The transience of transient overexpression

Toby J Gibson, Markus Seiler & Reiner A Veitia

Much of what is known about mammalian cell regulation has been achieved with the aid of transiently transfected cells. However, overexpression can violate balanced gene dosage, affecting protein folding, complex assembly and downstream regulation. To avoid these problems, genome engineering technologies now enable the generation of stable cell lines expressing modified proteins at (almost) native levels.

In recent years, proteomic analyses have revealed that regulatory proteins almost always pull down groups of other proteins[1]. It is but a short logical step to conclude that regulatory proteins spend most of their time entwined in macromolecular complexes and that, therefore, a complete description of cell signaling will be a description of the dynamic behavior within and between these complexes. Regulatory decisions themselves, as we have previously argued, are made by in-complex cooperative molecular switching[2]. Furthermore, as part of the spatial organization of the cell, proteins are often synthesized in the cellular locations where they are needed[3], and many regulatory proteins exist at remarkably low numbers per cell[4,5]. These observations collectively contradict the simplistic view that proteins are made somewhere in the cytosol, diffuse as free agents to some 'useful' location and then cascade around the cell, thereby generating signaling pathways[6]. This pervasively held belief comes at least in part from experiments involving massive transient overexpression.

Transient transfection techniques have been continually developed and applied for over 30 years to address key questions in cell biology. Indeed, many high-profile studies of cell signaling have almost exclusively made use of transient overexpression. Such an approach obliterates subtlety and hence might result in overly simplistic models of signaling, as these networks are pervaded by cross-talk and feedback control mechanisms[7]. Here we argue for a change in the way that cell signaling is studied. We consider it unsafe to develop models of cell regulation based solely on transient overexpression and that results should be confirmed at native expression and/or *in vitro*. In the near future, the requirement to work in-cell at close to native expression levels will be aided by switching from transient transfection to stably transfected cell lines.

Unbalancing gene dosage

Genetic studies have shown that mutations that affect the expression levels of regulatory proteins are often deleterious and dominantly inherited. For good experimental design, it is important to understand that the physical interactions of regulatory proteins are what make them so sensitive to gene dosage.

Cooperative interactions in macromolecular complexes. A remarkably large number of known pairwise regulatory interactions display affinities in the low micromolar range that are too weak to make stable complexes[8,9]. This implies that most binary macromolecular interactions cannot by themselves create a regulatory event. Rather, regulatory decisions are made within complexes by cooperative molecular switching[2]. As we have previously argued, essentially all regulatory interactions in the cell are cooperative. This means that the individually weak binding interactions work with or against each other to create or destabilize specific conformations within the complexes. Dynamic switches between these metastable conformations underlie signal transduction. The properties of regulatory switches can be formalized and captured, as, for example, in the new "switches.ELM" bioinformatics resource[2]. Cooperativity is as pervasive for receptors in the plasma membrane as it is for transcription factors in the chromatin[10,11].

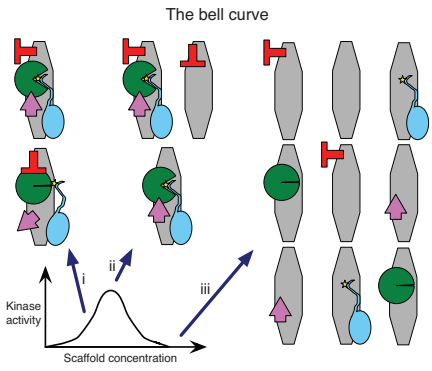

Figure 1 | Illustration of how varying a scaffold concentration can result in a bell-shaped curve of signaling protein activity. A scaffold (gray) may assemble a complex with kinase (green), kinase inhibitor (red), kinase activator (lilac) and substrate molecules (blue). Under normal cellular conditions, the complex is regulated by incoming signals (not shown) specifying activation or inactivation. When the complexes are activated (i, top), the phosphorylation site (yellow) can enter the kinase active-site cleft and be phosphorylated. Under inactive conditions (i, bottom), the site is excluded from the cleft. Increasing concentration of scaffold leads to partial complexes (ii), some of which may be constitutively active. Massive increase of scaffold (iii) dilutes the other components so that active complexes are no longer made.

Toby J. Gibson and Markus Seiler are in the Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. Reiner A. Veitia is at the Université Paris-Diderot–Paris 7 and Institut Universitaire de France and in Unité Mixte de Recherche 7592, Centre National de la Recherche Scientifique and Institut Jacques Monod, Paris, France.
e-mail: toby.gibson@embl.de or veitia.reiner@ijm.univ-paris-diderot.fr

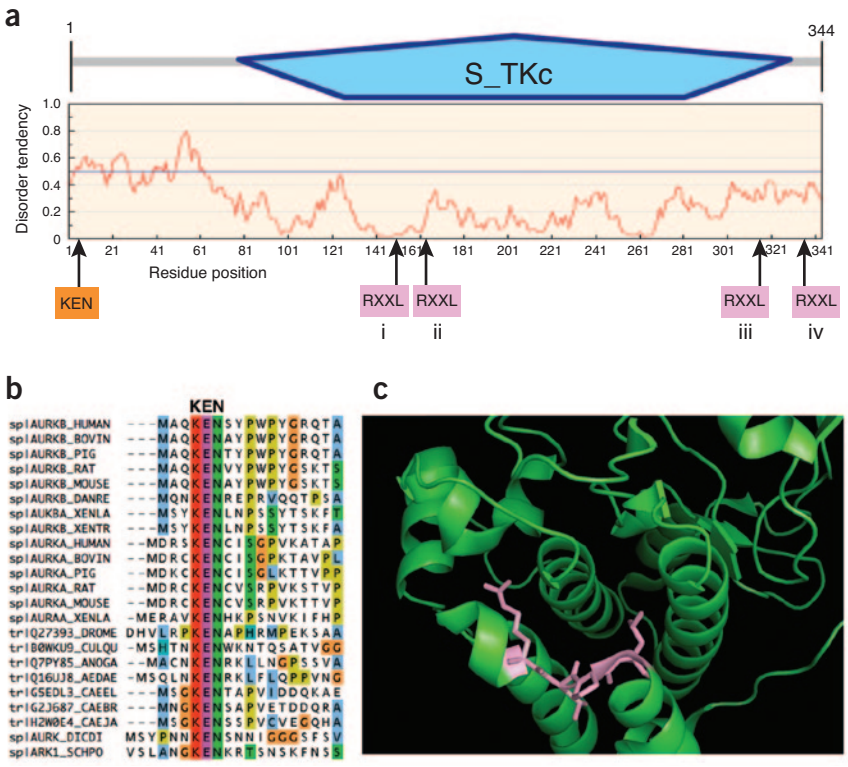

Figure 2 | Experimentally investigated candidates for the anaphase destruction motif in aurora B. (**a**) Intrinsically disordered protein (IUP; red) plot showing the order-disorder propensity for human aurora B. IUP values above 0.5 are likely to be intrinsically unstructured polypeptide[41]. The kinase domain (indicated by the kinase object S_TKc from the SMART resource[42]), has a globular fold, in agreement with a region predicted to be natively folded. Five candidate destruction motifs are indicated, three of which (i–iii) are in the kinase domain. (**b**) Alignment of the N-terminal KEN box showing conservation between vertebrate, insect, nematode, slime mold and yeast aurora kinases. (**c**) Structural context of RXXL motif iii (pink) reported to stabilize aurora B when mutated. The alignment was prepared with JalView[43], and the Protein Data Bank (PDB) structure file 2VGO (ref. 44) was displayed with PyMOL (http://www.pymol.org/).

Most gene products—RNAs and proteins—interact with multiple cellular components, often in a stoichiometric fashion. In agreement with this idea, in many organisms, components of stable protein complexes tend to have similar expression levels (are said to be 'isoexpressed')[12]. Perturbation of the stoichiometric balance in macromolecular complexes and cellular circuits can be a source of cellular malfunction and, therefore, of abnormal phenotypes. This is the basic claim of the gene dosage balance hypothesis[13].

Chromosomes highlight sensitivity to gene dosage. Studies of human aneuploidy syndromes suggest that the phenotypic effects of even moderate alterations in gene dosage are severe. It has long been known that trisomy 21 is the only viable autosomal aneuploidy in humans. Chromosome 21 is the shortest human autosome and also has the lowest number of dosage-sensitive genes[14]. Another well-known manifestation of a mechanism to maintain gene dosage balance is inactivation of the second X chromosome in female mammals[15].

Gene dosage artifacts. The overproduction of a protein that bridges two or more subunits of a complex may lead to the formation of inactive subcomplexes and alter biological output. This effect has been observed experimentally and has also been mathematically modeled in MAP kinase signaling, which is facilitated by scaffold proteins that place the molecular actors in close proximity. Experimental increase of the concentration of scaffold proteins to substoichiometric levels enhances signaling but at higher expression causes a decrease in signaling[16]. Beyond the optimal scaffold concentration, incomplete, inactive subcomplexes appear[17]. This leads to a skewed bell-shaped curve in a plot of kinase activity versus scaffold concentration (**Fig. 1**). Such bell curves are well known to pharmacologists, who use terms such as "biphasic," "inverted U-shaped" or "hormesis"[18] to describe these shapes.

In addition to the gene dosage balance artifacts, there can be other undesirable effects of overexpression. A key property of intrinsically disordered proteins (IUPs) is the capacity to make large numbers of interactions, which may further increase in number, but decrease in specificity, when IUPs are overexpressed[19]. Fluorophore tags may themselves have a preference for self-association such that, at high intracellular concentrations, they can cause expressed proteins to assemble into complexes[20]. Finally, flooding the cellular system with overexpressed proteins may result in ectopic cellular localizations[20]. Thus, many overexpressed proteins tend to show diffuse localizations, whereas at native expression they may display punctate staining patterns[20].

Systemic error caused by unbalanced gene dosage. Transcription factors (TFs) can be exquisitely sensitive to gene dosage effects, which may contribute to downstream system perturbations if the TFs are overexpressed. This is epitomized *in vivo* by eye-development TFs, such as the forkhead factor encoded by *FOXC1*, whose dosage alteration in any direction (typically heterozygous deletions or duplications) leads to similar ocular phenotypes[21]. It follows that experimental misexpression of such regulators may lead to unforeseen downstream side effects. TF overexpression can also give rise to global effects, as in the fly GAL4 system: this '*in vivo*' gene overexpression system involving the yeast transcription factor GAL4 is commonly used in *Drosophila melanogaster* to express genes driven by an upstream activator sequence (UAS). But as has been shown in a genomic analysis, high levels of GAL4 change the expression of many *Drosophila* genes in a UAS-independent manner[22]. Paradoxically, the overexpression of a potent transcriptional activator can suppress the transcription of its target genes. This phenomenon, called 'squelching', is considered to result from titration of limiting TFs[23], much in the way we have already discussed above in the context of activity bell curves.

Probing mutated proteins by overexpression

Overexpression is commonly used to study the effects of point mutations or trunc-

ations of a given protein. The abolition of a given function as a result of the mutation is often taken as logical proof that the function has been directly inactivated. Unfortunately, there may be alternative explanations for loss of function in this type of experiment. For instance, aggregation of misfolded protein in conditions of overexpression can lead to an oversimplified and misleading appraisal of biological reality as discussed in the following examples.

D-box difficulties: an unholy alliance between unfolding and overexpression. During anaphase, the anaphase-promoting complex, or cyclosome (APC/C), targets proteins for destruction, recognizing either KEN-box or D-box peptide tags[24]. The kinases aurora A and aurora B have an N-terminal KEN-box sequence that is strongly conserved and that exists in a region of intrinsically unstructured polypeptide (**Fig. 2**). Nguyen *et al.*[25] used overexpression transfections of GFP-tagged aurora B to demonstrate that the KEN box functions as a degron for APC/C-mediated degradation. They also mutated four predicted D boxes in aurora B and found that these mutations did not affect degradation. However, Stewart and Fang[26] used overexpression transfections to test mutations of two of the D-box motifs (but not the KEN box). They found that the mutation of one of them blocked in-cell protein degradation, from which they inferred that this was an APC/C-binding D box.

What might underlie the inconsistency of these results? The D box that was apparently functional in one of the overexpression transfection experiments is buried in the structural core of the kinase domain (**Fig. 2**), with the arginine and leucine from the motif facing inward (and in the case of the arginine, making buried salt-bridge interactions). Mutation of such buried residues in the kinase domain could unfold the kinase, inactivating it and resulting in aggregation: it is known that the exposure of the hydrophobic core in misfolded proteins often leads to aggregation[27].

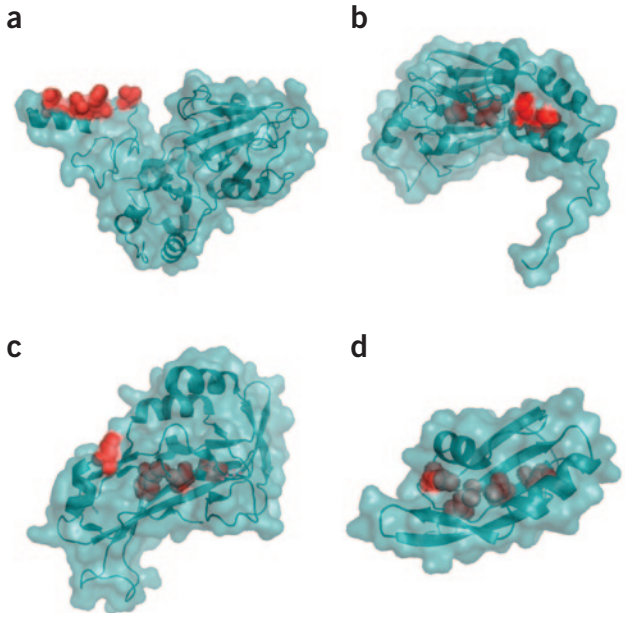

Figure 3 | Structural context of one true and three postulated NESs in human proteins. The protein surface is shown in transparent gray, with the main chain denoted as a cartoon. Red shapes mark amino acid positions matching the NES motif. (**a**) Structure of snurportin-1 as cocrystallized with the nuclear export factor CRM1 (PDB: 3GB8; ref. 45). (**b**–**d**) Postulated NES residues are part of the MH2 domain of SMAD1 (**b**), the T-box domain of TBX5 (**c**) and the ribonucleoprotein domain of UPF3 (**d**). NES residues are fully accessible only in **a**.

Nuclear export sequences in hiding. Nuclear export sequence (NES) motifs typically consist of a closely grouped tetrad of hydrophobic residues[28]. The NES motif needs to be accessible to bind to the export protein CRM1. Unsurprisingly, therefore, the well-studied canonical NESs such as those present in P53, PKI-α and HIV Rev are present in well-described IUP regions[29]. However, when protein databases are searched with such short hydrophobic motifs, these motifs overwhelmingly match buried residues of hydrophobic protein cores[29]. Thus, proposed NES motifs in c-ABL1, UPF3B and TBX5 do not have the possibility of ever being accessible to CRM1 under physiological conditions[30–32]. In **Table 1**, we provide a set of ten reported NESs, all of which are buried inside

Table 1 | Postulated nuclear export signals (NESs) in wrong structural contexts

UniProt protein name	UniProt accession	NES amino acid positions	NES amino acid sequence[a]	Correct hydrophobic spacing?[b]	Protein domain harboring this sequence	Ref.	Protein Data Bank structure	Status
ABL1	P00520	1083–1093	**L**ESN**L**RE**L**Q**I**C[c]	Yes	F-actin binding domain (FABD)	46	1ZZP	Rejected by Hantschel *et al.*[30]
AN32B	Q92688	109–120	**L**KK**L**EC**L**KS**L**D**L**	Yes	Histone chaperone domain	47	2ELL	This article
CTNA2	P26232	146–155	**V**MR**L**L**S**H**L**K**I**[c]	Yes	Inside four-helix bundle	48	1DOW	This article
FAK1	Q00944	90–102	**L**QSEE**V**HW**L**H**L**D**M**	No	FERM	49	2J0K	This article
LIMK1	P53667	231–242	**I**RN**V**P**L**DE**I**D**L**L[c]	No	PDZ	50	2YUB	Retracted article[d]
OAZ1	P54370	114–134	R**V**L**S**IQST**L**TEAKQ**V**TWRA**V**W	No	Inside β-fold	51	1ZO0	This article
SMAD1	Q15797	406–414	**L**TK**M**CT**I**R**M**	Yes	MH2	52	1KHU	This article
STAT1	P42224	302–314	WDRT**F**S**L**FQQ**L**IQ[c]	Yes	Inside four-helix bundle	53	1YVL	This article
TBX5	Q99593	152–160	**L**VS**F**QK**L**K**L**	Yes	T-box	54	2X6V	Rejected by Stirnimann *et al.*[32]
UPF3	P48412	88–97	**L**V**I**R**L**L**P**P**N**L[c]	No	RNP	55	1UW4	Rejected by Kadlec *et al.*[31]

[a]As defined in the original references; bold letters mark putative relevant hydrophobic residues of the postulated NES motif. [b]Spacing of hydrophobic residues is LX(2,3)LX(2,3)LXL as defined by Kutay and Güttinger[28]. [c]Different species were used to determine the postulated NES and to solve the structure; hence, the amino acid sequence may vary slightly. [d]This 'NES' was an artifact created by fabricated western blots. The article was retracted[50].

globular domains. A structural view (**Fig. 3**) contrasts the accessibility of NES residues in the true CRM1 interactor snurportin with three problematic NES candidates.

For most of these implausible NESs, the usual set of nuclear-export experiments (heavily dependent on overexpression) were performed: mutagenesis of the hydrophobic residues, leptomycin B inhibition of export, and assaying of the isolated NES peptide sequence in a strongly expressing reporter construct. The experimental results always appeared to support NES functionality. However, these results do not rule out an alternative explanation: mutation of buried residues may be expected to unfold the protein, and proteins bearing such unfolding mutations will probably be imported into the nucleus owing to the uncompromised nuclear localization signal. Potential aggregation of the unfolded protein may be restricted to the nucleus because this is the compartment where the overexpressed proteins accumulate. A combination of unfolding and aggregation will thus artifactually yield the expected result for abrogation of a bona fide NES.

Controlling overenthusiastic overexpression. There are some steps that researchers can take to identify artifactual results in transient transfection protocols. The most important control is testing the structural or functional integrity of mutant proteins by a different experimental approach, be it biochemical, biophysical or structural. For instance, if a protein kinase mutation is not supposed to destroy enzymatic activity, test the mutant kinase for its activity. If an SH2 domain is mutated to interfere with a binding interaction, check that it is still folded. A little biochemistry will go a long way. The second most important control is checking that the observation (that is, the interaction or activity) can also be detected at native expression levels. The third most important control is ruling out the presence of in-cell aggregates of a protein that is supposed to be soluble. In-cell protein precipitation can be revealed by the light microscope by, for example, birefringence of protein aggregates stained with Congo red[27]. Other microscopic approaches include molecular-rotor dye fluorescence[33], immunostaining to reveal the existence of non–wild-type protein distribution or aggregation and, for fluorophore-tagged proteins, fluorescence recovery after photobleaching[34]. A precipitated protein should result in termination of the experiment. Finally, protein function probed via transfection with presumed loss-of-function mutations should have results consistent with those of short interfering RNA knockdown of the endogenous gene.

Table 2 | Alternatives to transient overexpression

Technology/resource	Principle
Bacterial artificial chromosome (BAC) TransgeneOmics[36] http://hymanlab.mpi-cbg.de/bac_viewer/search.action http://www.mitocheck.org/ http://www.mitosys.org/	Efficient and scalable approach to stably integrate a second, tagged copy of a mammalian gene on a BAC; allows expression of the tagged protein at nearly native levels.
Recombinase-mediated cassette exchange[56]	Site-specific recombinases are used to exchange a DNA cassette flanked by appropriate sequences for a tagging cassette within the endogenous gene.
Dual recombinase-mediated cassette exchange[57] http://www.nature.com/protocolexchange/protocols/1906/	Uses *loxP* and FRT sites present in most conditional and many gene-trap alleles for site-specific targeting of a custom-designed cassette at the endogenous locus.
Zinc-finger nucleases[58]	Targeted nucleases that consist of zinc finger–based DNA-binding domains that can recognize a specific DNA sequence, coupled to a nuclease domain.
Transcription activator–like effector (TALE) nucleases[59,60]	Targeted nucleases that consist of DNA-binding domains from TALE proteins that can recognize a specific DNA sequence, coupled to a nuclease domain.
Targeted cleavage with clustered, regularly interspaced, short palindromic repeats (CRISPR)/Cas9 (refs. 61,62)	Targeted nucleases based on CRISPR/Cas9 systems. A custom guide RNA targets the Cas9 endonuclease activity to a desired DNA sequence.
ROSA26-based systems in mouse[63,64]	The *ROSA26* locus, commonly used to achieve wide expression, can be modified to allow controlled expression of a gene of interest from the *ROSA26* promoter, a minimal promoter or the gene's endogenous promoter; inducible if desired.
Flp-In System	Commercial Flp recombinase–based system (Invitrogen); provides an inducible promoter that can allow controlled perturbation by overexpression.
International Gene Trap Consortium[65] http://www.genetrap.org/	Gene trapping is used to create insertional mutations throughout the genome, generating banks of murine ES cell lines with traps in specific genes. Major gene-trapping groups have joined to centralize access to all publicly available gene-trap lines.
International Knockout Mouse Consortium (IKMC)[66] http://www.knockoutmouse.org/	Gene trapping and gene targeting in C57BL/6 ES cells to mutate all mouse protein-coding genes. The European Conditional Mouse Mutagenesis Program contributes the largest fraction of conditionally trapped and targeted genes to the IKMC.
Yeast GFP Clone Collection[67] http://clones.invitrogen.com/	Yeast strains expressing GFP fusion proteins from endogenous promoters.
Yeast deletion project[68] http://www-sequence.stanford.edu/group/yeast_deletion_project/	A genome-wide collection of yeast strains, each carrying a deletion of a single gene.
Drosophila transgenic RNAi library[69] http://stockcenter.vdrc.at/control/main/	A genome-wide library of *D. melanogaster* RNAi transgenes, allowing conditional gene knockdown in specific tissues of the intact organism.

This table includes a non-exhaustive list of techniques and resources in various experimental systems that can help overcome the limitations of transient overexpression.

Table 3 | Transient overexpression versus native expression

Biological property or feature	Issues with overexpression experiments	Issues with native-expression experiments
Low molecule number (<1,000 per cell)	If molecule number is important to function, expect disruption	May be difficult to visualize
Spatially restricted protein	May overload anchor sites; usually causes diffuse localization; protein may enter wrong cell compartment	None
Coupled mRNA transport and spatially restricted translation	Risk of overload to system; mistargeting, mistranslation	None
(Unknowingly) unfolded mutant proteins	Routinely expect amyloid/aggregation	Uncertain; possibility of aggregation
Balanced dosage-sensitive regulators	Abolished	None
Scaffolded kinases and kinase substrates	Scaffold overexpression dilutes components between scaffolds; nonlinear responses to overexpression, including bell curves	None
Laser bleaching to study motion of a signaling protein	Meaningless	None
Protein complexes by coimmunoprecipitation	Disrupts stoichiometry; may produce false interactors	None
Proteomic pulldowns	Disrupts stoichiometry; may produce false interactors	None
Reproducibility	Limited	None
Synchronized cell population	Limited time window	None
Differentiated cells from stem cells	Not feasible	None
Expression levels per cell	Variable DNA uptake leads to a wide range of expression levels and creates the temptation to cherry-pick cells	None
Alternatively spliced gene products	If a cDNA clone, only one variant is expressed	Ideally, no effect; however, this should be verified and not simply assumed
Off-target effects or nonspecific interactions	Almost certain	None expected

Here we contrast issues that can affect the efficacy or the interpretation of experiments that use overexpression versus native expression of a protein of interest.

The future of cell-signaling research: stably engineered cell lines

In the longer term it is desirable to move away from transient transfection techniques as a primary research tool. A well-established approach to modify an endogenous locus (whether for tagging, mutating or knocking-in) is homologous recombination. This allows the generation of engineered primary cells as well as immortalized ones, but it remains inefficient and time consuming in mammalian systems. It is also possible to place modified constructs regulated by their native promoter stably into embryonic stem (ES) cells that can undergo differentiation into other cell types of interest[35]. Tags may include a fluorescent protein for in-cell visualization and/or high-affinity epitopes for purification or other antibody targeting uses and may be placed at either the N or C terminus of the target protein.

One approach for generating stable lines involves the introduction of an extra tagged or otherwise modified copy of a gene by bacterial artificial chromosome (BAC) transgenesis[36]. This will result in mild overexpression and might therefore be unsuitable when tight gene dosage is critical, as with TFs in eye development such as FOXC1, PAX6 or SOX2. Moreover, BACs might not always contain all of the required regulatory sequences, leading to inappropriate expression patterns and levels. BAC transgenesis does have the compensatory advantage that there are two forms of the expressed protein, tagged and untagged, which can provide a useful quantitation control in proteomic analyses and can be used for rescue experiments to verify the specificity of RNAi experiments[37]. Additionally, stable BAC transgenic lines can be generated in a high-throughput manner[36] and be made available to other researchers, as with a recently published cell resource focused on motor proteins[38].

In a second approach, the gene of interest is engineered *in situ*; gene-trap methods are being adapted for this purpose[39]. A complete

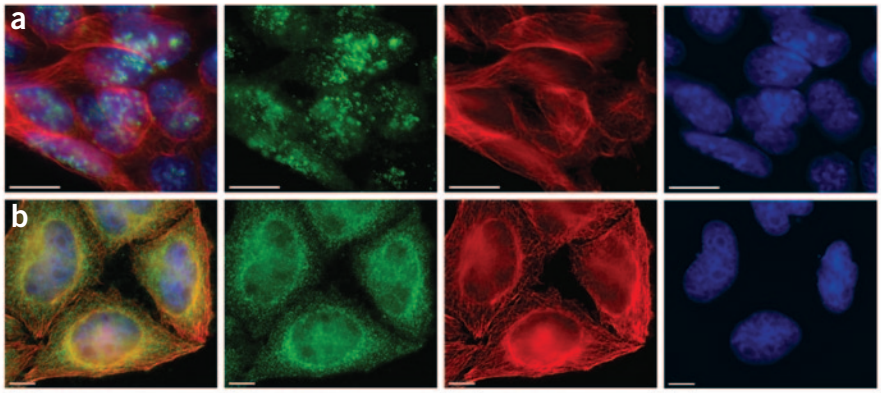

Figure 4 | Examples of discretely localized cellular proteins in engineered cell lines stably expressing GFP-tagged transgenes. (**a**) Mouse ES cells expressing tagged Ki67 (green), a nuclear protein used as a tumor proliferation marker. (**b**) HeLa cells expressing tagged mRae1 (green), an mRNA export factor. Green, GFP-tagged transgene; red, immunostaining of α-tubulin; blue, DAPI (4′,6-diamidino-2-phenylindole). Scale bars, 10 μm. Images courtesy of I. Poser, M. Augsburg and A. Nitzsche (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden).

review of genome engineering methods is beyond the scope of this paper; however, we provide a non-exhaustive list of techniques and resources currently available for modifying endogenous genes and for studying gene function in conditions closer to those *in vivo* (**Table 2**). In particular, targeted tools such as zinc-finger nucleases and transcription activator–like effector nucleases can be designed to cleave a genomic sequence of interest and thus substantially increase the efficiency of homologous recombination at a desired site. These tools are increasingly becoming cost effective, enabling many labs studying cell regulation to generate stable cell lines expressing modified proteins of interest. In parallel, these tools are now also beginning to be mass produced by both private and public enterprises[40].

As instruments of cell signaling research, stably tagged cell lines have fundamental properties that cannot be matched by transient approaches (**Table 3**) and that are often critical for experimental success. In contrast to GFP-tagged proteins in cells in which a gene is massively overexpressed[20], GFP-tagged proteins expressed in stable cell lines frequently show spatially restricted localizations (**Fig. 4**). We invite the reader to compare any of the images of GFP-tagged proteins expressed from BAC transgenes in Poser *et al.*[36] with images from an overexpression study of the same protein: even where some residual punctate structure is visible, overexpression typically causes dispersed fluorescence indicative of protein that is freely diffusing and/or involved in non-native interactions. With an engineered cell line expressing the tagged protein at native levels, by contrast, the protein can be followed under any condition. In particular, its behavior can be tracked during the cell cycle or in differentiating cells. It can be studied under a variety of cellular perturbations such as stress, RNAi knockdown or transient overexpression of interacting proteins. (Observe that transient overexpression will now be correctly treated as a perturbation!)

Genome-engineered cell lines don't just provide (approximately) native expression: they are also a tool to share between labs, and it follows that experiments involving such lines are closer to meeting desired goals of scientific reproducibility. Unfortunately, reproducibility has not been a characteristic of transient overexpression work[7].

We began this article by asking for a change in approach to the way that cell signaling is studied. The fulcrum upon which this change will be balanced is the use of stable genome-engineered cell lines to open up new avenues, some of them anticipated, others not. Banks of engineered cell lines will be generated over time and will complement gene knockout ES cell and mouse line collections already available from repositories such as the Sanger Centre (http://www.sanger.ac.uk/resources/mouse/) and the Jackson Laboratory (http://jaxmice.jax.org/). In parallel, genome engineering methods are increasingly accessible to nonspecialist labs, and the smaller-scale generation of lines driven by specific biological questions will undoubtedly accelerate as well.

Conclusion

Engineered cell lines now allow cell regulation to be studied with tagged proteins expressed at (almost) native levels. These tools can eliminate the need for transient overexpression, except as a 'quick and dirty' exploratory technique or when clearly designated as a deliberate cell perturbation. Transient transfection results can be suggestive but are not confirmatory. If we want to develop realistic models of cellular regulatory systems, we need to explore how the cellular components interact with each other under the normal range of cellular conditions. Used as the main tool, transient overexpression simply cannot deliver that knowledge, for it cannot reveal the subtleties and the complexities of cell regulation.

ACKNOWLEDGMENTS

We thank many colleagues at the European Molecular Biology Laboratory, Monod Institute and two research consortia, the German National Genome Research Network–funded DiGtoP and European Union–funded SyBoSS—which are focused on developing and applying genome engineered cell lines—for useful discussions. We apologize for not citing many important references because of space limitations. M.S. is funded by DiGtoP. R.A.V. is supported by Centre National de la Recherche Scientifique, La Ligue contre le Cancer (Comité de Paris), l'Université Paris Diderot–Paris 7 and Institut Universitaire de France. Special thanks to I. Poser, M. Augsburg and A. Nitzsche (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden) for providing the images of stably engineered cell lines.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Gingras, A.C., Gstaiger, M., Raught, B. & Aebersold, R. *Nat. Rev. Mol. Cell Biol.* **8**, 645–654 (2007).
2. Van Roey, K., Dinkel, H., Weatheritt, R.J., Gibson, T.J. & Davey, N.E. *Sci. Signal.* **6**, rs7 (2013).
3. Martin, K.C. & Ephrussi, A. *Cell* **136**, 719–730 (2009).
4. Beck, M. *et al. Mol. Syst. Biol.* **7**, 549 (2011).
5. Schwanhäusser, B. *et al. Nature* **473**, 337–342 (2011).
6. Gibson, T.J. *Trends Biochem. Sci.* **34**, 471–482 (2009).
7. Dumont, J.E., Dremier, S., Pirson, I. & Maenhaut, C. *Am. J. Physiol. Cell Physiol.* **283**, C2–C28 (2002).
8. Nooren, I.M. & Thornton, J.M. *J. Mol. Biol.* **325**, 991–1018 (2003).
9. Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G. & Orengo, C. *Structure* **18**, 1233–1243 (2010).
10. Volonté, C., D'Ambrosi, N. & Amadio, S. *Prog. Neurobiol.* **86**, 61–71 (2008).
11. Georges, A.B., Benayoun, B.A., Caburet, S. & Veitia, R.A. *FASEB J.* **24**, 346–356 (2010).
12. Ettwiller, L. & Veitia, R.A. *Comp. Funct. Genomics* **2007**, 58721 (2007).
13. Birchler, J.A. & Veitia, R.A. *Proc. Natl. Acad. Sci. USA* **109**, 14746–14753 (2012).
14. Makino, T. & McLysaght, A. *Proc. Natl. Acad. Sci. USA* **107**, 9270–9274 (2010).
15. Okamoto, I. *et al. Nature* **472**, 370–374 (2011).
16. Cacace, A.M. *et al. Mol. Cell. Biol.* **19**, 229–240 (1999).
17. Levchenko, A., Bruck, J. & Sternberg, P.W. *Proc. Natl. Acad. Sci. USA* **97**, 5818–5823 (2000).
18. Calabrese, E.J. & Baldwin, L.A. *Trends Pharmacol. Sci.* **22**, 285–291 (2001).
19. Vavouri, T., Semple, J.I., Garcia-Verdugo, R. & Lehner, B. *Cell* **138**, 198–208 (2009).
20. Rizzo, M. A., Davidson, M. W. & Piston, D. W. *Cold Spring Harb. Protoc.* **2009**, pdb.top64 (2009).
21. Lehmann, O.J. *et al. Am. J. Hum. Genet.* **67**, 1129–1135 (2000).
22. Liu, Y. & Lehmann, M. *Fly (Austin)* **2**, 92–98 (2008).
23. Natesan, S., Rivera, V.M., Molinari, E. & Gilman, M. *Nature* **390**, 349–350 (1997).
24. Peters, J.M. *Nat. Rev. Mol. Cell Biol.* **7**, 644–656 (2006).
25. Nguyen, H.G., Chinnappan, D., Urano, T. & Ravid, K. *Mol. Cell. Biol.* **25**, 4977–4992 (2005).
26. Stewart, S. & Fang, G. *Cancer Res.* **65**, 8730–8735 (2005).
27. Dobson, C.M. *Nature* **426**, 884–890 (2003).
28. Kutay, U. & Güttinger, S. *Trends Cell Biol.* **15**, 121–124 (2005).
29. Diella, F. *et al. Front. Biosci.* **13**, 6580–6603 (2008).
30. Hantschel, O. *et al. Mol. Cell* **19**, 461–473 (2005).
31. Kadlec, J., Izaurralde, E. & Cusack, S. *Nat. Struct. Mol. Biol.* **11**, 330–337 (2004).
32. Stirnimann, C.U., Ptchelkine, D., Grimm, C. & Müller, C.W. *J. Mol. Biol.* **400**, 71–81 (2010).
33. Shen, D. *et al. Cell Biochem. Biophys.* **60**, 173–185 (2011).
34. Roberti, M.J., Jovin, T.M. & Jares-Erijman, E. *PLoS ONE* **6**, e23338 (2011).
35. Ciotta, G. *et al. Methods* **53**, 113–119 (2011).
36. Poser, I. *et al. Nat. Methods* **5**, 409–415 (2008).
37. Ding, L., Poser, I., Paszkowski-Rogacz, M. & Buchholz, F. *Stem Cell Rev.* **8**, 32–42 (2012).
38. Maliga, Z. *et al. Nat. Cell Biol.* **15**, 325–334 (2013).
39. Schebelle, L. *et al. Nucleic Acids Res.* **38**, e106 (2010).
40. DeFrancesco, L. *Nat. Biotechnol.* **29**, 681–684 (2011).
41. Dosztányi, Z., Csizmok, V., Tompa, P. & Simon,

I. *Bioinformatics* **21**, 3433–3434 (2005).
42. Letunic, I., Doerks, T. & Bork, P. *Nucleic Acids Res.* **37**, D229–D232 (2009).
43. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. *Bioinformatics* **25**, 1189–1191 (2009).
44. D'Alise, A.M. *et al. Mol. Cancer Ther.* **7**, 1140–1149 (2008).
45. Dong, X. *et al. Nature* **458**, 1136–1141 (2009).
46. Taagepera, S. *et al. Proc. Natl. Acad. Sci. USA* **95**, 7457–7462 (1998).
47. Fries, B. *et al. J. Biol. Chem.* **282**, 4504–4515 (2007).
48. Giannini, A. *et al. Exp. Cell Res.* **295**, 150–160 (2004).
49. Ossovskaya, V., Lim, S.T., Ota, N., Schlaepfer, D.D. & Ilic, D. *FEBS Lett.* **582**, 2402–2406 (2008).
50. Roovers, K., Klein, E.A., Castagnino, P. & Assoian, R.K. *Dev. Cell* **5**, 273–284 (2003).
51. Murai, N., Murakami, Y. & Matsufuji, S. *J. Biol. Chem.* **278**, 44791–44798 (2003).
52. Xiao, Z., Watson, N., Rodriguez, C. & Lodish, H.F. *J. Biol. Chem.* **276**, 39404–39410 (2001).
53. Begitt, A., Meyer, T., van Rossum, M. & Vinkemeier, U. *Proc. Natl. Acad. Sci. USA* **97**, 10418–10423 (2000).
54. Kulisz, A. & Simon, H.G. *Mol. Cell. Biol.* **28**, 1553–1564 (2008).
55. Shirley, R.L., Ford, A.S., Richards, M.R., Albertini, M. & Culbertson, M.R. *Genetics* **161**, 1465–1482 (2002).
56. Turan, S. *et al. J. Mol. Biol.* **407**, 193–221 (2011).
57. Osterwalder, M. *et al. Nat. Methods* **7**, 893–895 (2010).
58. Kim, Y.G., Cha, J. & Chandrasegaran, S. *Proc. Natl. Acad. Sci. USA* **93**, 1156–1160 (1996).
59. Zhang, F. *et al. Nat. Biotechnol.* **29**, 149–153 (2011).
60. Boch, J. *Nat. Biotechnol.* **29**, 135–136 (2011).
61. Cong, L. *et al. Science* **339**, 819–823 (2013).
62. Mali, P. *et al. Science* **339**, 823–826 (2013).
63. Casola, S. *Methods Mol. Biol.* **667**, 145–163 (2010).
64. Soriano, P. *Nat. Genet.* **21**, 70–71 (1999).
65. Nord, A.S. *et al. Nucleic Acids Res.* **34**, D642–D648 (2006).
66. Skarnes, W.C. *et al. Nature* **474**, 337–342 (2011).
67. Huh, W.K. *et al. Nature* **425**, 686–691 (2003).
68. Winzeler, E.A. *et al. Science* **285**, 901–906 (1999).
69. Dietzl, G. *et al. Nature* **448**, 151–156 (2007).