Methods xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

TransgeneOmics – A transgenic platform for protein localization based function exploration

Susanne Hasse, Anthony A. Hyman, Mihail Sarov *

Max Planck Institute of Molecular Cell Biology and Genetics Dresden, Germany

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form 9 October 2015
Accepted 10 October 2015
Available online xxxx

Keywords:
BAC transgenesis
Recombineering
Protein localization
Genome engineering
Protein tagging

ABSTRACT

The localization of a protein is intrinsically linked to its role in the structural and functional organization of the cell. Advances in transgenic technology have streamlined the use of protein localization as a function discovery tool. Here we review the use of large genomic DNA constructs such as bacterial artificial chromosomes as a transgenic platform for systematic tag-based protein function exploration.

1. Introduction

For decades, immunostaining with protein specific antibodies has provided a reliable method for protein localization and antibodies are now available for a large number of proteins. Over the last 10 years the Human Protein Atlas program (HPA) has generated a collection of over 50,000 polyclonal antibodies targeting 19,100 (95%) human protein-coding genes [1]. Many of them have already been used to generate validated protein expression profiles and to systematically map the subcellular localization in fixed cells [2–5]. Antibodies are a versatile affinity reagent and can be applied in any cell line or tissue from the target species and even across species when the epitope is conserved. However, antibodies have a number of known limitations that have to be taken into account in every experiment. Although *in vivo* localization approaches based on fluorescently labeled single chain Fab fragments (scFabs) or camelid single chain antibodies (often referred as nanobodies) is possible [6–8], the vast majority of the available antibodies require cell fixation and permeabilization, which can cause cell shrinkage or leakage of endogenous components [9,10]. Antibodies are also prone to cross-reactivity and often cannot discriminate between protein isoforms that have different subcellular localizations [11].

The ability to observe protein localization *in vivo* can often reveal functionally relevant dynamics that cannot be inferred from immunostaining [12–14]. Since the first use of the *Aequorea victoria* green fluorescent protein (GFP) over 20 years ago [15] a large array of fluorescent proteins with various properties and methods for expression of fluorescently tagged proteins (Table 1) have been developed [16,17]. The cloning of cDNA derived open reading frames (ORFs) into standard expression vectors was one of the early approaches for systematic tag-based protein localization mapping [18]. A proteome-scale localization map was generated in budding yeast through the systematic cloning of ORFs into plasmid vectors for inducible overexpression of V5 tagged proteins, followed by immunostaining with an anti-V5 antibody [19] and similar approaches have been used in other systems [20–22]. Due to its simplicity this approach remains very popular and large scale ORF resources are now available for many commonly used model systems, typically in vector formats that allow the easy shuffling to expression vectors containing fluorescent or epitope tags [20,23–26]. However, the heterologous promoters and 3′ regulatory elements often used with these type of vectors do not reflect the endogenous expression levels of most proteins and can disturb cellular functions. Since certain diseases can be caused by abnormally high protein levels [27], cDNA transgenes have been used in systematic screens for such phenotypes, for example in an activator screen for the antioxidant response element in human cells [28]. In addition, overexpression may lead to saturation of the specific binding sites of a protein leading to lead to mislocalization and/or obscuring of the normal localization pattern.

The most reliable way to ensure endogenous regulatory control of the tagged protein expression is the direct targeting of the genomic locus. Two major strategies, the transposon or viral vector mediated random insertion and the homology directed precise

* Corresponding author.

http://dx.doi.org/10.1016/j.ymeth.2015.10.005
1046-2023/

Table 1
Comparison of various methods for system-scale protein localization.

	Cost	Scalability	Signal strength	Isoform selectivity	Major advantages	Major limitations
Immunostaining	High	Low	Varies	Low	Applicable to any sample	Cross-reactivity, cost
cDNA transgenes	Low	Good	High	Good	Ease of cloning	Position effects, overexpression
gDNA transgenes	Low	Good	Near native levels	Good	Comparable with targeting at a lower cost, hypomorphic or lethal protein variants can be studied	Position effects, fragmentation
Gene targeting	High	Low	Native levels	Good	Endogenous expression levels	High costs, the tag can affect gene function

targeting, have been developed in parallel with their own advantages and limitations. The main advantage of the random approach is the ability to rapidly generate and map a large number of insertions. The tag can be either inserted directly or through an exchange of a previously inserted gene trap cassette by recombinase mediated cassette exchange (RMCE) [19,29–35]. Some of the technical limitations of this approach, including a insertion bias towards a particular sequence content, were mostly overcome with later generations of these vectors. However, all random approaches suffer from diminishing returns, where the number of repeatedly hit genes grows and the number of new genes rapidly declines, and only a fraction of the inserts can result in a functional tagged protein. Until recently, homology directed repair (HDR) was only practical in a small number of model systems and has only been used for proteome-scale localization studies in a few species including *Saccharomyces cerevisiae* [36]. The discovery of the CRISPR/Cas technology has dramatically increased the efficiency of gene targeting in systems where such tools were already available and has made genome engineering feasible in species that so far appeared resistant to this approach. However, a number of technical and logistic challenges still need to be addressed before systematic genome-wide HDR targeting based tagging in mammalian systems becomes reality.

Genomic DNA (gDNA) transgenes provide an alternative that combines the ease and efficiency of cDNA transgenics with preservation of the endogenous regulatory expression control comparable with gene targeting. In this review, we explore various applications of the gDNA transgene approach and compare it to other methods for protein localization.

2. gDNA transgenesis as platform for protein function exploration

2.1. Methods for gDNA transgene engineering

The first attempts to use gDNA for large scale GFP tagging was cloning of random fragments of the fission yeast genome into expression vectors, resulting in one of the early protein localization maps [37]. This approach is not practical for metazoan genes, which can be orders of magnitude bigger than in yeast. However, large insert gDNA libraries of bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs), or yeast artificial chromosomes (YACs) that can cover most genes with their endogenous cis regulatory context are available for many model systems. When used as transgenes, these clones typically exhibit near-physiological levels and patterns of gene expression and are routinely used in mapping mutations by complementation [38]. YACs can be engineered *in vivo* by homologous recombination in yeast [39,40] and the development of efficient methods for homologous recombination mediated engineering in *Escherichia coli* [41,42] made it possible to apply this approach to the easier to handle fosmid/BAC/PAC constructs. Most *E. coli* cloning hosts exhibit very low levels of spontaneous recombination, which ensures a stable maintenance of large and repetitive clones. A simple two-component recombination system, consisting of an exonuclease and a strand annealing protein from the phage lambda or related phages, is sufficient for an efficient and precise homologous recombination in *E. coli* with only 30–50 bp of homology [41,43,44]. As recombineering is independent of the availability of restriction sites and can be used to insert, delete, change or retrieve any sequence of interest to and from any DNA that can be propagated in *E. coli* it was soon adapted to a wide spectrum of applications. BAC recombineering provided an efficient way to engineer point mutations, create translational or transcriptional reporters and targeting constructs for genome editing [45–48] for almost any gene of interest in a range of model organisms (mammals: [49–51]; fly: [52,53]; worm: [54–57]). With recombineering, the tag coding sequence can be inserted into the gene of interest at the N- or the C-terminus or at any internal position of choice, allowing the tagging of specific splice forms. The tag is typically inserted as a cassette with a selectable marker, which can be removed by site-specific excision through recombinases like Cre or Flp. Methods have been developed to allow the easy exchange of the tag in the transgene by either homologous recombination [45] or RMCE [58].

2.2. System-scale gDNA transgenesis

The high recombineering efficiency, near absence of unintended recombination background and the ease of handling *E. coli* in liquid culture lead to the development of high throughput pipelines for multi-step recombineering in a 96 well plate format for parallel engineering of a large number of constructs in *Caenorhabditis elegans* [59], mammals [60] and Drosophila [52,53,61] (Fig. 1). Using this approach a *C. elegans* library of 14637 tagged fosmid transgenes was constructed, which enabled the *in vivo* localization or affinity purification of 73% of the proteome [54]. The fosmid collection was used to create over 580 transgenic *C. elegans* lines. All constructs, lines and associated localization patterns are accessible at https://transgeneome.mpi-cbg.de. Since the release of the *C. elegans* TrangeneOme resource, 2806 constructs have been distributed to labs around the world in a community-wide effort to generate a proteome scale resource of tagged transgenic lines. Using a similar approach [52,53,61] a collection covering approximately 10000 *Drosophila* genes was recently created, from which transgenic lines for 826 genes were already established. The 'tagged FlyTransgeneOme' (fTRG) lines are available at the VDRC stock centre (http://stockcenter.vdrc.at) [61]. Systematic BAC transgenesis in mammalian cell lines [60] has led to the generation of over 6000 BAC-tagged cell lines, including HeLa, U2OS, mouse embryonic stem cells and MDCK (http://hymanlab.mpi-cbg.de/bac). Over 5800 lines have been sent to 160 labs in 26 countries. Their applications spanned from the investigation of small amounts of different lines [62,63] to studies involving several hundred BAC-lines [49–51,64]. Hutchins et al. studied 696 tagged human proteins that carried a C-terminal localization and

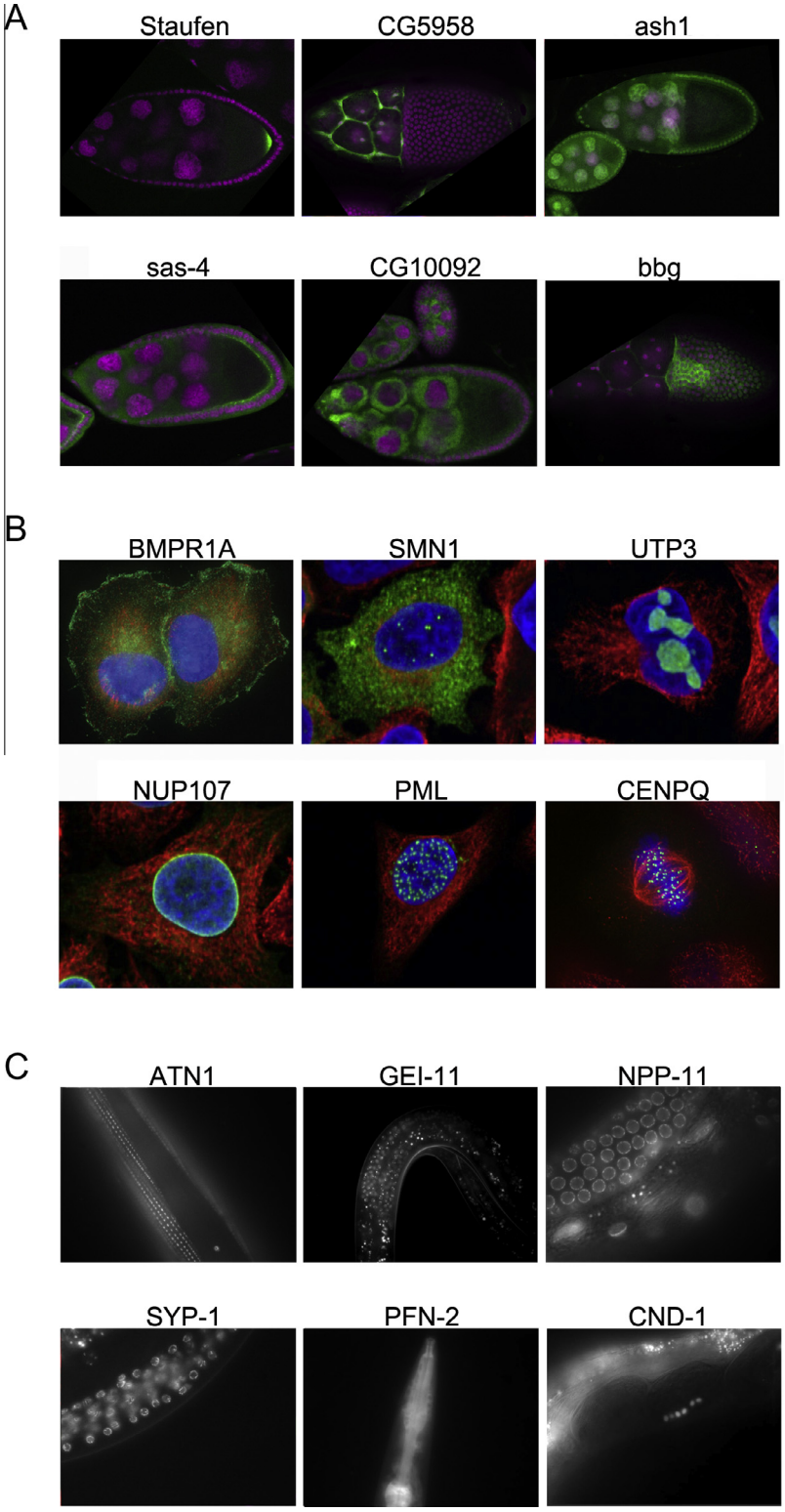

Fig. 1. Selected examples of gDNA transgene derived specific subcellular localization patterns. (A) *D. melanogaster*. GFP fluorescence is shown in green, DAPI staining for DNA in magenta. All proteins show expression patterns consistent with their known function, for example the maternal effect protein Staufen localizes to the posterior pole of the egg, while Ash-1 localizes in the nucleus. (B) Transgenic HeLa cells. An antibody staining enhanced GFP-signal (green), Actin (red) and DAPI stained DNA (blue). The receptor protein Bmpr1a localizes to the plasma membrane; as one of the core subunits of the SMN complex, SMN1 forms granules in the cytoplasm corresponding to the cytosolic assembly of small nuclear ribonucleoprotein particles (snRNPs) by the SMN complex and subnuclear bodies consistent with the final maturation of the snRNPs within Cajal bodies of the nucleus [93]; Small dots are formed as well by Pml, presumably Pml-bodies; Utp3 localizes specifically to the nucleolus; the kinetochore protein Cenpq localizes where it is required for proper kinetochore function and mitotic progression [94]; while the nuclear pore complex protein Nup-107 localizes to the nuclear membrane. (C) *C. elegans*. GFP fluorescence is shown. Actinin-1 (Atn-1) and Profilin-2 (Pfn-2) localize to muscles, while the nuclear pore complex protein Npp-1 can be found at the nuclear envelope. Gei-11 is a transcription factor that specifically regulates the expression of the snRNA genes, which are clustered on chromosome IV, resulting in specific punctate subnuclear localization; Syp-1 is a member of the synaptonemal complex; Cnd-1 is the worm NeuroD ortholog.; Images taken (with permission) from: http://tomancak-srv1.mpi-cbg.de/DOT/main.html, http://hymanlab.mpi-cbg.de/bac/login.jsp, https://transgeneome.mpi-cbg.de.

Please cite this article in press as: S. Hasse et al., Methods (2015), http://dx.doi.org/10.1016/j.ymeth.2015.10.005

affinity-purification tag (LAP-tag) for investigating protein complexes in mitosis. The exploration of their localization and physical interaction partners led to the discovery of complexes essential for spindle assembly and chromosome segregation [50]. Maliga et al. used BAC transgenics to explore the function of 243 mouse and human motor proteins [51].

2.3. Advantages and known limitations of the gDNA transgene approach

The main advantage of the gDNA/BAC approach is the high efficiency of engineering and transgenesis, which can reach the 80–90% range in mammalian tissue culture applications [60]. Combined with the low costs of this approach, gDNA transgenesis provides a way to quickly explore the function for large sets of genes. Due to their large size, BACs appear more resistant to position effects than smaller plasmid transgenes, but are prone to fragmentation and rearrangement during transgenesis and independent lines should be obtained and compared to verify the observed patterns. The integrity and localization of the BAC can be determined by targeted locus amplification (TLA) [65]. This issue can be overcome through various techniques that ensure intact, single copy transgene integration. In Drosophila, the efficient phiC31 integrase mediated insertion of a fosmid/BAC transgene at a pre-defined genomic position is reliable and a large number of insertion sites are available [35,66,67]. In mammalian cells the *PiggyBac* and the *Sleeping Beauty* transposons have been shown to efficiently jump into the genome with intact BAC sized cargo, typically as a single copy [68,69]. A similar approach using the *Drosophila Mauritania* transposon *Mos1* works efficiently in *C. elegans* [70]. Recently, an approach was developed that utilizes programmable nucleases like Zn fingers, TALENs or Cas9 to induce NHEJ mediated insertion of transgenes at any point of interest in the target genome that appears suitable for large inserts [71].

Typically, gDNA transgenes do not suffer from the strong overexpression common for cDNA transgenes [11], however, they are still present in addition to the endogenous alleles. As many proteins tend to have similar expression levels as their binding partners, in some cases even a slight overexpression can perturb this balance causing mislocalisation and abnormal phenotypes [72]. On the other hand, a slight overexpression would in most cases be tolerated by the cell and may even be beneficial for the localization of proteins expressed at levels close to the detection limit for *in vivo* imaging.

A known limitation of any tagging approach is the potential for the tag to interfere with the normal protein function by altering its charge and size or by affecting protein folding, degradation, the binding of interaction partners *etc*. However, improvements in tag design, the use of flexible linkers and rational design of the tag insertion points (to avoid interference with signaling peptides or catalytic domains for example) have helped to minimize this problem, and systematic comparisons with immunofluorescent labeling or rescue of mutant phenotypes [61] with tagged transgenes [11] show that this is a minor issue that can be controlled for and typically overcome. This is particularly straightforward with gDNA transgene recombineering as there are practically no restrictions on the tag insertion position.

2.4. Applications of BAC transgenics beyond localization

Once generated, the BAC transgenes can be used with any cell line or a model organism that can be efficiently transfected and have many additional applications (Fig. 2). A large variety of tagging cassettes are now available that reflect the specifics of the model organism and the biological question they are used to address. Most tagging cassettes use a combination of fluorescent

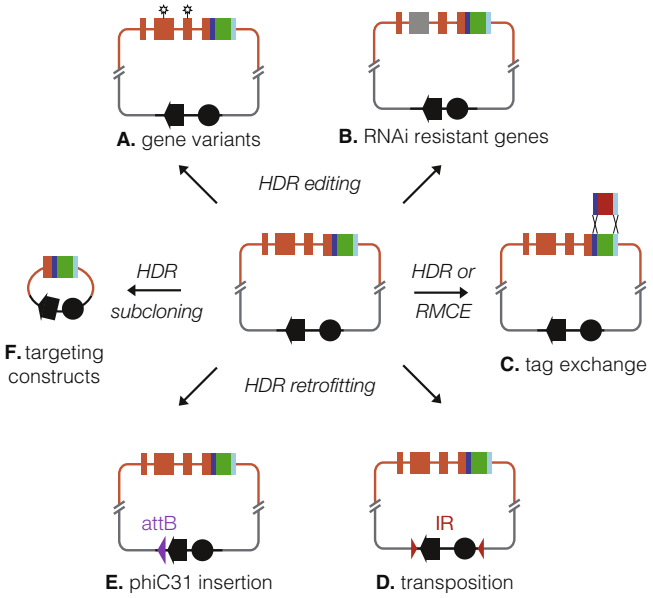

Fig. 2. BAC transgenes as a versatile resource. (A, B) The tagged BAC transgenes can be easily engineered for various downstream applications that increase their utility. Homology directed recombineering (HDR) allows the insertion of mutations (A) to investigate the function of protein variants and/or to make the transgenes resistant to RNAi (B) targeting the endogenous counterpart, which can be used to rescue RNAi phenotypes to confirm their specificity or for structure–function studies. (C) Several methods have been generated that allow the easy exchange of the tag coding sequence in the transgene through HDR or site-specific recombinase mediated cassette exchange (RMCE). (D, E) Various methods for random or site-specific single copy insertion of intact BAC transgenes have been developed. BACs can be easily retrofitted for these approaches by either inserting inverted repeats for transposons (D) or the phiC31 recombinase recognition site *attB* (E). The tagged BACs are also an intermediate step in the generation of targeting constructs (F). Subcloning by recombineering is efficient as homology arms of any designed length can be subcloned. It is not affected by the specific sequence composition (high/low GC content) as with other PCR-based methods.

and affinity epitopes and can be used for both localization and purification [45,52,60,73]. Earlier affinity purification protocols favored a multi-step approach to ensure highly pure complexes, but with the development of stable isotope labeling (SILAC) or label-free quantitative mass-spec methods, a single step purification is often sufficient and can preserve weaker integrations that might otherwise be missed [74–76]. Tag specific purification makes other affinity applications such as chromatin immunopurification (ChIP) [45,60,77–80], photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-Clip) [81] more comparable and reliable as it removes variability in non-specific antibody interaction.

Beyond the simple tagging with the commonly used fluorescent proteins and affinity epitopes the tagging cassettes can add useful properties that cannot be achieved with the antibody based techniques. For example, poly-cistronic expression cassettes have been developed for various species including mammalian cells, *C. elegans* [55,82] and *Drosophila* [83,84], which can be used to drive the expression of a selection marker or to facilitate the identification of the expressing cell by targeting a fluorescent reporter to a specific and easy to identify compartment like the cell nucleus [55]. Identifying the expressing cell in complex tissues with antibody staining can be non-trivial, especially for highly polarized and irregularly shaped cells like neurons where a protein can localize far away from the cell body. The expression of multiple proteins from the same mRNA can be facilitated by the use of viral 2A and internal ribosomal entry site (IRES) sequences. The 2A sequence induces a "ribosomal skip" during translation, which results in a release of the first peptide without interrupting translation of the

following peptide [85]. IRES sequences work by causing additional initiation of translation from the same transcript. In *C. elegans* a bi-cistronic expression has been achieved using the trans-splice leader 2 (SL2) sequence, which results in the expression of 2 mRNAs [55]. As the expression levels may vary depending on the inter-cistronic region used, the viral 2A peptides should be preferred over the IRES [86] or SL2 [55] sequences when near-stoichiometric expression levels are required [87].

BACs and fosmids can be used to explore the changes in the localization dynamics of gene variants (SNPs, insertions, deletions) that can be easily introduced into the transgene by recombineering. Since the endogenous gene locus is not disrupted, the function of gene mutations that would otherwise be lethal could also be studied. RNA interference against the endogenous protein in combination with an RNAi-resistant fusion protein can be used to knock down the endogenous protein for structure–function studies. This approach can be used in reverse to confirm the specificity of an RNAi experiment in human cells using RNAi-resistant BACs carrying the orthologous mouse gene [88]. A similar strategy can be applied to make a BAC-transgene resistant to Cas9 cleavage of the endogenous locus.

3. Summary and perspective

In the last two decades, we have seen an enormous progress in our ability to understand protein function though localization and many tools and resources have made these studies much more straightforward and routine. So far the only comprehensive localization pattern resources in model organisms more complex than yeast are based on antibody staining, which limits our ability to understand protein localization dynamics *in vivo*. The large-scale BAC transgenesis applications demonstrate that systems scale protein tagging is feasible and these efforts will likely accelerate with the arrival of the CRISPRS/Cas technology for direct genome editing, allowing the *in vivo* interrogation of any protein of interest at near-physiological expression levels. In recent years new techniques for organ-like induced pluripotent stems cells [89–91] and organoids [92] have made it possible to address a number of experimental questions in human cells that would have previously required the use of potentially less disease relevant model organisms. In combination with modern imaging techniques like SPIM that allow for long-term live imaging of mesoscale samples with minimal photodamage, these tools would allow us to understand the restructuring of the cellular proteome in development and disease, and to assign functions to thousands of previously uncharacterized human proteins.

Acknowledgements

We are grateful to Helena Jambor and Ina Poser (MPI-CBG) for providing protein localization examples for Fig. 1. Some of the resources mentioned in this paper were generated with the support of grants from the Max Planck Society (MPG Vorhaben 8070 “A reverse genetic toolkit for systematic study of gene function and protein localization in Drosophila”) and the German Federal Ministry of Education and Research (BMBF 031A099 “Collective organization of cells and tissues: Systems biology of tissue size and shape”).

References

[1] C. Stadler et al., RNA- and antibody-based profiling of the human proteome with focus on chromosome 19, J. Proteome Res. 13 (4) (2014) 2019–2027.
[2] L. Fagerberg et al., Mapping the subcellular protein distribution in three human cell lines, J. Proteome Res. 10 (8) (2011) 3766–3777.
[3] F. Ponten, K. Jirstrom, M. Uhlen, The Human Protein Atlas – a tool for pathology, J. Pathol. 216 (4) (2008) 387–393.
[4] M. Uhlen et al., A Human Protein Atlas for normal and cancer tissues based on antibody proteomics, Mol. Cell. Proteomics 4 (12) (2005) 1920–1932.
[5] M. Uhlen et al., Proteomics. Tissue-based map of the human proteome, Science 347 (6220) (2015) 1260419.
[6] U. Rothbauer et al., Targeting and tracing antigens in live cells with fluorescent nanobodies, Nat. Methods 3 (11) (2006) 887–889.
[7] P. Panza et al., Live imaging of endogenous protein dynamics in zebrafish using chromobodies, Development 142 (10) (2015) 1879–1884.
[8] A. Burgess, T. Lorca, A. Castro, Quantitative live imaging of endogenous DNA replication in mammalian cells, PLoS One 7 (9) (2012) e45726.
[9] U. Schnell et al., Immunolabeling artifacts and the need for live-cell imaging, Nat. Methods 9 (2) (2012) 152–158.
[10] D.R. Whelan, T.D. Bell, Image artifacts in single molecule localization microscopy: why optimization of sample preparation protocols matters, Sci. Rep. 5 (2015) 7924.
[11] C. Stadler et al., Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells, Nat. Methods 10 (4) (2013) 315–323.
[12] R.M. Hoffman, Application of GFP imaging in cancer, Lab. Invest. 95 (4) (2015) 432–452.
[13] R.M. Hoffman, Imaging metastatic cell trafficking at the cellular level *in vivo* with fluorescent proteins, Methods Mol. Biol. 1070 (2014) 171–179.
[14] O. Filhol et al., Live-cell fluorescence imaging reveals the dynamics of protein kinase CK2 individual subunits, Mol. Cell. Biol. 23 (3) (2003) 975–987.
[15] M. Chalfie et al., Green fluorescent protein as a marker for gene expression, Science 263 (5148) (1994) 802–805.
[16] A.S. Mishin et al., Novel uses of fluorescent proteins, Curr. Opin. Chem. Biol. 27 (2015) 1–9.
[17] J. Zhang et al., Creating new fluorescent probes for cell biology, Nat. Rev. Mol. Cell Biol. 3 (12) (2002) 906–918.
[18] J.C. Simpson et al., Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing, EMBO Reports 1 (3) (2000) 287–292.
[19] A. Kumar et al., Subcellular localization of the yeast proteome, Genes Dev. 16 (6) (2002) 707–719.
[20] J. Bischof et al., Generation of a transgenic ORFeome library in Drosophila, Nat. Protoc. 9 (7) (2014) 1607–1620.
[21] P. Lamesch et al., hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes, Genomics 89 (3) (2007) 307–315.
[22] J. Reboul et al., *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression, Nat. Genet. 34 (1) (2003) 35–41.
[23] P. Uetz et al., A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*, Nature 403 (6770) (2000) 623–627.
[24] K. Yamada et al., Empirical analysis of transcriptional activity in the Arabidopsis genome, Science 302 (5646) (2003) 842–846.
[25] G. Temple et al., The completion of the Mammalian Gene Collection (MGC), Genome Res. 19 (12) (2009) 2324–2333.
[26] P. Lamesch et al., *C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions, Genome Res. 14 (10B) (2004) 2064–2069.
[27] T. Santarius et al., A census of amplified and overexpressed human cancer genes, Nat. Rev. Cancer 10 (1) (2010) 59–64.
[28] Y. Liu et al., A genomic screen for activators of the antioxidant response element, Proc. Natl. Acad. Sci. U.S.A. 104 (12) (2007) 5205–5210.
[29] P. Ross-Macdonald et al., Large-scale analysis of the yeast genome by transposon tagging and gene disruption, Nature 402 (6760) (1999) 413–418.
[30] X. Morin et al., A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila, Proc. Natl. Acad. Sci. U.S.A. 98 (26) (2001) 15050–15055.
[31] M. Buszczak et al., The carnegie protein trap library: a versatile tool for Drosophila developmental studies, Genetics 175 (3) (2007) 1505–1531.
[32] A.T. Quinones-Coello et al., Exploring strategies for protein trapping in Drosophila, Genetics 175 (3) (2007) 1089–1104.
[33] S. Nagarkar-Jaiswal et al., A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila, Elife (2015) 4.
[34] S. Ding et al., Efficient transposition of the *piggyBac* (PB) transposon in mammalian cells and mice, Cell 122 (3) (2005) 473–483.
[35] K.J. Venken et al., MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes, Nat. Methods 8 (9) (2011) 737–743.
[36] W.K. Huh et al., Global analysis of protein localization in budding yeast, Nature 425 (6959) (2003) 686–691.
[37] D.Q. Ding et al., Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library, Genes Cells 5 (3) (2000) 169–190.
[38] D.L. Janke et al., Interpreting a sequenced genome: toward a cosmid transgenic library of *Caenorhabditis elegans*, Genome Res. 7 (10) (1997) 974–985.
[39] Y. Le, M.J. Dobson, Stabilization of yeast artificial chromosome clones in a rad54-3 recombination-deficient host strain, Nucleic Acids Res. 25 (6) (1997) 1248–1253.
[40] G. Palmieri et al., Construction of a pilot human YAC library in a recombination-defective yeast strain, Gene 188 (2) (1997) 169–174.
[41] Y. Zhang et al., A new logic for DNA engineering using recombination in *Escherichia coli*, Nat. Genet. 20 (2) (1998) 123–128.
[42] S.K. Sharan et al., Recombineering: a homologous recombination-based method of genetic engineering, Nat. Protoc. 4 (2) (2009) 206–223.

[43] N.G. Copeland, N.A. Jenkins, D.L. Court, Recombineering: a powerful new tool for mouse functional genomics, Nat. Rev. Genet. 2 (10) (2001) 769–779.
[44] G. Testa et al., Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles, Nat. Biotechnol. 21 (4) (2003) 443–447.
[45] H. Hofemeister et al., Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis, Methods 53 (4) (2011) 437–452.
[46] G. Ciotta et al., Recombineering BAC transgenes for protein tagging, Methods 53 (2) (2011) 113–119.
[47] W.C. Skarnes et al., A conditional knockout resource for the genome-wide study of mouse gene function, Nature 474 (7351) (2011) 337–342.
[48] C. Pfander et al., A scalable pipeline for highly effective genetic modification of a malaria parasite, Nat. Methods 8 (12) (2011) 1078–1082.
[49] N.C. Hubner et al., Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions, J. Cell Biol. 189 (4) (2010) 739–754.
[50] J.R. Hutchins et al., Systematic analysis of human protein complexes identifies chromosome segregation proteins, Science 328 (5978) (2010) 593–599.
[51] Z. Maliga et al., A genomic toolkit to investigate kinesin and myosin motor function in cells, Nat. Cell Biol. 15 (3) (2013) 325–334.
[52] R.K. Ejsmont et al., Recombination-mediated genetic engineering of large genomic DNA transgenes, Methods Mol. Biol. 772 (2011) 445–458.
[53] R.K. Ejsmont et al., A toolkit for high-throughput, cross-species gene engineering in Drosophila, Nat. Methods 6 (6) (2009) 435–437.
[54] M. Sarov et al., A genome-scale resource for *in vivo* tag-based protein function exploration in *C. elegans*, Cell 150 (4) (2012) 855–866.
[55] B. Tursun et al., A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*, PLoS One 4 (3) (2009) e4625.
[56] C.T. Dolphin, I.A. Hope, Caenorhabditis elegans reporter fusion genes generated by seamless modification of large genomic DNA clones, Nucleic Acids Res. 34 (9) (2006) e72.
[57] N. Hirani et al., A simplified counter-selection recombineering protocol for creating fluorescent protein reporter constructs directly from *C. elegans* fosmid genomic clones, BMC Biotechnol. 13 (2013) 1.
[58] S. Minorikawa, M. Nakayama, Recombinase-mediated cassette exchange (RMCE) and BAC engineering via VCre/VloxP and SCre/SloxP systems, Biotechniques 50 (4) (2011) 235–246.
[59] M. Sarov et al., A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*, Nat. Methods 3 (10) (2006) 839–844.
[60] I. Poser et al., BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals, Nat. Methods 5 (5) (2008) 409–415.
[61] M. Sarov et al., A genome-wide resource for the analysis of protein localisation in Drosophila, BioRxiv (2015).
[62] F. Morra et al., FBXW7 and USP7 regulate CCDC6 turnover during the cell cycle and affect cancer drugs susceptibility in NSCLC, Oncotarget 6 (14) (2015) 12697–12709.
[63] P.R. Andersen et al., The human cap-binding complex is functionally connected to the nuclear RNA exosome, Nat. Struct. Mol. Biol. 20 (12) (2013) 1367–1376.
[64] B. Hegemann et al., Systematic phosphorylation analysis of human mitotic protein complexes, Sci. Signal. 4 (198) (2011) rs12.
[65] P.J. de Vree et al., Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping, Nat. Biotechnol. 32 (10) (2014) 1019–1025.
[66] J. Bischof et al., An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases, Proc. Natl. Acad. Sci. U.S.A. 104 (9) (2007) 3312–3317.
[67] J.M. Knapp, P. Chung, J.H. Simpson, Generating customized transgene landing sites and multi-transgene arrays in Drosophila using phiC31 integrase, Genetics 199 (4) (2015) 919–934.
[68] M. Rostovskaya et al., Transposon-mediated BAC transgenesis in human ES cells, Nucleic Acids Res. 40 (19) (2012) e150.
[69] L. Zhao, E.T. Ng, P. Koopman, A *piggyBac* transposon- and gateway-enhanced system for efficient BAC transgenesis, Dev. Dyn. 243 (9) (2014) 1086–1094.
[70] C. Frokjaer-Jensen et al., Random and targeted transgene insertion in *Caenorhabditis elegans* using a modified *Mos1* transposon, Nat. Methods 11 (5) (2014) 529–534.
[71] M. Maresca et al., Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining, Genome Res. 23 (3) (2013) 539–546.
[72] T.J. Gibson, M. Seiler, R.A. Veitia, The transience of transient overexpression, Nat. Methods 10 (8) (2013) 715–721.
[73] I.M. Cheeseman, A. Desai, A combined approach for the localization and tandem affinity purification of protein complexes from metazoans, Sci. STKE (266) (2005) pl1.
[74] S.E. Ong et al., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, Mol. Cell. Proteomics 1 (5) (2002) 376–386.
[75] S.E. Ong, M. Mann, Stable isotope labeling by amino acids in cell culture for quantitative proteomics, Methods Mol. Biol. 359 (2007) 37–52.
[76] J.W. Wong, G. Cagney, An overview of label-free quantitation methods in proteomics by mass spectrometry, Methods Mol. Biol. 604 (2010) 273–283.
[77] H. Lei et al., A widespread distribution of genomic CeMyoD binding sites revealed and cross validated by ChIP-Chip and ChIP-Seq techniques, PLoS One 5 (12) (2010) e15898.
[78] M.B. Gerstein et al., Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project, Science 330 (6012) (2010) 1775–1787.
[79] W. Niu et al., Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*, Genome Res. 21 (2) (2011) 245–254.
[80] M. Zhong et al., Genome-wide identification of binding sites defines distinct functions for *Caenorhabditis elegans* PHA-4/FOXA in development and environmental response, PLoS Genet. 6 (2) (2010) e1000848.
[81] A.C. Jungkamp et al., *In vivo* and transcriptome-wide identification of RNA binding protein target sites, Mol. Cell 44 (5) (2011) 828–840.
[82] A. Ahier, S. Jarriault, Simultaneous expression of multiple proteins under a single promoter in *Caenorhabditis elegans* via a versatile 2A-based toolkit, Genetics 196 (3) (2014) 605–613.
[83] F. Diao, B.H. White, A novel approach for directing transgene expression in Drosophila: T2A-Gal4 in-frame fusion, Genetics 190 (3) (2012) 1139–1144.
[84] M. Gonzalez et al., Generation of stable Drosophila cell lines using multicistronic vectors, Sci. Rep. 1 (2011) 75.
[85] J.F. Atkins et al., A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go), RNA 13 (6) (2007) 803–810.
[86] M.F. Underhill et al., Transient gene expression levels from multigene expression vectors, Biotechnol. Prog. 23 (2) (2007) 435–443.
[87] V. Torres et al., A bicistronic lentiviral vector based on the 1D/2A sequence of foot-and-mouth disease virus expresses proteins stoichiometrically, J. Biotechnol. 146 (3) (2010) 138–142.
[88] R. Kittler et al., RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells, Proc. Natl. Acad. Sci. U.S.A. 102 (7) (2005) 2396–2401.
[89] T. Ueda et al., Generation of functional gut-like organ from mouse induced pluripotent stem cells, Biochem. Biophys. Res. Commun. 391 (1) (2010) 38–42.
[90] A. Mathur et al., Human induced pluripotent stem cell-based microphysiological tissue models of myocardium and liver for drug development, Stem Cell Res. Ther. 4 (Suppl. 1) (2013) S14.
[91] J.L. Sterneckert, P. Reinhardt, H.R. Scholer, Investigating human disease using stem cell models, Nat. Rev. Genet. 15 (9) (2014) 625–639.
[92] R.E. Hynds, A. Giangreco, Concise review: the relevance of human stem cell-derived organoid models for epithelial translational medicine, Stem Cells 31 (3) (2013) 417–422.
[93] U. Fischer, C. Englbrecht, A. Chari, Biogenesis of spliceosomal small nuclear ribonucleoproteins, Wiley Interdiscip. Rev. RNA 2 (5) (2011) 718–731.
[94] M. Okada et al., The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres, Nat. Cell Biol. 8 (5) (2006) 446–457.