Unusual intragenic suppression of an IFT52 gene disruption links hypoxia to the intraflagellar transport in *Tetrahymena thermophila*.

Drashti Dave, Gautham Pandiyan, Dorota Wloga[1] and Jacek Gaertig*

Department of Cellular Biology, University of Georgia, Athens, GA 30602.

[1]Present address: Laboratory of Cytoskeleton and Cilia Biology, Nencki Institute of Experimental Biology, Warsaw 02-093, Poland

*Correspondance to: jgaertig@uga.edu

Abstract

IFT52 protein is a conserved intraflagellar transport protein (a part of the IFT complex B) that is essential for assembly and maintenance of cilia. *Tetrahymena* null mutants with an insertion of a *neo* gene cassette into the *IFT52* gene undergo frequent suppressions that lead to conditional assembly of cilia only under hypoxic conditions (Brown et al. 2003). Here we show that these conditional suppressions are intragenic and occur by a novel mechanism. First, the non-native (bacterial) portion of the DNA sequence of the *neo* cassette is deleted during the process of genome rearrangement that occurs in the developing macronucleus of conjugating *Tetrahymena*. Next, the residual sequences of the *neo* cassette (of *Tetrahymena* origin) within the IFT52 mRNA are recognized as multiple introns and undergo splicing, leading to a restoration of the translational frame of *IFT52*. The resulting hypoxia-dependent IFT52 protein contains an insertion of 43 new amino acids that replace 7 original amino acids. Taken together with a study in *Chlamydomonas reinhardtii* showing a hypoxia-dependence of another IFT subunit mutant, IFT46, (Hou et al. 2007), our observations generalize that defective IFT complex subunits can regain functionality under hypoxia.

Results and Discussion

Intraflagellar transport (IFT) is a bidirectional motility of ciliary precursors that occurs inside cilia (Kozminski et al. 1993). Kinesin-2 is the anterograde IFT motor, whereas cytoplasmic dynein1b is responsible for the retrograde IFT (Kozminski et al. 1995; Pazour et al. 1999; Porter et al. 1999). These motors move IFT trains, that are composed

of two protein complexes, A and B (Cole et al. 1998; Piperno and Mead 1997). IFT52 is a complex B protein that is required for the assembly and maintenance of cilia (Brazelton et al. 2001; Brown et al. 2003; Deane et al. 2001).

The ciliate *Tetrahymena thermophila* has two nuclei, a transcriptionally silent micronucleus (Mic) and a transcriptionally active macronucleus (Mac) (reviewed in (Yao and Chao 2005)). Earlier (Brown et al. 2003), *IFT52* was disrupted by insertion of the *neo2* marker within exon 4 (Figure 1D). Heterokaryons were constructed with Macs carrying wild-type *IFT52* alleles and Mics homozygous for the disrupted alleles. Most progeny cells of mating *IFT52* heterokaryons are completely paralyzed due to the lack of cilia and cannot complete cytokinesis since they are unable to rupture the connecting cytoplasmic bridge (rotokinesis). Surprisingly, 3% of the heterokaryon progeny recover partial motility due to spontaneous suppressions. Importantly, the suppressed cells (IFT52Δsm) assemble motile cilia when grown at either a lower temperature or in hypoxia. In a single suppressed strain, an additional event produced IFT52Δmov cells, which are capable of assembling cilia independently of temperature or hypoxia (Brown et al. 2003).

The high frequency of the IFT52Δsm conditional suppressions and the fact that these suppressions occur only during conjugation (Brown et al. 2003), suggested that the mechanism of suppression is based on processes that occur inside the developing new Mac. Conjugating *Tetrahymena* cells undergo a series of nuclear events that culminate in replacement of the parental Mac by a new Mac that develops by differentiation from a zygotic Mic (reviewed in (Coyne et al. 1996)). About 15% of the Mic genome is removed from the new Mac, by a pathway that involves an RNAi-dependent sequence recognition

and degradation (reviewed in (Yao and Chao 2005)). Yao and colleagues showed that a foreign sequence, *neo2*, inserted into multiple loci, undergoes RNAi-mediated deletion (Yao et al. 2003). Thus, we tested whether *neo2* inserted into *IFT52* also undergoes deletions that could be a cause of the conditional suppressions.

The *IFT52* knockout was done by inserting the *neo2* disruption cassette into exon 4 (Figure 1D). *neo2* consists of the bacterial neomycin phosphotransferase (*neo*) coding region placed between DNA fragments of *Tetrahymena* origin; the *HHF4* promoter and the *BTU2* transcription terminator (Gaertig et al. 1994; Kahn et al. 1993). We isolated total genomic DNA from wild-type, IFT52Δ, IFT52Δsm and IFT52Δmov cells and amplified the *IFT52* locus across the *neo2* insertion site (Figure 1D). Amplification of genomic DNA of wild-type cells produced a fragment of expected size (1.3kb). The same primers used with IFT52Δ (non-suppressed) DNA produced a larger fragment (~2.7kb) consistent with presence of an intact *neo2* cassette (Figure 1A). Strikingly, the same primers amplified a smaller fragment (~1.9kb) from the genomic DNA of both conditional and non-conditional suppressors (IFT52Δsm and IFT52Δmov). This suggested that the suppressions are associated with deletions around the *neo2* insertion site. Sequencing of fragments amplified from multiple independent suppressor strains showed deletions of a portion of *neo* (~0.8kb) with deletion junctions at exactly the same positions, while the flanking sequences of *neo2* (of *Tetrahymena* origin) remained largely intact (Figure 1D). Specifically, all deletions analyzed were between the nucleotide at position +45 in the *neo* coding sequence and the fifth nucleotide downstream of the stop codon within the *BTU2* segment (Figure S1). These observations are consistent with

earlier reports on deletions of *neo2* sequences during macronuclear development (Liu et al. 2005; Yao et al. 2003).

The observed *neo* sequence deletions, do not explain the mechanism of suppression because the sequence of the *neo2* cassette remnant has stop codons in all forward translational frames. An Ift52p translated from the predicted mRNA containing the *neo2* remnant would be severely truncated; lacking 5 out of 7 exons, all containing conserved sequences (Cole 2003) (Figure S1). Nevertheless, the suppressions correlate with deletions of the *neo* coding sequence.

To establish whether the suppressed *IFT52* locus with the residual *neo2* is sufficient to restore partial motility, we introduced the rearranged fragment of the *IFT52Δmov* genomic DNA into IFT52Δ cells by biolistic bombardment (Figure 1D). As a control we mock-transformed the same number of IFT52Δ cells ($9x10^6$). After 7-9 days of incubation at room temperature, we obtained 2 clones that regained motility in the population bombarded with the rearranged (*IFT52Δmov)* fragment and none in the mock-transformed IFT52Δ cells. We confirmed that the targeting fragment replaced the corresponding region of the fully disrupted *IFT52* locus by PCR (results not shown). The rescued cells showed the conditional suppression phenotype, a cell density (pericellular hypoxia)-dependent ciliary motility (Brown et al. 2003) (Figure 1C and results not shown). These data indicate that the IFT52Δmov cells underwent an additional, unknown genetic or epigenetic change that resulted in a non-conditional suppression.

The rearranged *IFT52Δsm/mov* gene contains a residual *neo2* sequence that somehow provides a partially functional Ift52p. Either an extremely truncated Ift52p is sufficient for conditional ciliary assembly or an additional mechanism restores the

translational frame across the residual *neo2*. To determine the sequence of the translated Ift52p in IFT52Δsm cells, we used RT-PCR to amplify the *IFT52* cDNA obtained from mRNA of IFT52Δsm cells (Figure 1D). For a spliced wild-type *IFT52* mRNA, the amplified fragment was expected to be ~0.3kb. For the *IFT52Δsm* mRNA with residual *neo2* cassette, the cDNA fragment was expected to be ~0.9kb. However, the size of the amplified product from the IFT52Δsm cDNA was ~0.4kb, indicating that an additional splicing event occurs in the IFT52Δsm mRNA (Figure 1B). The sequencing of a cloned IFT52Δsm cDNA revealed that ~0.8kb of the residual *neo2* was absent. Most of the residual *neo2* sequence, mainly comprising of the *HHF4* and *BTU2* sequences, was removed from the mRNA as 3 (artificial) introns. The artificial intron junctions have sequences consistent with the native intron junctions observed in ciliates such as *Paramecium* and *Tetrahymena* (Figure S1) (Jaillon et al. 2008). The processing of the residual *neo2* as a set of artificial introns restores the translational frame across the site of *neo2* insertion (Figure S2). Hence, the predicted suppressor Ift52p has 43 additional amino acids but lacks 7 original amino acids as a result of the *neo2* cloning procedure (Figures 1D and S2). Either the presence of these extra amino acids or the absence of the 7 endogenous amino acids in Ift52p-sm (or both) results in the intragenic conditional suppression.

To conclude, we reveal a novel mechanism for intragenic suppression in *Tetrahymena* that consists of two steps: 1) foreign DNA within the inserted disruption cassette is deleted during macronuclear development, and 2) the remaining AT-rich *Tetrahymena* native sequences of the disruption marker are processed as introns during

mRNA splicing. The first step almost certainly occurs via the RNAi-mediated developmental genome rearrangement pathway (Mochizuki et al. 2002; Mochizuki and Gorovsky 2004; Yao et al. 2003). This form of genomic DNA deletion is thought to have evolved as a means of genome surveillance to eliminate transposon DNA from the transcriptionally active Mac (reviewed in (Yao and Chao 2005)). In 4 independent suppressor clones we detected a genomic deletion at precisely identical positions. Previous studies showed a variability in the deletion sites (Liu et al. 2005; Yao et al. 2003). It is likely that other deletions occur in the disrupted *IFT52* locus but they do not create potential splice junctions that restore the translational frame. When *IFT52* heterokaryons undergo conjugation, the majority (97%) of the progeny has a non-suppressed phenotype. In these cells, the deletions of *neo2* either do not occur, or occur on an insufficient number of macronuclear chromosomes to achieve a phenotypic threshold for suppression (there are 45 copies of each chromosome in the G1 macronucleus).

Chlamydomonas cells carrying an insertional mutation in IFT46 (encoding another complex B protein), also underwent a spontaneous intragenic mutation that led to a hypoxia-dependent cilia assembly (Hou et al. 2007). Both studies taken together ((Hou et al. 2007) and this work), allow for a generalization; that hypoxic conditions can restore the functionality of mutated IFT complex B components. Hou and colleagues observed the assembly of complex B in the flagella of suppressed *Chlamydomonas* IFT46 mutants. It is likely that the suppressed IFT52Δsm *Tetrahymena* cells also assemble complex B. Hou and colleagues proposed that the IFT complex B subunits are folded by a chaperone whose levels increase under hypoxia. Thus, a partly damaged IFT component may still

fold properly when the chaperone activity is increased. Another possibility is that the IFT complex B assembly is regulated directly by an oxygen-dependent post-translational modification of one or more subunits. Regardless of the exact mechanism, both studies indicate that a hypoxia-dependent modulation of the activity of IFT complex B subunits is a conserved mechanism.

Materials and Methods

Cells, cultures and media

For the maintenance, IFT52Δ, IFT52Δsm and IFT52Δmov cells were grown at the room temperature in MEPP medium (Orias and Rasmussen 1976) with an antibiotic-antimycotic mixture (Invitrogen, Carlsbad, CA).

DNA preparation and cloning

Isolation of total genomic DNA was done as described (Dave et al. 2009). The genomic region across exons 3 and 4 was amplified using the following primers: 5'-ATGCCCTCAAATAAT-3' and 5'-TAGAGTTGGTTTAGATTT-3'. The resulting fragments were cloned into pGEM-T-vector (Promega Corp, Madison, WI) and sequenced.

Biolistic transformation of Tetrahymena

To determine whether a genomic fragment of IFT52 of suppressor origin is sufficient to confer suppression of the IFT52Δ phenotype (lack of cilia), IFT52Δ cells were

biolistically bombarded with a genomic fragment of IFT52Δmov origin, that was earlier separated from the pGEM-T-vector plasmid with NcoI and SalI digestion. Bombarded cells were grown at the room temperature and transformants were identified based on recovery of cell motility (Cassidy-Hanley et al. 1997).

cDNA preparation

Cells were grown to a concentration of 2 x 10^5 cells/ml in MEPP medium (Gorovsky et al. 1975), washed with 10 mM Tris-HCl buffer pH 7.5 and used for total RNA extraction with TRI-reagent (MRC Inc, Cincinnati, OH) according to manufacturer's instructions. Total cDNA was prepared using the SMART IV-forward and CDS III-reverse primers from the RT-PCR kit (Clontech Inc, Mountainview, CA).

Acknowledgements

This work was supported by the National Science Foundation grant MCB-063994 to JG.

References

Brazelton, W. J., Amundsen, C. D., Silflow, C. D., Lefebvre, P. A., (2001). The bld1 mutation identifies the Chlamydomonas osm-6 homolog as a gene required for flagellar assembly. Curr Biol **11**(20): 1591-4.

Brown, J. M., Hardin, C., Gaertig, J., (1999). Rotokinesis, a novel phenomenon of cell locomotion-assisted cytokinesis in the ciliate Tetrahymena thermophila. Cell Biol Int **23**(12): 841-8.

Brown, J. M., Fine, N. A., Pandiyan, G., Thazhath, R., Gaertig, J., (2003). Hypoxia regulates assembly of cilia in suppressors of Tetrahymena lacking an intraflagellar transport subunit gene. Mol Biol Cell **14**(8): 3192-207.

Cassidy-Hanley, D., Bowen, J., Lee, J. H., Cole, E., VerPlank, L. A., Gaertig, J., Gorovsky, M. A., Bruns, P. J., (1997). Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. Genetics **146**(1): 135-47.

Cole, D. G., Diener, D. R., Himelblau, A. L., Beech, P. L., Fuster, J. C., Rosenbaum, J. L., (1998). Chlamydomonas kinesin-II-dependent intraflagellar transport (IFT): IFT particles contain proteins required for ciliary assembly in Caenorhabditis elegans sensory neurons. J Cell Biol **141**(4): 993-1008.

Cole, D. G., (2003). The intraflagellar transport machinery of Chlamydomonas reinhardtii. Traffic **4**(7): 435-42.

Coyne, R. S., Chalker, D. L., Yao, M. C., (1996). Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring. Annu Rev Genet **30**: 557-78.

Dave, D., Wloga, D., Gaertig, J., (2009). Manipulating ciliary protein-encoding genes in Tetrahymena thermophila. Methods Cell Biol **93**: 1-20.

Deane, J. A., Cole, D. G., Seeley, E. S., Diener, D. R., Rosenbaum, J. L., (2001). Localization of intraflagellar transport protein IFT52 identifies basal body transitional fibers as the docking site for IFT particles. Curr Biol **11**(20): 1586-90.

Gaertig, J., Gu, L., Hai, B., Gorovsky, M. A., (1994). High frequency vector-mediated transformation and gene replacement in Tetrahymena. Nucleic Acids Res **22**(24): 5391-8.

Gorovsky, M. A., Yao, M. C., Keevert, J. B., Pleger, G. L., (1975). Isolation of micro- and macronuclei of Tetrahymena pyriformis. Methods Cell Biol **9**(0): 311-27.

Hou, Y., Qin, H., Follit, J. A., Pazour, G. J., Rosenbaum, J. L., Witman, G. B., (2007). Functional analysis of an individual IFT protein: IFT46 is required for transport of outer dynein arms into flagella. J Cell Biol **176**(5): 653-65.

Jaillon, O., Bouhouche, K., Gout, J. F., Aury, J. M., Noel, B., Saudemont, B., Nowacki, M., Serrano, V., Porcel, B. M., Segurens, B., Le Mouel, A., Lepere, G., Schachter, V., Betermier, M., Cohen, J., Wincker, P., Sperling, L., Duret, L., Meyer, E., (2008). Translational control of intron splicing in eukaryotes. Nature **451**(7176): 359-62.

Kahn, R. W., Andersen, B. H., Brunk, C. F., (1993). Transformation of Tetrahymena thermophila by microinjection of a foreign gene. Proc Natl Acad Sci U S A **90**(20): 9295-9.

Kozminski, K. G., Johnson, K. A., Forscher, P., Rosenbaum, J. L., (1993). A motility in the eukaryotic flagellum unrelated to flagellar beating. Proc Natl Acad Sci U S A **90**(12): 5519-23.

Kozminski, K. G., Beech, P. L., Rosenbaum, J. L., (1995). The Chlamydomonas kinesin-like protein FLA10 is involved in motility associated with the flagellar membrane. J Cell Biol **131**(6 Pt 1): 1517-27.

Liu, Y., Song, X., Gorovsky, M. A., Karrer, K. M., (2005). Elimination of foreign DNA during somatic differentiation in Tetrahymena thermophila shows position effect and is dosage dependent. Eukaryot Cell **4**(2): 421-31.

Mochizuki, K., Fine, N. A., Fujisawa, T., Gorovsky, M. A., (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell **110**(6): 689-99.

Mochizuki, K., Gorovsky, M. A., (2004). Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. Genes Dev **18**(17): 2068-73.

Orias, E., Rasmussen, L., (1976). Dual capacity for nutrient uptake in Tetrahymena. IV. Growth without food vacuoles and its implications. Exp Cell Res **102**(1): 127-37.

Pazour, G. J., Dickert, B. L., Witman, G. B., (1999). The DHC1b (DHC2) isoform of cytoplasmic dynein is required for flagellar assembly. J Cell Biol **144**(3): 473-81.

Piperno, G., Mead, K., (1997). Transport of a novel complex in the cytoplasmic matrix of Chlamydomonas flagella. Proc Natl Acad Sci U S A **94**(9): 4457-62.

Porter, M. E., Bower, R., Knott, J. A., Byrd, P., Dentler, W., (1999). Cytoplasmic dynein heavy chain 1b is required for flagellar assembly in Chlamydomonas. Mol Biol Cell **10**(3): 693-712.

Yao, M. C., Fuller, P., Xi, X., (2003). Programmed DNA deletion as an RNA-guided system of genome defense. Science **300**(5625): 1581-4.

Yao, M. C., Chao, J. L., (2005). RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. Annu Rev Genet **39**: 537-59.

Figure legends

Figure 1: Two subsequent sequence deletions lead to intragenic suppression of an insertional *IFT52* mutation. A. Results of PCR amplifications of the genomic region of IFT52 locus with primers corresponding to sequences in exons 3 and 4 in wildtype (*IFT52*), gene knockout (*IFT52Δ*) and the suppressor (*IFT52Δsm/mov*) cells. Amplified fragments were separated on an agarose gel. An asterisk marks an apparent non-specific amplification product. B. PCR amplifications of total cDNA obtained from mRNA using wildtype (*IFT52*), knockout (*IFT52Δ*) and suppressors (*IFT52Δsm/mov*) cells using the same primers as in panel A. The amplification products were separated on an agarose gel. C. IFT52Δ cells rescued with *IFT52Δsm/mov* DNA and wild-type cells both at a concentration of 3x10^5 were diluted down to different concentrations (1x to 100x) on a

96-well microtiter plate, and incubated at 30°C. After 12 hours of incubation, the number of motile cells (cells showing detectable displacement and lacking cytokinesis defects) was determined. D. A schematic diagram detailing the *IFT52* locus in the wild-type (*IFT52*), knockout (*IFT52Δ*) and the suppressors (*IFT52Δsm/mov*) cells and the cDNA in the suppressors (*IFT52Δsm/mov processed mRNA)* cells. F and R represent primers used for the various PCR reactions.

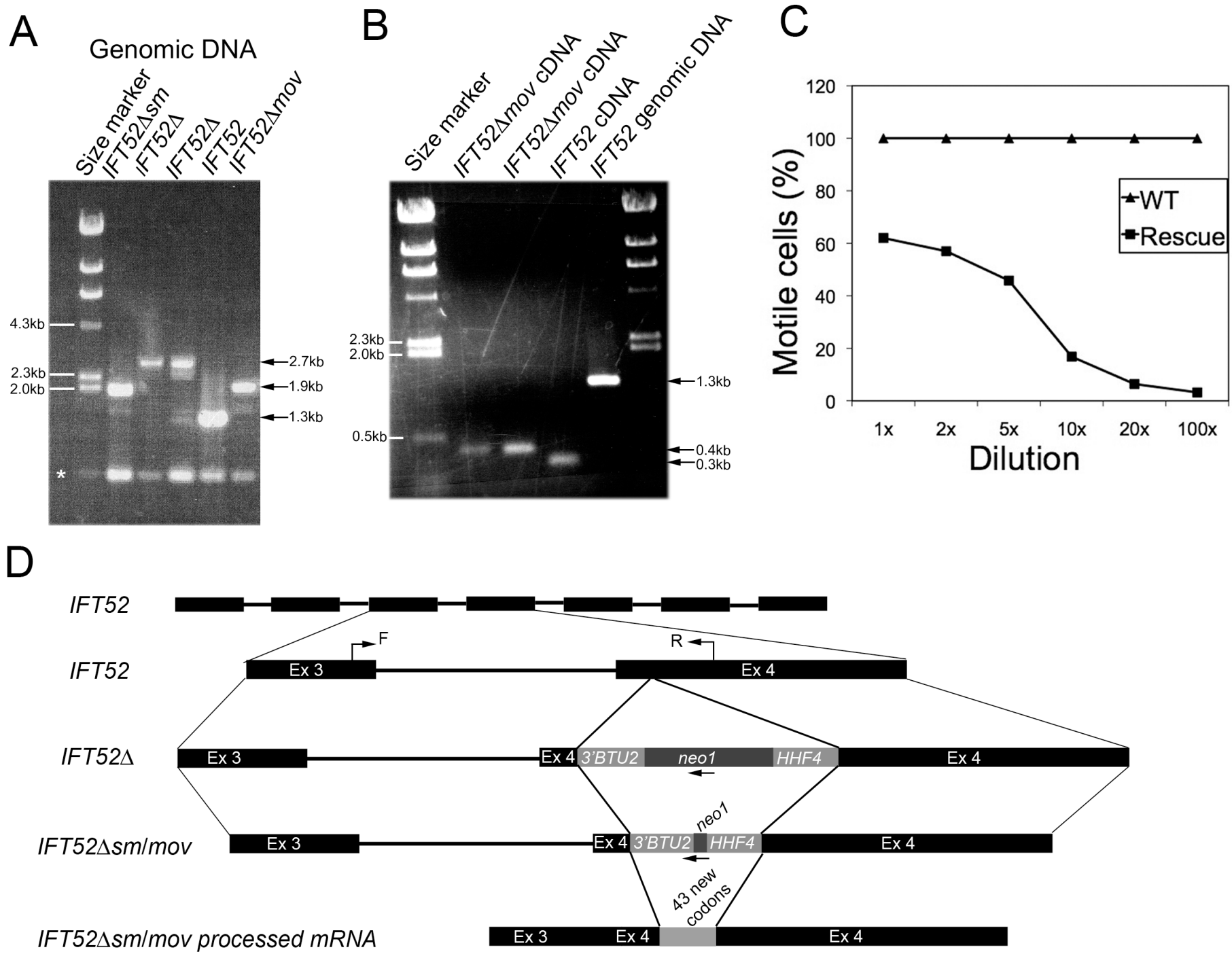
A
Genomic DNA
Size marker
IFT52Δsm
IFT52Δ
IFT52Δ
IFT52
IFT52Δmov
4.3kb
2.3kb
2.0kb
*
2.7kb
1.9kb
1.3kb
B
Size marker
IFT52Δmov cDNA
IFT52Δmov cDNA
IFT52 cDNA
IFT52 genomic DNA
2.3kb
2.0kb
0.5kb
1.3kb
0.4kb
0.3kb
C
Motile cells (%)
120
100
80
60
40
20
0
WT
Rescue
1x
2x
5x
10x
20x
100x
Dilution
D
IFT52
IFT52
F
R
Ex 3
Ex 4
IFT52Δ
Ex 3
Ex 4
3'BTU2
neo1
HHF4
Ex 4
IFT52Δsm/mov
Ex 3
Ex 4
neo1
3'BTU2
HHF4
Ex 4
43 new codons
IFT52Δsm/mov processed mRNA
Ex 3
Ex 4
Ex 4

Figure 1

Supplemental data

Figure S1: Sequence of genomic DNA reconstructed from 4 independent IFT52Δsm/mov suppressor strains shows a deletion within the *neo2* cassette. The sequence shown corresponds to the region between exons 3 and 4 of the suppressed *IFT52* genomic DNA. The three segments of the *neo2* cassette are marked in blue (*BTU2*), pink (*neo*) and green (*HHF4*). Natural intron junctions are marked as grey boxes. Artificial intron junctions are marked as open boxes. Note: Within the residual *neo* cassette left after genomic deletion, there are stop codons (red *) in every translational frame.

Figure S2: The cDNA sequence of *IFT52Δsm/mov* has 43 extra codons from residual *neo2* cassette. The residual *neo2* cassette consists of bacterial and *Tetrahymena* (*BTU2* and *HHF4*) sequences after being processed as artificial introns. The sequence shown corresponds to the region between exons 3 and 4 of the suppressed *IFT52* cDNA.

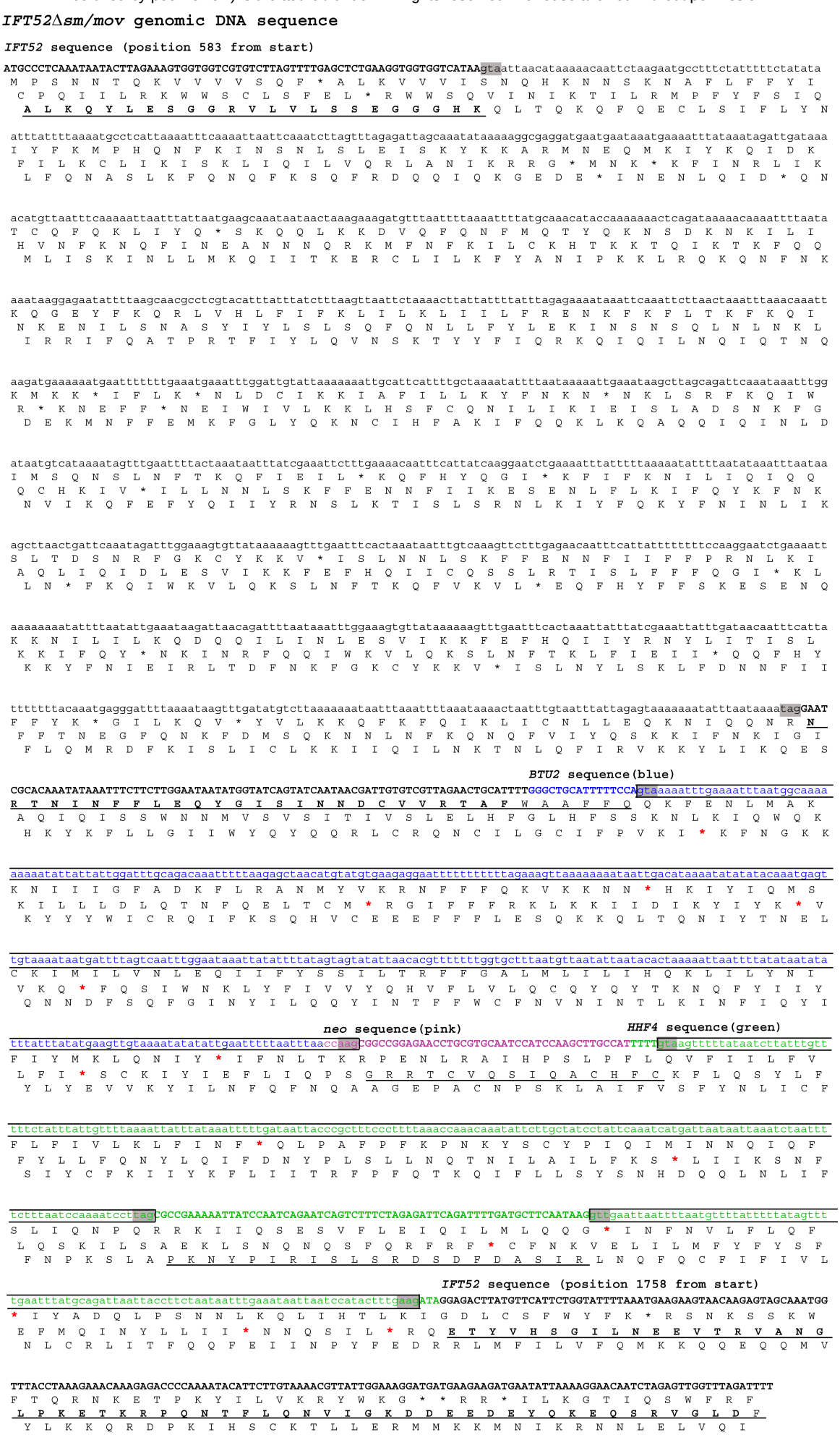

Figure S1

IFT52Δsm/mov cDNA sequence (470 codons)

IFT52 cDNA(black)

```
ATG AGT GGA GAA TAG AAA ATT ATT GTC TTC AAC GCT TCA AAG AAG GAG GCT GGT AAC CCT AGT ACA AAT ATT AAA AAG ATT ATC AAG AAA
M   S   G   E   Q   K   I   I   V   F   N   A   S   K   K   E   A   G   N   P   S   T   N   I   K   K   I   I   K   K

TAT AAA GAA ACA TAT AAG TGC GGC AGA AAT AAA GAA GAT ATC ACA TAT GAT AGA CTA AAG ATG GCT TCT TTG GTC ATA TTT TTT TGC CCC
Y   K   E   T   Y   K   C   G   R   N   K   E   D   I   T   Y   D   R   L   K   M   A   S   L   V   I   F   F   C   P

AAA GAA ATG TTT ACC AAG GAG GAA TTC GAT GCC CTC AAA TAA TAC TTA GAA AGT GGT GGT CGT GTC TTA GTT TTG AGC TCT GAA GGT GGT
K   E   M   F   T   K   E   E   F   D   A   L   K   Q   Y   L   E   S   G   G   R   V   L   V   L   S   S   E   G   G

GGT CAT AAG AAT CGC ACA AAT ATA AAT TTC TTC TTG GAA TAA TAT GGT ATC AGT ATC AAT AAC GAT TGT GTC GTT AGA ACT GCA TTT Tgg
G   H   K   N   R   T   N   I   N   F   F   L   E   Q   Y   G   I   S   I   N   N   D   C   V   V   R   T   A   F   W
```

BTU2 cDNA(blue) *neo* cDNA(pink) *HHF4* cDNA(green)

```
gct gca ttt ttc cac ggc cgg aga acc tgc gtg caa tcc atc caa gct tgc cat ttt tcg ccg aaa aat tat cca atc aga atc agt ctt
A   A   F   F   H   G   R   R   T   C   V   Q   S   I   Q   A   C   H   F   S   P   K   N   Y   P   I   R   I   S   L
```

IFT52 cDNA(black)

```
tct aga gat tca gat ttt gat gct tca ata aga taG GAG ACT TAT GTT CAT TCT GGT ATT TTA AAT GAA GAA GTA ACA AGA GTA GCA AAT
S   R   D   S   D   F   D   A   S   I   R   Q   E   T   Y   V   H   S   G   I   L   N   E   E   V   T   R   V   A   N

GGT TTA CCT AAA GAA ACA AAG AGA CCC CAA AAT ACA TTC TTG TAA AAC GTT ATT GGA AAG GAT GAT GAA GAA GAT GAA TAT TAA AAG GAA
G   L   P   K   E   T   K   R   P   Q   N   T   F   L   Q   N   V   I   G   K   D   D   E   E   D   E   Y   Q   K   E

CAA TCT AGA GTT GGT TTA GAT TTT GTT TAT GCC TTT GGT GCT ACC TTG ACT GTT TAA TAA CCT GCA CAC GCT ATT TTA GGT TCT GGT CCT
Q   S   R   V   G   L   D   F   V   Y   A   F   G   A   T   L   T   V   Q   Q   P   A   H   A   I   L   G   S   G   P

CTT TCT TAC CCT TCT AAT AGA CCA GTT TCT GCT ATC GTC TAA ACT AAA AAT AAT GGT AGA CTT GCA GTT ATT GGT TCA TTC GAA ATG TTT
L   S   Y   P   S   N   R   P   V   S   A   I   V   Q   T   K   N   N   G   R   L   A   V   I   G   S   F   E   M   F

ACA GAT GAA TAT TTT GAC AAT GAA GAT AAC TCC AAG ATT TTT GAT TTC TTT ATA AAA TAT TTG CTC ACA AAT GAG TGC GAA TTT GAA TTT
T   D   E   Y   F   D   N   E   D   N   S   K   I   F   D   F   F   I   K   Y   L   L   T   N   E   C   E   F   E   F

AGT CCT AAA GAA CCT GAT GTT GAA TAC TTC AAG GTT CCT GAT ATT GCT GAA TTA GCT GAT AAC CTC AAG AGT TGC TTA CAA GAA AGT GAC
S   P   K   E   P   D   V   E   Y   F   K   V   P   D   I   A   E   L   A   D   N   L   K   S   C   L   Q   E   S   D

CCA TTA CCA TTT GAT AGC AAG CAA TTA TTT ATG ACA GAC TTG TTT AAG TAT GAT GTA GAC TTA GTT CCA GAA GCT GTA AAA TTG TAT GAA
P   L   P   F   D   S   K   Q   L   F   M   T   D   L   F   K   Y   D   V   D   L   V   P   E   A   V   K   L   Y   E

ACT CTT GGA GTA AAG CAC GAT CCT CTT GCT CTT ATA GTT CCT TAA TTC GAA ACT CCA CTC CTT GGA CTT GTT TCA GCT GTT TTC CCA CCT
T   L   G   V   K   H   D   P   L   A   L   I   V   P   Q   F   E   T   P   L   L   G   L   V   S   A   V   F   P   P

ATT TTA AAA GAA TTA GCT CCT CCA AGT TTA GAA TTG TTT GAT TTA GAT GAT GAA TTT GCT TCA GAA AAA GTA AGA CTG GCC TAA CTC ACA
I   L   K   E   L   A   P   P   S   L   E   L   F   D   L   D   D   E   F   A   S   E   K   V   R   L   A   Q   L   T

AAT AAA TGC AAC AAC AAC GAT TTA GAT TAT TAC ATT AAA GAA TCA GGT GAT ATC TTG GGT GTA ACA GAT AAA GTT AAG AAC AAA CAT GAT
N   K   C   N   N   N   D   L   D   Y   Y   I   K   E   S   G   D   I   L   G   V   T   D   K   V   K   N   K   H   D

GCC AAA GCT ATT TTA AGA TAT GTT TTA GAA GAA TTA ATA AAT TTC AAG AAG CTC AAT AAC TGA
A   K   A   I   L   R   Y   V   L   E   E   L   I   N   F   K   K   L   N   N   *
```

Figure S2