

# ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*

Akihisa Matsuyama<sup>1,2,6</sup>, Ritsuko Arai<sup>1,2,6</sup>, Yoko Yashiroda<sup>1,2,6</sup>, Atsuko Shirai<sup>1,2,3</sup>, Ayako Kamata<sup>1,2,3</sup>, Shigeko Sekido<sup>1</sup>, Yumiko Kobayashi<sup>1</sup>, Atsushi Hashimoto<sup>1</sup>, Makiko Hamamoto<sup>1,5</sup>, Yasushi Hiraoka<sup>2,4</sup>, Sueharu Horinouchi<sup>3</sup> & Minoru Yoshida<sup>1,2,3</sup>

**Cloning of the entire set of an organism's protein-coding open reading frames (ORFs), or 'ORFeome', is a means of connecting the genome to downstream 'omics' applications. Here we report a proteome-scale study of the fission yeast *Schizosaccharomyces pombe* based on cloning of the ORFeome. Taking advantage of a recombination-based cloning system, we obtained 4,910 ORFs in a form that is readily usable in various analyses. First, we evaluated ORF prediction in the fission yeast genome project by expressing each ORF tagged at the 3' terminus. Next, we determined the localization of 4,431 proteins, corresponding to ~90% of the fission yeast proteome, by tagging each ORF with the yellow fluorescent protein. Furthermore, using leptomycin B, an inhibitor of the nuclear export protein Crm1, we identified 285 proteins whose localization is regulated by Crm1.**

The ORFeome of an organism is useful for various reverse-proteomics studies, including global analysis of protein localization (the 'localizome'). So far, the most comprehensive localizome established has been for a collection of budding yeast *Saccharomyces cerevisiae* strains (each strain expressing a chromosomally tagged green fluorescent protein (GFP)-fusion protein), covering two-thirds of the proteome<sup>1</sup>. The fission yeast *S. pombe*, which has the smallest genome of all eukaryotes analyzed to date, shares many traits with cells of higher eukaryotes; its proteome will therefore be valuable for functional and comparative studies of eukaryotic cell processes.

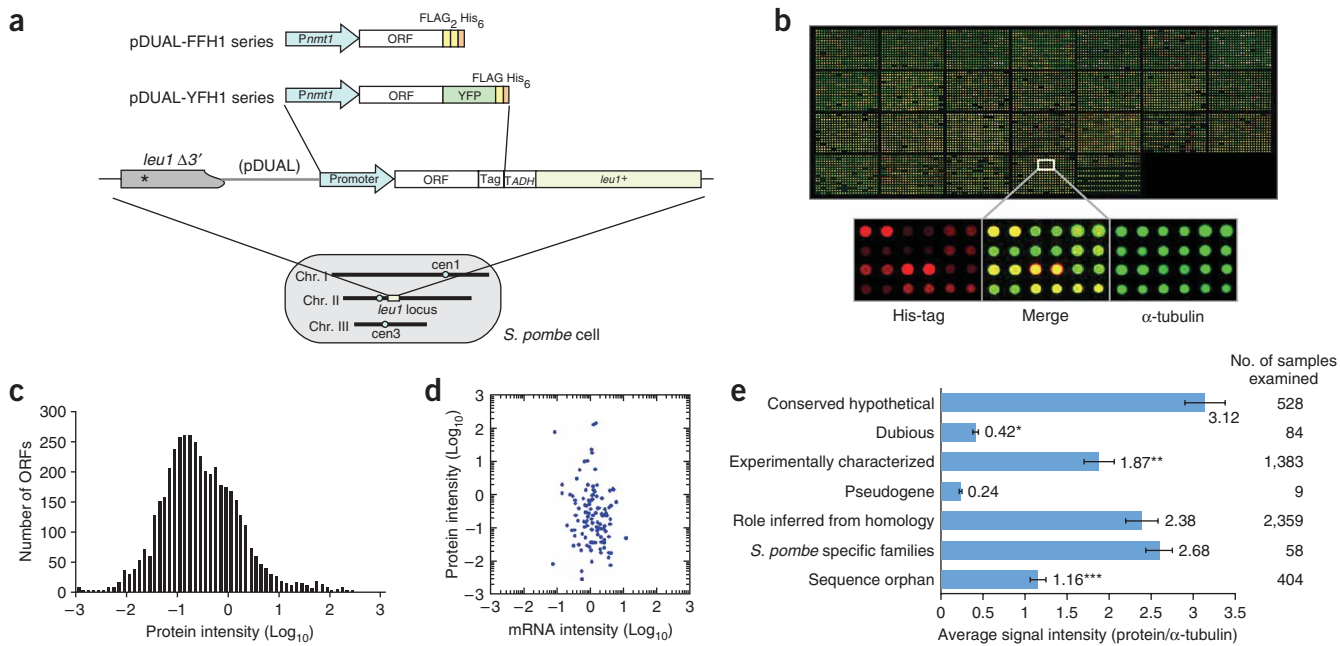
We constructed the fission yeast ORFeome by recombination-based cloning, as this procedure enables systematic ORF cloning and expression<sup>2</sup>. Although full-length complementary DNA (cDNA) libraries have been used for ORF cloning of eukaryotic genes, this method is not well suited to collecting a complete ORFeome because a subset of genes is not expressed or is expressed at low levels under normal culture conditions. However, a high coverage rate can be achieved by cloning the ORFeome from the predicted genes in the

genome, which are equally represented in a genomic DNA preparation. Correct splicing of these cloned genes will likely result in expression of the fission yeast proteins. Using genomic DNA from the wild-type fission yeast strain JY3 as the template, we amplified each predicted ORF by PCR from the initiation to the penultimate codon to allow C-terminal fusion. After eliminating 45 duplicated genes, 40 predicted pseudogenes, 13 transposons and 2 partial genes, we designed primer pairs for a total of 4,948 ORFs. Universal sequences required for the recombination-based Gateway cloning system were tailed (fused) to the gene-specific primers. We cloned the resulting PCR products into the universal donor vector by the BP recombination reaction for generating the so-called 'entry' clones. Once an entry clone is created, we can transfer the cloned ORF to any Gateway-designed expression vector by carrying out the LR recombination reaction.

To confirm that the ORFs were correctly amplified and cloned, we carried out full-length sequencing of the entire entry-clone library. Whenever an error, such as a frameshift or nonsense mutation, was found in the cloned ORF, we recloned the ORF and resequenced the clone. Thus far, we have cloned 4,910 ORFs corresponding to 99.2% of the predicted ORFeome without frameshift or nonsense mutations (**Supplementary Table 1** online). During sequence analyses, we found that 15 clones contained the same frameshifts as in the first trial (**Supplementary Table 2** online). These frameshifts might be due to sequencing errors in the genome project. Upon cloning these 15 ORFs from the genome of the strain used in the genome project (the 972 *h*<sup>-</sup> strain), we demonstrated by sequence analysis that 14 out of the 15 clones contained a sequence that was identical to the JY3 sequence. Most of these corrected sequences resulted in elongated ORFs. For example, correction of the closely aligned ORFs SPBC14C8.08c and SPBC14C8.09c led to one extended ORF. However, the sequence of one of the fifteen clones (SPAC1782.09c) from 972 *h*<sup>-</sup> was identical to the one reported in the genome project, suggesting the presence of a single nucleotide polymorphism (SNP) between JY3 and 972 *h*<sup>-</sup>.

<sup>1</sup>Chemical Genetics Laboratory, RIKEN, Wako, Saitama 351-0198, Japan. <sup>2</sup>CREST Research Project, JST, Kawaguchi, Saitama 332-0012, Japan. <sup>3</sup>Department of Biotechnology, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan. <sup>4</sup>Kansai Advanced Research Center, National Institute of Information and Communications Technology, Kobe, Hyogo 651-2492, Japan. <sup>5</sup>Present address: School of Agriculture, Meiji University, Tama-ku, Kanagawa 214-8571, Japan. <sup>6</sup>These authors contributed equally to this work. Correspondence should be addressed to M.Y. (yoshidam@riken.jp).

Received 6 February; accepted 4 May; published online 25 June; corrected after print 25 July 2006; doi:10.1038/nbt1222



**Figure 1** Cloning and expression of the fission yeast ORFeome. **(a)** Schematic representation of integrants constructed for expression and localization studies of the fission yeast proteome. **(b)** Fission yeast reverse array. Tagged proteins were detected with an anti-His-tag antibody (red), and the internal control  $\alpha$ -tubulin was simultaneously quantified with an anti- $\alpha$ -tubulin antibody (green). Strains with integrated empty vector pDUAL-FFH1 were used as a negative control. The total cell lysate of the control strain was spotted at least every 192 spots in this array. **(c)** Expression levels of all proteins measured by reverse arrays. The histogram shows the number of ORFs binned by protein intensity as described in **b**. Strains with integrated empty vector pDUAL-FFH1 were also used as a negative control. **(d)** Distribution of the protein and mRNA expression levels. Abundance of 120 mRNAs determined by quantitative RT-PCR analysis. **(e)** Protein expression levels categorized by their annotation status. The average expression levels of tagged proteins in each indicated set of ORFs measured by reverse arrays are shown. Error bars represent standard errors. \*,  $P < 1 \times 10^{-8}$ ; \*\*,  $P = 7.8 \times 10^{-6}$ ; \*\*\*,  $P = 2.6 \times 10^{-7}$ , by the Mann-Whitney test.

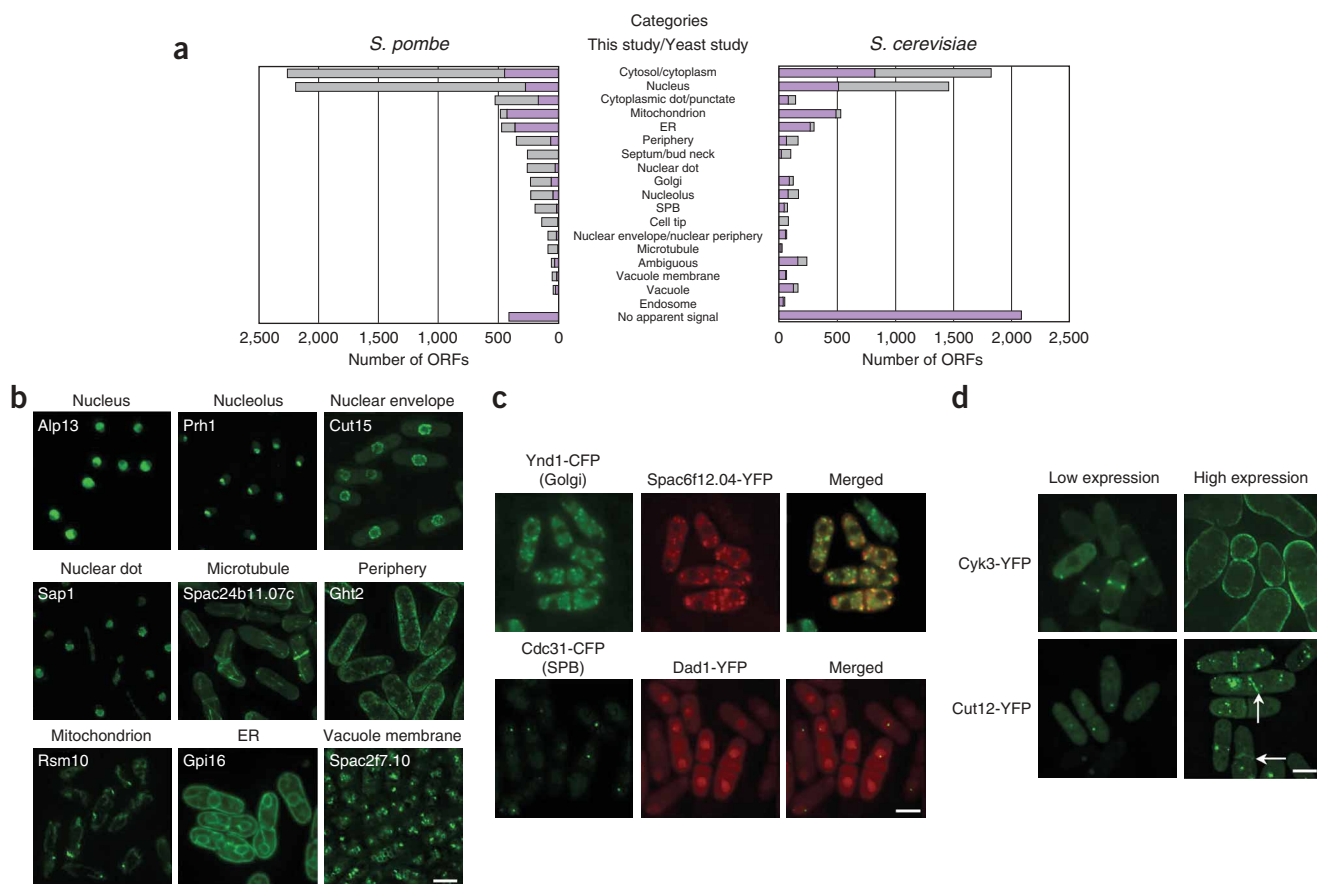
Although we detected many mismatches in ORFs, additional analyses are needed to determine whether these mutations are caused by sequencing and/or PCR errors, or the presence of SNPs between both fission yeast strains.

Almost half of the genes in fission yeast are predicted to have one or more introns, although many of these introns have not been experimentally confirmed<sup>3</sup>. As reported in the construction of the *Caenorhabditis elegans* nematode ORFeome<sup>2</sup>, incorrect prediction of exon-intron organization might hamper downstream 'omics' applications. To estimate the intron prediction accuracy, we constructed a series of plasmids containing an ORF with two 3' FLAG epitopes and one hexahistidine (His<sub>6</sub>) tag. If intron prediction is incorrect, a premature termination codon or a frameshift in the 3'-terminal tags will likely occur and give no signal in western blotting analysis. Computational analysis indicated that in the absence of proper splicing of all the predicted introns, only 14 of the 2,286 intron-containing genes (0.6%) would produce in-frame tagged proteins. To express each 3'-tagged ORF, we cloned each ORF into the *S. pombe* expression vector pDUAL-FFH1c, a plasmid that enables both multi-copy maintenance and chromosomal integration of the transgene<sup>4</sup>. We expressed the cloned ORFs under the control of the thiamine-regulatable *nmt1* promoter<sup>5</sup>, because this promoter allows controlled expression of the cloned ORFs and reduces expression of toxic ORFs. We integrated the 3'-tagged ORF plasmids individually into the *leu1* locus of the *leu1-32* strain AM2 (**Fig. 1a**). After induction of tagged ORF expression in minimal medium (MM), we prepared total cell lysates for each strain and spotted these lysates onto nitrocellulose membranes to make genome-wide protein macroarrays,

so-called 'reverse arrays'<sup>6</sup>. We detected protein expression from the cloned ORFs on the reverse arrays by immunoblotting with an anti-His-tag antibody. The expression level of each ORFeome protein was determined by normalization against an internal control,  $\alpha$ -tubulin (**Fig. 1b**).

Compared with the intron-less genes, the intron-containing genes showed no significant difference in protein expression (**Supplementary Fig. 1a** online). Lack of significant difference between both classes indirectly validates intron prediction accuracy. Although all ORFs were expressed under the same promoter and terminator (**Fig. 1c**), we found that the protein amounts varied widely. This variation might be due to the rate of translation or protein degradation, because the mRNA levels of 120 randomly selected ORFs were not as dispersed ( $P < 1 \times 10^{-8}$ , using F test) and never correlated with the protein levels (Spearman rank correlation coefficient  $r_s = 0.0729$ ) (**Fig. 1d** and **Supplementary Fig. 1b** online). In addition, ORF size did not correlate with protein level (**Supplementary Fig. 1c** online). Statistical analysis revealed, however, that proteins expressed at low levels were evolutionarily less conserved (e.g., orphan proteins, which have no recognizable homolog in distantly related species, or dubious proteins, which are the nonconserved ORFs that might not actually code for proteins<sup>3</sup>), whereas proteins expressed at high levels were evolutionarily more conserved (including hypothetically conserved proteins) (**Fig. 1e**), suggesting that the levels of the less conserved proteins are kept low at the post-transcriptional level by an unidentified mechanism.

Comprehensive analysis of protein subcellular localization will provide deeper insight into systems-level eukaryotic cell biology.



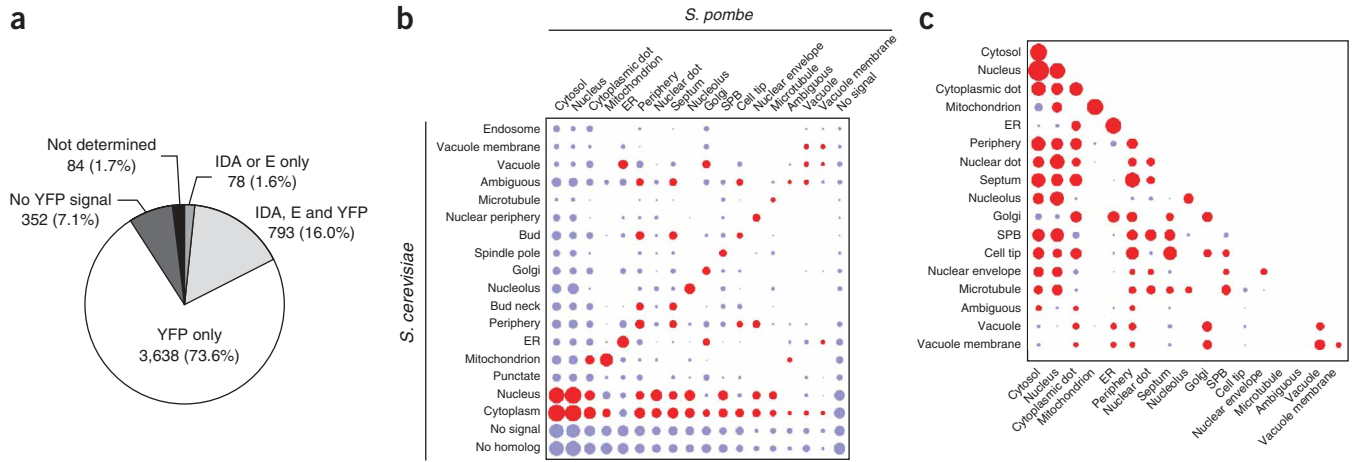
**Figure 2** Microscopic analysis of *S. pombe* strains expressing ORF-YFP fusions. **(a)** Distribution of YFP fusions in the localization categories and comparison with the budding yeast study—a data set of the budding yeast proteins available from the yeast GFP fusion localization database<sup>1</sup> (<http://yeastgfp.ucsf.edu/>). The purple and gray regions in each bar represent the numbers of ORFs classified into a single category and multiple categories, respectively. Because certain proteins are classified into multiple localization categories, the total ORF count exceeds the number of genomic ORFs. SPB, spindle pole body; ER, endoplasmic reticulum. **(b)** Representative YFP images. SPB, spindle pole body; ER, endoplasmic reticulum. **(c)** Representative colocalization study using markers of Golgi (Ynd1-CFP) and SPB (Cdc31-CFP). **(d)** Examples of morphological abnormalities induced by overexpression. Cyk3-YFP localized at the site of septum formation; its overexpression induced peripheral localization and aberrant cell morphology. Overexpression altered localization of Cut12-YFP from SPBs to the abnormal septum (arrows). Bars, 5  $\mu$ m.

Cells are organized into complex networks of membranes and compartments that are specialized for various biological functions. To systematically determine protein localization in living fission yeast cells, we fused each ORF at the 3' end to the yellow fluorescent protein (YFP)<sup>7</sup> by using the pDUAL-YFH1c vector (Fig. 1a). We integrated each plasmid into the *S. pombe* chromosome and expressed the ORF-YFP fusions under the control of the *nmt1* promoter<sup>5</sup>. Forced expression of proteins that are normally not abundant in the cell from the *nmt1* promoter might result in detectable expression levels. Indeed, the number of fission yeast proteins with no fluorescence signal (411 proteins, 8.3% of total ORFs) was much smaller than the number obtained in the *S. cerevisiae* localization study (2,078 proteins, 33.3% of total ORFs)<sup>1</sup> (Fig. 2a). The higher protein expression levels achieved by the *nmt1* promoter, however, could also raise a risk of incorrect localization (see below).

Among the 4,948 ORFs, 4,431 (89.6%) showed positive signals (Supplementary Table 1 online); representative images are shown in Figure 2b and images of all the ORF-YFP fusions are available on our website (<http://cgl.riken.go.jp>). To depict the localization of proteins with nuclear and cytoplasmic dots more precisely, we used Cdc31-CFP and Ynd1-CFP as a spindle pole body (SPB)<sup>8</sup> and a Golgi<sup>9</sup> marker,

respectively (Fig. 2c). Colocalization of some ORF-YFP fusions with these two markers indicated that some of the nuclear and cytoplasmic ORF-YFP dots correspond to the SPB and the Golgi apparatus, respectively. When strains showed morphological alterations or suspect localizations of the YFP-fusion because of the strong *nmt1* promoter, we reanalyzed these strains under low expression conditions (Fig. 2d). Ultimately, we determined the subcellular localization of ~90% of the fission yeast proteome, including a total of 3,638 previously uncharacterized proteins (>70% of the proteome) (Fig. 3a).

After expressing certain fusions under reduced expression conditions and carrying out colocalization studies, we classified all the gene products with a positive YFP signal into 17 distinct localization categories (Fig. 2a). Similar to those of *S. cerevisiae*<sup>1</sup>, the majority of proteins localized in the cytosol and the nucleus (Fig. 2a). *S. pombe* has 4,280 proteins that are homologous to those present in *S. cerevisiae*. The subcellular distribution of *S. pombe* proteins was very similar to that of *S. cerevisiae* homologs (Fig. 3b). Furthermore, we evaluated the *S. pombe* localizome data using the reported or predicted data available from the Gene Ontology (GO) database<sup>10</sup> (GO\_cellular component) (Supplementary Fig. 2 online). Although



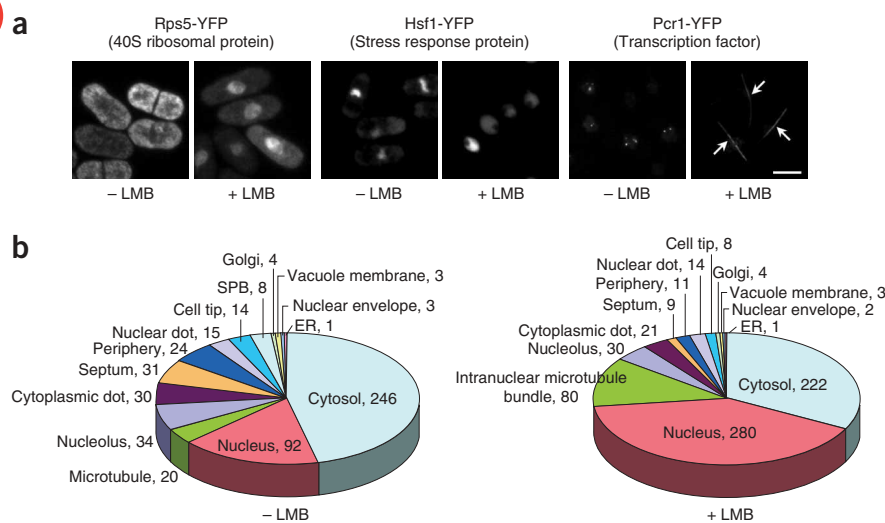
**Figure 3** Comparative analysis of localization pattern of *S. pombe* proteins. **(a)** Rate of newly determined localization. IDA, a GO term assigned to 'inferred by direct assay' by the *S. pombe* model organism database GeneDB\_Spombe using Gene Ontology (GO) descriptors. E, 'experimentally determined' assigned by the PombePD™ database<sup>30</sup>. **(b)** Localizome comparison between fission and budding yeast. The subcellular localization of fission yeast proteins is compared with that of the budding yeast homologs using the same protein data sets<sup>1</sup> as in **Fig. 2a**. The radii of the circles indicate the number (log) of the proteins in each category. Red circles indicate >99% confidence in the one-sided binominal test. Proteins categorized in 'no signal' or 'no homolog' are eliminated in this test. SPB, spindle pole body; ER, endoplasmic reticulum. **(c)** Protein distribution patterns in the cellular compartments of fission yeast. Each circle shows the number (log) of proteins localized in both indicated organelles. The circles on the diagonal line denote the number (log) of proteins localized in only one organelle. Calculation as described in **c**. SPB, spindle pole body; ER, endoplasmic reticulum.

GO classification has its limitations, mainly because of its dependency on homologs from other organisms, our data coincided well with the GO database. Among the 4,431 products with positive YFP signals, 2,495 proteins showed localization in multiple compartments (**Fig. 2a**). The overlapping locations of each protein were analyzed by the one-sided binominal test (**Fig. 3c**; same analysis for *S. cerevisiae*, **Supplementary Fig. 3** online). In both yeasts, many proteins simultaneously localized at 'microtubule' and 'SPB/spindle pole.' We observed a similar relationship between 'periphery' and 'septum/bud neck' or 'vacuole.' In contrast, 'mitochondrion' showed a weak relationship with other subcellular compartments (**Figs. 2a, 3c**

and **Supplementary Fig. 3** online). Remarkably, we found a relation between 'SPB' and 'septum,' indicating the probable interaction between the SPB/centrosome and the cell division plane, as predicted in *S. pombe* and animal cells<sup>11</sup>. In addition, our colocalization study distinguished 'SPB' from 'non-SPB' nuclear dots. The proteins that localized at the 'non-SPB' nuclear dots were predominantly sorted into GO biological process categories related to DNA-dependent functions (**Supplementary Table 3** online). This suggests the existence of various functional domains in the *S. pombe* nucleus, similar to the mammalian nucleus<sup>12</sup>. Whereas certain cytoplasmic dots localized to the Golgi apparatus, other cytoplasmic dots were often simultaneously

localized at the periphery and at the endoplasmic reticulum (**Fig. 3c**). These dots may include membrane-associated structures, such as intermediates in the membrane traffic pathway. Thus, localization pattern analysis indicates functional relationships between subcellular compartments.

Crm1 was the first identified, evolutionarily conserved exportin that mediates nuclear export of proteins containing the short nuclear export signal (NES) sequence<sup>13</sup>. Although the sequence requirements for the NES are only loosely defined, the consensus sequence for a functional NES is proposed to be L-X<sub>2-3</sub>-L/I/F/V/M-X<sub>2-3</sub>-L-X-L/I, a short leucine-rich motif<sup>14</sup>. However, a large population of the proteome contains this sequence motif (see below). In most cases, it has not been examined whether localization is indeed regulated by the motif and Crm1. In addition, several proteins that do not possess any canonical NES sequence are exported by Crm1<sup>15,16</sup>. To examine which proteins in our ORF-YFP fusion library are transported



**Figure 4** Regulation of protein localization by Crm1-mediated nuclear export. **(a)** Images of localization-altered proteins by LMB treatment. Examples of nuclear accumulation (left and middle panels) and relocalization to a microtubule structure (right panel, arrows) of YFP-fusions after LMB treatment are shown. Bars, 5  $\mu$ m. **(b)** Subcellular localization of 285 proteins before (-) and after (+) LMB treatment. SPB, spindle pole body; ER, endoplasmic reticulum.

in a Crm1-dependent manner, we treated our yeast strains with the specific Crm1 inhibitor, leptomycin B (LMB)<sup>17,18</sup>. Whereas *S. pombe* and human cells are LMB sensitive<sup>19</sup>, *S. cerevisiae* is LMB resistant, because the budding yeast variant lacks a cysteine residue in the Crm1 central conserved region.

We identified 285 proteins whose localization was altered upon Crm1 inhibition (**Supplementary Table 4** online). After addition of LMB, most of these proteins accumulated in the nucleus/nucleolus (**Fig. 4a**). Among the five known Crm1-dependent proteins<sup>15,20–23</sup>, Pap1 and Wis1 relocated upon LMB treatment, whereas Ste11, Cut3 and Mid1 did not, because these three proteins were already accumulated in the nucleus in the absence of LMB. A computational search revealed that ~45% of the localization-altered proteins contain the canonical NES sequence (L-X<sub>2-3</sub>-L/I/F/V/M-X<sub>2-3</sub>-L-X-L/I), although it is unclear whether this sequence functions as a NES. The remaining proteins may either possess partially matched or atypical NES sequence motifs, or be transported by an adapter that binds Crm1. Because proteins bearing the canonical NES sequence were not enriched in this screen (analyzed by the Fisher's exact probability test,  $P = 0.069$ ), NES sequence variation might be greater than expected. Almost half of the proteins were originally localized in the cytosol (**Fig. 4b**). A significant portion of the proteins that mainly localized in the nucleus/nucleolus and partially localized in the cytosol accumulated exclusively in the nucleus upon LMB treatment. Surprisingly, a considerable number of proteins that were localized to the cell tip, septum and microtubules relocated after LMB treatment, suggesting that proper localization of not only cytosolic proteins but also those at these specialized domains might require nuclear export (and/or import). Based on grouping analysis of the 285 proteins using the GO\_biological process, we suggest that Crm1 is involved in various cellular processes (**Supplementary Table 5** online). We also found that 80 proteins relocated to an intranuclear microtubule bundle upon LMB treatment (**Supplementary Table 4** online and **Fig. 4a**). Therefore, Crm1-dependent dynamic interchange of proteins between the nucleus and the microtubules may regulate microtubule organization during the cell cycle in fission yeast. Further analysis on individual proteins will reveal the involvement and importance of Crm1 in various biological processes.

We have described the construction of the fission yeast ORFeome and its application to determining the localizome. Using this reverse-proteomic approach, we successfully collected localizome data that covered nearly 90% of the *S. pombe* proteome, including 3,638 gene products whose localization was previously unknown. Use of the *nmt1* promoter enabled detection of genes that are normally tightly regulated or undetectable when expressed by their endogenous promoters. However, our study lacks temporal expression data. Although the C-terminal YFP fusion partner might impair correct localization of some proteins, such as  $\alpha$ -tubulin, and forced expression by the strong *nmt1* promoter might cause incorrect localization, the data set presented here still provides useful information for the identification of novel components or regulators of subcellular architecture. In addition, our comparison of fission and budding yeast demonstrates that many homologous proteins share common localization patterns despite the evolutionary disparity between these organisms<sup>24</sup>. For example, although both yeasts divide differently, many fission yeast homologs of the budding yeast proteins that localize in the bud neck are likely to localize at the septum. This localization result suggests an evolutionary link between the bud in budding yeast and the septum in fission yeast. Finally, we show in our LMB screen that Crm1 regulates the export of more than 280 gene products, some of which may be responsible for the LMB-induced phenotype. Because the molecular

system that directs subcellular localization of disease-related proteins is one of the more promising drug targets, the localizome might allow the design of screening systems for the discovery of drugs or drug targets.

In this study all the constructed pDUAL-based<sup>4</sup> plasmids contain the fission yeast markers *leu1* and *ura4* in one plasmid. Because most existing fission yeast strains usually contain either a marker of *leu1-32* or *ura4-D18*, these strains can be used as hosts for the pDUAL-based plasmids. For example, systematic introduction of the ORF-YFP plasmids into the *Δimp1* mutant strain that contains a disruption in one of the importin- $\alpha$  genes<sup>25</sup> might enable identification of proteins that are transported in an Imp1-dependent manner. Thus, the fission yeast ORFeome should enable more complex reverse-proteomic studies.

## METHODS

**Cloning of the *S. pombe* ORFeome using the Gateway system.** Cloning of each gene predicted in fission yeast was carried out as follows. Sequencing information for each gene predicted in fission yeast was retrieved from the GenBank database. Since *S. pombe* has many duplicated genes that have exactly the same sequences or those encoding proteins of which amino acid sequences are identical, only one of them was randomly chosen for cloning. Besides the duplicated genes, *S. pombe* has 14 transposons that encode a reverse transcriptase, including pseudo-type retrotransposable elements, but only one of them (SPBC9B6.02c) was chosen for cloning. In addition, there are some ORFs having the same 5'- and 3'- sequences, although the internal sequences are distinct, which is typical in retrotransposons containing long terminal repeats wtf<sup>3</sup>. These ORFs were simultaneously amplified by the same set of primers and distinguished by sequencing after cloning. Sequence data were automatically processed by the software originally developed for this study using the Script Editor (Apple). A set of primers to amplify each gene was also designed automatically with this software. Essentially, twenty bases starting from an initiation codon were attached to the universal sequence AAAAAGC AGGCTCTCAT containing a part of *attB1* and a spacer sequence to generate a forward primer. Similarly, a complementary sequence of twenty bases just upstream of the stop codon was attached to the other universal sequence AGAAAGCTGGGTA containing a part of *attB2*, resulting in the generation of a reverse primer. Using these primers and DNA extracted from JY3 (*h*<sup>90</sup> wild-type), each gene was amplified by two-step PCR according to the manufacturer's instructions in a 96-well format. *Pfu* DNA Polymerase (TaKaRa BIO) or Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) was used for amplification. The resultant PCR products were checked by agarose gel electrophoresis, mixed with the Donor vector pDONR201 (Invitrogen), and subjected to the recombination reaction named the 'BP' reaction followed by transformation of the *E. coli* strain DH5 $\alpha$  to generate *S. pombe* entry clones. When an extra band(s) was observed in the electrophoretic analysis, the correct band was gel-extracted, purified and subjected to the BP reaction. Plasmid DNAs were prepared from a 1-ml culture of one isolated colony for each ORF by the alkaline method in a 96-well format. For extremely long ORFs such as SPCC738.08 and SPAC23G3.02c, ORFs were amplified as two divided fragments at first and they were subsequently connected by the standard method using restriction endonucleases and a DNA ligase. All inserts were systematically subjected to full-length sequence analysis (minimum PHRED score of 20) using the RISA-384 sequencer (SHIMADZU BIOTECH). We detected 4,540 mismatches in 7.2-Mbp insert sequences, corresponding to one mismatch at every 1,590 bp. These may be a sum of errors in PCR or sequence analyses, and due to substantial differences in strains used in the genome project and our study. If the mutations were shown to cause a nonsense mutation or a frameshift, we checked another colony for those ORFs, or reamplified and cloned them. Clones containing deletions or insertions were also subjected to the reamplification. Any mutations that are supposed to be missense changes, and those that are located within intron and are not supposed to affect splicing were accepted. Information on these mutations is disclosed on our website. Thus, it should be noted that further analyses for each mutation-containing ORF are required to determine whether or not each mutation affects

the function of the ORF. All the strains and clones described here will be available from RIKEN BioResource Center ([http://www.brc.riken.jp/lab/dna/en/yoshidayeast\\_en.html](http://www.brc.riken.jp/lab/dna/en/yoshidayeast_en.html)).

**Vector construction.** pDUAL<sup>4</sup>, a multicopy vector capable of chromosomal integration was modified to allow expression of a YFP- or CFP-fused gene at the 3' terminus of an ORF. The FLAG epitope tag and the His<sub>6</sub> tag were also contained in these vectors. The coding sequence for each fluorescence protein, the FLAG tag, and the His<sub>6</sub> tag are positioned in tandem between the polylinker sequences in the vector. The CFP and YFP moieties were amplified by PCR using the following primer sets, BamHI-EGFP-C-F (5'-AGGCCTGGATCCGG ATATCATGGTGAGCAAGGGCGAGGAG-3') and NcoI-EGFP-C-R (5'-GTATGCCATGGGCTTGACAGCTCGTCCATGC-3'). The PCR products were digested by BamHI and NcoI, and then inserted into the same sites of the pDUAL-FFH1 vector, replacing one copy of the FLAG epitope. The *ccdB*-*cm<sup>r</sup>* cassette was inserted upstream of each fluorescent protein-coding sequence, resulting in the generation of a series of vectors named pDUAL-YFH1c or pDUAL-CFH1c, respectively, ready to use in the Gateway system.

**Construction of *S. pombe* strains expressing tagged proteins.** General methods to handle fission yeast cells were as described<sup>26</sup>. To create ORF-FLAG<sub>2</sub>-His<sub>6</sub>, ORF-YFP-FLAG-His<sub>6</sub> or ORF-CFP-FLAG-His<sub>6</sub> expression plasmids, universal *S. pombe* entry clones were mixed with pDUAL-FFH1c, pDUAL-YFH1c or pDUAL-CFH1c vectors, respectively, and subjected to the LR recombination reaction. A correct insertion of each ORF was confirmed from a fragment size by *EcoRV* digestion. These expression plasmids were digested with *NotI* (*ApaI* or *SacII*, if the *NotI* site exists within an insert region) and the resultant fragments were individually integrated into the *leu1* locus of the strain AM2 (*h<sup>90</sup> leu1-32*) by a slightly modified version of a high efficiency transformation protocol<sup>27</sup>. *Leu<sup>+</sup>* transformants were selected on SD solid medium.

**Measurement of the protein levels using reverse arrays.** Total cell lysates were prepared for reverse arrays by a slightly modified protocol of the post-alkaline method<sup>28</sup>. Cultures of tagged strains grown on MM solid medium for 22 h at 30 °C were harvested in 25 µl of MM liquid medium within 96-well plates. Forty microliters of 0.4 N NaOH were added into the suspension. Cells were incubated for 5 min at room temperature, pelleted, resuspended in 30 µl of SDS gel-loading buffer and boiled for 5 min. The resultant supernatants were directly spotted on a nitrocellulose membrane (BIO-RAD) using a 96-pin replicator (V&P SCIENTIFIC). The expression levels of each protein were calculated by immunoblotting using an anti-His-tag antibody (MBL) compared to those of  $\alpha$ -tubulin detected by an anti- $\alpha$ -tubulin antibody (Oxford) as an internal control. Each signal was analyzed using the Odyssey Imaging System (LI-COR).

**The quantitative RT-PCR analysis.** Total cellular RNA was prepared from each sample by disrupting the cells with glass beads and following a standard extraction protocol<sup>29</sup>. The quantitative RT-PCR was carried out using the SuperScript III Platinum Two-Step qRT-PCR Kit with SYBR Green (Invitrogen) according to the manufacturer's instructions. The expression level of each ORF was determined as the ratio against that of the  $\alpha$ -tubulin gene *tub1*.

**Fluorescence microscopy of *S. pombe* expressing YFP-tagged proteins.** Two to four individual transformants for each ORF were initially grown on YE solid medium (0.5% yeast extract, 2% glucose, 5 µg/ml adenine) at 26.5 °C for 2–3 d. To allow expression of genes driven by the *nmt1* promoter<sup>5</sup>, they were subsequently streaked on MM solid medium and grown overnight at 26.5 °C. For the colocalization study, strains expressing Cdc31-CFP or Ynd1-CFP under the control of the *tub1* promoter were constructed, and selected constructs for the ORF-YFP expression were individually integrated into these strains. For observation under conditions for the reduced expression, cells were streaked on a YE plate naturally containing 2.6 µg/ml of thiamine overnight at 26.5 °C. Living cells expressing YFP-tagged proteins grown on the MM or YE plates were suspended in the MM or YE liquid medium. Each suspension was mounted on a slide glass and covered. To collect YFP and CFP images, we mainly used a DeltaVision system (Applied Precision) with an Olympus IX70 fluorescence microscope equipped with an UPlan Apo 100 × lens.

**Chemical genetic study using leptomycin B.** The transformants of the ORF-YFP library were incubated in MM liquid medium using a 96-well plate overnight with agitation at 26 °C. These cultures were split in two, to which LMB (final concentration of 100 ng/ml) and ethyl alcohol (solvent for LMB), respectively, were added, and then incubated for 3–6 h at 26 °C before observation.

**Data source.** DNA and protein information were retrieved from the *S. pombe* GeneDB database (The Sanger Center), the GenBank database and the PombePD database<sup>30</sup>. Functional grouping of ORFs was based on the Gene Ontology (GO) data as of July 1, 2005.

**Statistical analyses.** SYSTAT 10.2 (Systat Software) was mainly used for statistical analyses. We estimated *P* values using the Mann-Whitney test for the relationship between the expression levels and the annotation status of proteins. The results were further analyzed by the Bonferroni correction. The relationship between the expression levels of proteins and the presence of introns, and between the alteration of the subcellular localization and the presence of NES were analyzed by the Fisher's exact probability test. The distribution of the protein and mRNA expression levels was analyzed by the F test.

**URLs.** Information about the ORFeome and the localizome is available at <http://cgl.riken.go.jp>. Users are free to browse and search basic gene information. Sequence information and localizome data regarding the entry clones constructed in this study is available in **Supplementary Table 6** online and on the RIKEN website.

*Note: Supplementary information is available on the Nature Biotechnology website.*

#### ACKNOWLEDGMENTS

We thank Megumi Takase, Kunio Watanabe and Ryuji Yoshinari for technical assistance and the members of Chemical Genetics Laboratory, RIKEN, for discussions. We also thank Daniel McCollum for critical reading of the manuscript, and Jürg Bähler for useful comments. This work was supported in part by the CREST Research Project, Japan Science and Technology Agency, The Strategic Research Programs for R&D, RIKEN and a Grant-in-Aid for Scientific Research on Priority Area 'Cancer' from The Ministry of Education, Culture, Sports, Science and Technology, Japan.

#### AUTHOR CONTRIBUTIONS

The order of listing of the authors A.M., R.A. and Y.Y. in no way reflects their relative contribution to this work. M.Y. is responsible for the project planning and experimental design with support from Y.H. and S.H.; A.M., Y.Y. and A.H. carried out cloning and sequencing; R.A., Y.Y. and M.H. collected the localizome data; R.A. and A.K. carried out the chemical genetic screen; S.S. and Y.K. generated plasmids and transformants; and A.M. and A.S. carried out reverse array and statistic analyses.

#### COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/naturebiotechnology/>  
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Huh, W.K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Reboul, J. *et al.* *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* **34**, 35–41 (2003).
- Wood, V. *et al.* The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).
- Matsuyama, A. *et al.* pDUAL, a multipurpose, multicopy vector capable of chromosomal integration in fission yeast. *Yeast* **21**, 1289–1305 (2004).
- Maudrell, K. *nmt1* of fission yeast. A highly transcribed gene completely repressed by thiamine. *J. Biol. Chem.* **265**, 10857–10864 (1990).
- Poetz, O. *et al.* Protein microarrays: catching the proteome. *Mech. Ageing Dev.* **126**, 161–170 (2005).
- Tsien, R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
- Paoletti, A. *et al.* Fission yeast *cdc31p* is a component of the half-bridge and controls SPB duplication. *Mol. Biol. Cell* **14**, 2793–2808 (2003).
- D'Alessio, C., Trombetta, E.S. & Parodi, A.J. Nucleoside diphosphatase and glycosyltransferase activities can localize to different subcellular compartments in *Schizosaccharomyces pombe*. *J. Biol. Chem.* **278**, 22379–22387 (2003).

10. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
11. Burgess, D.R. & Chang, F. Site selection for the cleavage furrow at cytokinesis. *Trends Cell Biol.* **15**, 156–162 (2005).
12. Zimber, A., Nguyen, Q.D. & Gespach, C. Nuclear bodies and compartments: functional roles and cellular signalling in health and disease. *Cell. Signal.* **16**, 1085–1104 (2004).
13. Ullman, K.S., Powers, M.A. & Forbes, D.J. Nuclear export receptors: from importin to exportin. *Cell* **90**, 967–970 (1997).
14. Bogerd, H.P., Fridell, R.A., Benson, R.E., Hua, J. & Cullen, B.R. Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel *in vivo* randomization-selection assay. *Mol. Cell. Biol.* **16**, 4207–4214 (1996).
15. Kudo, N., Taoka, H., Toda, T., Yoshida, M. & Horinouchi, S. A novel nuclear export signal sensitive to oxidative stress in the fission yeast transcription factor Pap1. *J. Biol. Chem.* **274**, 15151–15158 (1999).
16. Hoshino, H. *et al.* Oxidative stress abolishes leptomycin B-sensitive nuclear export of transcription repressor Bach2 that counteracts activation of Maf recognition element. *J. Biol. Chem.* **275**, 15370–15376 (2000).
17. Kudo, N. *et al.* Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1. *Exp. Cell Res.* **242**, 540–547 (1998).
18. Wolff, B., Sanglier, J.J. & Wang, Y. Leptomycin B is an inhibitor of nuclear export: inhibition of nucleo-cytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA. *Chem. Biol.* **4**, 139–147 (1997).
19. Kudo, N. *et al.* Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc. Natl. Acad. Sci. USA* **96**, 9112–9117 (1999).
20. Nguyen, A.N., Ikner, A.D., Shiozaki, M., Warren, S.M. & Shiozaki, K. Cytoplasmic localization of Wis1 MAPKK by nuclear export signal is important for nuclear targeting of Spc1/Sty1 MAPK in fission yeast. *Mol. Biol. Cell* **13**, 2651–2663 (2002).
21. Qin, J., Kang, W., Leung, B. & McLeod, M. Ste11p, a high-mobility-group box DNA-binding protein, undergoes pheromone- and nutrient-regulated nuclear-cytoplasmic shuttling. *Mol. Cell. Biol.* **23**, 3253–3264 (2003).
22. Sutani, T. *et al.* Fission yeast condensin complex: essential roles of non-SMC subunits for condensation and Cdc2 phosphorylation of Cut3/SMC4. *Genes Dev.* **13**, 2271–2283 (1999).
23. Paoletti, A. & Chang, F. Analysis of mid1p, a protein required for placement of the cell division site, reveals a link between the nucleus and the cell surface in fission yeast. *Mol. Biol. Cell* **11**, 2757–2773 (2000).
24. Sipiczki, M. Where does fission yeast sit on the tree of life? *Genome Biol.* **1** REVIEWS1011 (2000).
25. Umeda, M., Izaddoost, S., Cushman, I., Moore, M.S. & Sazer, S. The fission yeast *Schizosaccharomyces pombe* has two importin- $\alpha$ , Imp1p and Cut15p, which have common and unique functions in nucleocytoplasmic transport and cell cycle progression. *Genetics* **171**, 7–21 (2005).
26. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
27. Matsuyama, A., Yabana, N., Watanabe, Y. & Yamamoto, M. *Schizosaccharomyces pombe* Ste7p is required for both promotion and withholding of the entry to meiosis. *Genetics* **155**, 539–549 (2000).
28. Kushnirov, V.V. Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857–860 (2000).
29. Elder, R.T., Loh, E.Y. & Davis, R.W. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. *Proc. Natl. Acad. Sci. USA* **80**, 2432–2436 (1983).
30. Csank, C. *et al.* Three yeast proteome databases: YPD, PombePD, and CalPD (MycPathPD). *Methods Enzymol.* **350**, 347–373 (2002).

## Corrigendum: ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*

Akihisa Matsuyama, Ritsuko Arai, Yoko Yashiroda, Atsuko Shirai, Ayako Kamata, Shigeko Sekido, Yumiko Kobayashi, Atsushi Hashimoto, Makiko Hamamoto, Yasushi Hiraoka, Sueharu Horinouchi & Minoru Yoshida

*Nat. Biotechnol.* 24, 841–847 (2006); published online 25 June 2006; corrected after print 25 July 2006

In the print version of this article and the version initially published online, the last sentence of the URLs section in Methods, p.846, reads, “Sequence information regarding the entry clones constructed in this study is also available on our website.” The sentence should read, “Sequence information and localizome data regarding the entry clones constructed in this study are available in **Supplementary Table 6** online and on the RIKEN website.” The sequence information and localizome data for the entry clones has been posted on the *Nature Biotechnology* website. The error have been corrected in the HTML and PDF versions of the article.

## Corrigendum: Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR

Cynthia H Collins, Jared R Leadbetter & Frances H Arnold

*Nat. Biotechnol.* 24, 708–712 (2006); published online 21 May 2006; corrected after print 25 July 2006

In the Methods section, p.711, “Library construction and selection,” line 11 of the print version of this article and the version initially published online, the plates used for rounds of ON selection were incorrectly said to contain 3OC6HSL. The text should read “...plated onto LB agar plates containing 50 µg/ml kanamycin, 150 µg/ml chloramphenicol and 100 nM C10HSL.” The error has been corrected in the HTML and PDF versions of the article.

## Corrigendum: Design of a genome-wide siRNA library using an artificial neural network

Dieter Huesken, Joerg Lange, Craig Mickanin, Jan Weiler, Fred Asselbergs, Justin Warner, Brian Meloon, Sharon Engel, Avi Rosenberg, Dalia Cohen, Mark Labow, Mischa Reinhardt, François Natt & Jonathan Hall

*Nat. Biotechnol.* 23, 995–1001 (2005), published online 17 July 2005; corrected after print 25 July 2006

In the Methods section, p. 1,000, col. 2, paragraph 2, the text beginning: “The whole plate was discarded if for any time point...” and ending, “The final data set contained 2,431 sequences.” inaccurately described the procedure. The text should read: “Data points for 2,675 siRNAs were carried forward from successful transfections and for which inhibition to ≤60% of residual levels by the positive control siRNA was reached at any time point on either plate, compared to the negative control. Subsequently, 9% of this data was discarded because data points showed a standard deviation of ≥30%. The final data set contained 2,431 sequences with Pearson coefficients of correlation of ≥0.7 between duplicate plates. The data points for each siRNA were averaged and used for neural net training and testing. Inhibition data before normalization will be made available on request to the authors.” This correction does not alter any of the data, the tables, the figures or the text in the remainder of the publication and supplementary material. The error has been corrected in the PDF versions of the article.