al.*, 2008 Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster. Proc Natl Acad Sci U S A **105:** 10045-10050.

LAU, C. K., J. L. BACHORIK and G. DREYFUSS, 2009 Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. Nat Struct Mol Biol **16:** 486-491.

LEFEBVRE, S., L. BURGLEN, S. REBOULLET, O. CLERMONT, P. BURLET *et al.*, 1995 Identification and characterization of a spinal muscular atrophy-determining gene. Cell **80:** 155-165.

LEMM, I., C. GIRARD, A. N. KUHN, N. J. WATKINS, M. SCHNEIDER *et al.*, 2006 Ongoing U snRNP biogenesis is required for the integrity of Cajal bodies. Mol Biol Cell **17:** 3221-3231.

LOBO, S. M., W. F. MARZLUFF, A. C. SEUFERT, W. L. DEAN, G. A. SCHULTZ *et al.*, 1988 Localization and expression of U1 RNA in early mouse embryo development. Dev Biol **127:** 349-361.

LORSON, M. A., A. M. DICKSON, D. J. SHAW, A. G. TODD, E. C. YOUNG *et al.*, 2008 Identification and characterisation of a nuclear localisation signal in the SMN associated protein, Gemin4. Biochem Biophys Res Commun **375:** 33-37.

MA, Y., J. DOSTIE, G. DREYFUSS and G. D. VAN DUYNE, 2005 The Gemin6-Gemin7 heterodimer from the survival of motor neurons complex has an Sm protein-like structure. Structure (Camb) **13:** 883-892.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

MARSHALLSAY, C., and R. LUHRMANN, 1994 In vitro nuclear import of snRNPs: cytosolic factors mediate m3G-cap dependence of U1 and U2 snRNP transport. EMBO J **13:** 222-231.

MATERA, A. G., R. M. TERNS and M. P. TERNS, 2007 Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol **8:** 209-220.

MATERA A.G. and Z. WANG, 2014 A day in the life of the spliceosome. Nature Reviews Molecular Cell Biology **15**: 108–121.

MONANI, U. R., M. SENDTNER, D. D. COOVERT, D. W. PARSONS, C. ANDREASSI *et al.*, 2000 The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in Smn(-/-) mice and results in a mouse with spinal muscular atrophy. Hum Mol Genet **9:** 333-339.

MOUILLET, J. F., X. YAN, Q. OU, L. JIN, L. J. MUGLIA *et al.*, 2008 DEAD-box protein-103 (DP103, Ddx20) is essential for early embryonic development and modulates ovarian morphology and function. Endocrinology **149:** 2168-2175.

MOURELATOS, Z., J. DOSTIE, S. PAUSHKIN, A. SHARMA, B. CHARROUX *et al.*, 2002 miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev **16:** 720-728.

NARAYANAN, U., T. ACHSEL, R. LUHRMANN and A. G. MATERA, 2004 Coupled in vitro import of U snRNPs and SMN, the Spinal Muscular Atrophy protein. Mol Cell **16:** 223-234.

NARAYANAN, U., J. K. OSPINA, M. R. FREY, M. D. HEBERT and A. G. MATERA, 2002 SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. Hum Mol Genet **11:** 1785-1795.

NELSON, P. T., A. G. HATZIGEORGIOU and Z. MOURELATOS, 2004 miRNP:mRNA association in polyribosomes in a human neuronal cell line. RNA **10:** 387-394.

OHNO, M., A. SEGREF, A. BACHI, M. WILM and I. W. MATTAJ, 2000 PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell **101:** 187-198.

OTTER, S., M. GRIMMLER, N. NEUENKIRCHEN, A. CHARI, A. SICKMANN *et al.*, 2007 A comprehensive interaction map of the human survival of motor neuron (SMN) complex. J Biol Chem **282:** 5825-5833.

PALACIOS, I., M. HETZER, S. A. ADAM and I. W. MATTAJ, 1997 Nuclear import of U snRNPs requires importin beta. EMBO J **16:** 6783-6792.

PARK, J. W., P. G. VOSS, S. GRABSKI, J. L. WANG and R. J. PATTERSON, 2001 Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein. Nucleic Acids Res **29:** 3595-3602.

PELLIZZONI, L., J. YONG and G. DREYFUSS, 2002 Essential role for the SMN complex in the specificity of snRNP assembly. Science **298:** 1775-1779.

SCHRANK, B., R. GOTZ, J. M. GUNNERSEN, J. M. URE, K. V. TOYKA *et al.*, 1997 Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. Proc Natl Acad Sci U S A **94:** 9920-9925.

SHPARGEL, K. B., and A. G. MATERA, 2005 Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. Proc Natl Acad Sci U S A **102:** 17372-17377.

SHPARGEL, K. B., K. PRAVEEN, T. K. RAJENDRA and A. G. MATERA, 2009 Gemin3 is an essential gene required for larval motor function and pupation in Drosophila. Mol Biol Cell **20:** 90-101.

bioRxiv preprint doi: https://doi.org/10.1101/242529; this version posted January 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

STRZELECKA, M., S. TROWITZSCH, G. WEBER, R. LUHRMANN, A. C. OATES *et al.*, 2010 Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. Nat Struct Mol Biol **17:** 403-409.

YAN, X., J. F. MOUILLET, Q. OU and Y. SADOVSKY, 2003 A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity. Mol Cell Biol **23:** 414-423.

YONG, J., M. KASIM, J. L. BACHORIK, L. WAN and G. DREYFUSS, 2010 Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. Mol Cell **38:** 551-562.

ZAMBROWICZ, B. P., G. A. FRIEDRICH, E. C. BUXTON, S. L. LILLEBERG, C. PERSON *et al.*, 1998 Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature **392:** 608-611.