Evolving proteins in mammalian cells using somatic hypermutation

Lei Wang[1] & Roger Y Tsien[2]

[1]The Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, California 92037, USA. [2]Departments of Pharmacology and Chemistry and Biochemistry, Howard Hughes Medical Institute, University of California at San Diego, 9500 Gilman Dr., La Jolla, California 92093-0647, USA. Correspondence should be addressed to L.W. (lwang@salk.edu).

Published online 26 October 2006; doi:10.1038/nprot.2006.243

We describe a new method to mutate target genes through somatic hypermutation (SHM) and to evolve proteins directly in living mammalian cells. Target genes are expressed under the control of an inducible promoter in a B-cell line that hypermutates its immunoglobulin (Ig) V genes constitutively. Mutations can be introduced into the target gene through SHM upon transcription. Mutant genes are then expressed and selected or screened for desired properties in cells. Identified cells are subjected to another round of mutation and selection or screening. This process can be iterated easily for numerous rounds, and multiple reinforcing mutations can be accumulated to produce desirable phenotypes. This approach bypasses labor-intensive *in vitro* mutagenesis and samples a large protein sequence space. In this protocol a monomeric red fluorescent protein (mRFP1.2) was evolved in Ramos cells to afford a mutant (mPlum) with far-red emission. This method can be adapted to evolve other eukaryotic proteins and to be used in other cells able to perform SHM. For each round of evolution, it takes ~1 d to mutate the target gene, ~0.5–1 d to select or screen, and 2–4 d to propagate the cells for the next round depending on how many cells are collected.

INTRODUCTION

To evolve new protein properties, large protein diversities need to be generated and selected or screened efficiently[1]. Protein diversities are often generated from the expression of gene libraries, which can be prepared using *in vitro* methods, such as random mutagenesis and *in vitro* gene recombination[2,3]. When the selection or screen has to be done on transfected cells or organisms, mutant genes must be introduced into cells before selection and then isolated from selected cells for the next round of *in vitro* mutagenesis. The process becomes laborious when applied iteratively. Moreover, the size of the gene library that can be introduced into cells is limited by the efficiency of transformation, transfection or infection. Creating genetic diversity directly in intact living cells should mitigate these problems and thus allow larger protein sequence spaces to be sampled for desired protein properties.

This protocol provides a method for mutating target genes through SHM and for evolving proteins directly in mammalian cells (**Fig. 1**). SHM is a process through which B lymphocytes in the immune system mutate the Ig genes to produce antibodies when lymphocytes are activated by antigens[4–7]. SHM introduces point mutations into the rearranged V regions of Ig genes at a rate of $\sim 1 \times 10^{-3}$ mutations per base pair per generation[8]. SHM can be used to evolve antibodies *in vitro*[9] and to rescue single point mutations deliberately introduced into non-immunoglobulin genes[10,11]. We demonstrated that SHM could introduce multiple mutations throughout a target non-immunoglobulin gene, and beneficial mutations could be accumulated after rounds of directed evolution to generate useful phenotypes and protein properties[12].

The Burkitt lymphoma cell line Ramos, a human B-cell line that hypermutates its Ig V gene constitutively during culture, is used as the host cell[13]. The target foreign gene is expressed under the control of the inducible Tet-on promoter. After addition of the inducer doxycycline, transcription of the target gene and mutation through SHM are initiated. Mutated genes are expressed by host cells and are selected or screened for desired properties. Identified cells are then subjected to the next round of mutation and selection. The process can be iterated easily for numerous rounds until cell hits with desired properties are identified. These cells can be propagated in the absence of the inducer to avoid further mutations, and contributing mutations are subsequently identified by DNA sequencing[12].

As an example of this method, we describe the directed evolution of mRFP[14], a monomeric red fluorescent protein, through iterative SHM in Ramos cells. Fluorescence-activated cell sorting (FACS) was used to seek mRFP mutants with far-red emission as a means to improve optical imaging in thick tissues and animals. After 23 rounds of SHM and FACS, mPlum was evolved: this has an emission maximum at 649 nm and increased photostability as compared to the parental mRFP (whose emission maximum is 612 nm). Sequencing of mPlum revealed 7 substitutions out of 225

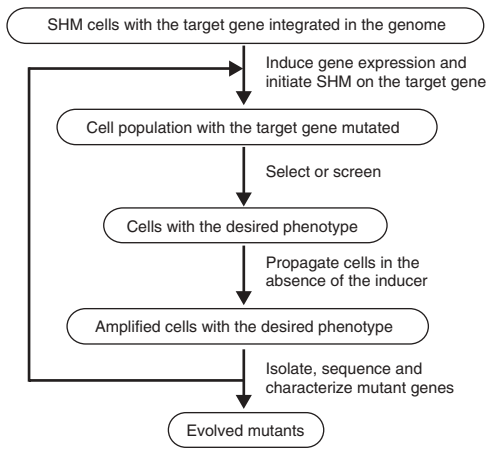

Figure 1 | Schematic illustration of the process of evolving proteins in mammalian cells via somatic hypermutation.

amino acids; saturation mutagenesis to try other amino acids at these 7 positions produced no improvements. Control experiments performed in *Escherichia coli* using conventional *in vitro* mutagenesis and evolution did not yield any mutant emitting beyond 632 nm, suggesting that SHM can sample a larger sequence space of proteins. Indeed, 15 mutations were found in 117 nucleotides within the 3′ LTR sequence downstream of mPlum (which should not be biased by the selection pressure), showing that SHM introduced a large number of mutations during the directed evolution process.

SHM can introduce multiple mutations into the target genes in each round, and these mutations will be scattered throughout the gene. Neighboring nucleotides can be mutated at the same time or within several rounds, which in some cases is critical in generating new properties and which rarely occurs with other mutagenesis methods. Together with directed evolution, SHM accumulates reinforcing beneficial mutations to produce new and desirable phenotypes that would be difficult or impossible to find by conventional mutagenesis. SHM works on single-copy integrants in the genome of mammalian cells, in contrast to the multicopy plasmids in bacteria mutated by an engineered error-prone DNA polymerase I (Ref. 15). SHM should provide a general strategy for iteratively evolving many other proteins, as long as a high-throughput selection or screen can be devised to assess the protein of interest. It should be especially valuable for eukaryotic proteins whose functions necessitate a native mammalian cell environment. Besides Ramos cells, other cell lines possessing SHM ability should be useful for this approach as well.

Like many other mutagenesis methods, SHM has a mutation bias, preferring G/C base pairs in the hot spot RGYW/WRCY (where R = A or G, Y = C or T, and W = A or T) and some specific di- and trinucleotide motifs[5,16]. However, we noted that beneficial mutations can be found at nucleotides that are relatively disfavored by SHM. For example, although T is not favored for SHM, it was mutated in the mRFP gene at two positions encoding amino acid residue 16 and 65, respectively, which were critical for creating mPlum[12]. SHM predominantly introduces point mutations, but deletions or duplications are occasionally found. A limitation of this method is that no gene recombination is involved. Although SHM can occur outside the immunoglobulin locus, the cells that evolved mPlum had integrated the gene into the Ig heavy chain locus on chromosome 14. We suspect that this location was not a coincidence, but may reflect a somewhat higher mutation rate within Ig loci, giving such integrants a competitive advantage over many generations.

MATERIALS

REAGENTS

- Ramos cells (CRL-1596, American Type Culture Collection (ATCC))
- HEK293 cells (ATCC)
- Modified RPMI 1640 medium (see REAGENT SETUP)
- Dulbecco's Modified Eagle Medium (DMEM, Invitrogen)
- Hank's Balanced Salt Solution (HBSS, Invitrogen)
- FuGENE 6 (Roche)
- Doxycycline (Sigma)
- Dimethyl sulfoxide (DMSO, Sigma)
- RNeasy mini kit (Qiagen)
- Retroviral plasmids derived from pCLNCX (Imgenex)[17]
- pBAD expression vector (Invitrogen)
- Inducible Tet-on promoter ($[tetO]_7/P_{minCMV}$, Clontech)
- Optimized reverse tet transactivator (rtTAM2.2)
- Polybrene
- Ni-NTA agarose (Qiagen)

EQUIPMENT

- Cell culture incubator and hood
- 0.45-μm syringe filters

For the directed evolution and characterization of mRFP (other instruments may be needed for the evolution of different proteins, depending on the selection and screening methods):

- BD FACSVantage SE flow cytometry system (BD Biosciences)
- Spex Fluorolog-3 fluorometer (HORIBA Jobin Yvon)
- Safire plate reader (Tecan)

REAGENT SETUP

Modified RPMI 1640 medium Mix nine parts RPMI 1640 medium containing 2 mM L-glutamine, 1.5 g per liter sodium bicarbonate, 4.5 g per liter glucose, 10 mM HEPES and 1.0 mM sodium pyruvate, with one part heat-inactivated fetal bovine serum. (Suggested by ATCC.)

PROCEDURE

Make retroviral constructs for stably expressing the target mRFP gene and the optimized reverse tet transactivator (rtTAM2.2)

1| Replace the cytomegalovirus (CMV) promoter near the *Hin*dIII restriction site in pCLNCX with the inducible Tet-on promoter to make the pCLT plasmid. Clone the mRFP1.2 gene into pCLT between the *Hin*dIII and *Cla*I sites to make pCLT-mRFP. Clone the gene for rtTAM2.2 (ref. 18) into pCLNCX to make pCL-Tet-on (**Fig. 2**). rtTAM2.2 requires doxycycline for activation of the Tet-on promoter. Although we could have made a single virus containing both the mRFP1.2 gene and the reverse tet transactivator, we chose to leave them in two separate viruses for modular flexibility. For example, if SHM were to be performed on genes longer than that of mRFP1.2, the combination with the reverse tet transactivator in one virus might exceed the size limitation of the retroviral genome.

Prepare retroviruses

2| Prepare mRFP virus by cotransfecting pCLT-mRFP and pCL-Ampho (Imgenex) into HEK293 cells grown in DMEM using FuGENE 6. Replace medium after 12–18 h.

3| Prepare Tet virus by cotransfecting pCL-Tet-on and pCL-Ampho into HEK293 cells using FuGENE 6. Replace medium after 12–18 h.

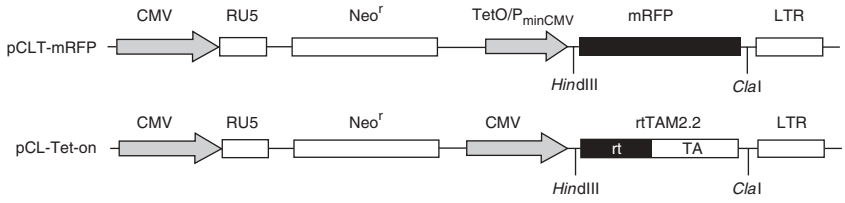

Figure 2 | Retroviral constructs for expressing the target gene mRFP (pCLT-mRFP) and the optimized reverse tet transactivator (pCL-Tet-on).

4| Harvest both viruses 24 h after the medium change and filter sterilize the virus-containing supernatant using a syringe filter to remove any cells. Both viruses can be used for infection of Ramos cells directly, or stored at –80 °C until needed.

■ **PAUSE POINT** Virus supernatant can be rapidly frozen in liquid nitrogen and stored at –80 °C for up to 1 year.

5| Determine the viral titer by following the protocol in the RetroMax instruction manual (Imgenex)[17].

Infect Ramos cells

6| To a 6-cm tissue culture dish add 1×10^6 Ramos cells, 0.2 ml mRFP virus supernatant, 0.2 ml Tet virus supernatant and 8 μg ml^{-1} polybrene. Add modified RPMI 1640 medium to a final volume of 4 ml.

▲ **CRITICAL STEP** The amount of mRFP virus supernatant should be adjusted to ensure the multiplicity of infection (MOI) is below 1, so that most infected cells will have only one copy of the mRFP gene. The amount of Tet virus supernatant should be adjusted so that most cells will be infected (MOI close to or greater than 1). Incubate the cells at 37 °C for at least 12 h.

? TROUBLESHOOTING

Establish a stable cell population harboring mRFP

7| Change medium, and add 2 μg ml^{-1} of doxycycline to induce the expression of mRFP gene. After 24 h, spin down cells and resuspend them in 2 ml HBSS.

8| Sort for red fluorescent cells. Excite cells at 568 nm, and use an emission filter of 615/40 nm emission for detection. In our initial sorting, <5% of cells became red, indicating an MOI for the mRFP virus well below 1.

9| Propagate sorted cells in the absence of doxycycline to ~10^9 cells. Add 2 μg ml^{-1} of doxycycline to induce for 24 h and sort again to enrich red fluorescent cells. After six rounds of sorting, a red fluorescent cell population can be established in which more than 96% of cells are fluorescent.

? TROUBLESHOOTING

Evolve mRFP mutants with red-shifted emission through SHM and ratio sorting

10| For each round of evolution, add 2 μg ml^{-1} of doxycycline to ~10^9 propagated Ramos cells. The induction time can be 18–36 h. To apply ratio sorting, excite Ramos cells at 568 nm, and use two emission filters (660/60 and 615/40 nm) in two different channels (**Fig. 3a**). The ratio of intensity at 660 nm to that at 615 nm is plotted against the intensity at 660 nm in real time. Collect cells with the highest ratio and sufficient intensity at 660 nm (**Fig. 3b**). Usually 10^6–10^7 cells are collected in each round, for which 10^8–10^9 cells are sorted depending on cell viability, the fluorescence ratio distribution of the population and the availability of the flow cytometer.

11| Grow the collected cells in the absence of doxycycline until the next round of induction and sorting. Freeze three vials of non-induced cells (~10^6 each) in growth medium containing 5% DMSO for emergency recovery and future analysis.

▲ **CRITICAL STEP** During multiple rounds of SHM and cell sorting, the collected cells may grow slowly or die. It is important to keep everything that comes in contact with the cells sterile, including the cell collection area of the flow cytometer. At some point, if even conditioned medium cannot rescue cell growth, re-propagate cells stored from the previous round to restart the process. Expired components of the growth medium, such as sodium pyruvate, can affect cell growth.

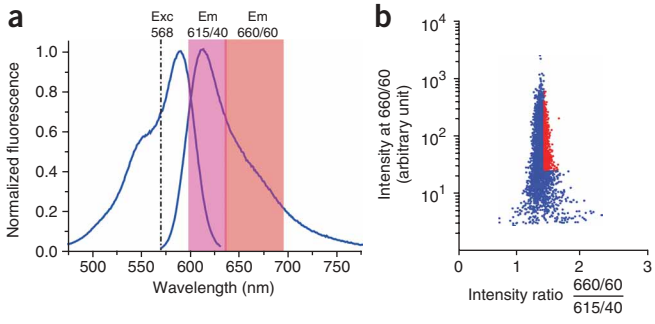

Figure 3 | Setup for fluorescence ratio sorting using the BD FACSVantage flow cytometer. (**a**) Fluorescence emission spectrum of mRFP1.2. Two emission filters centered at 660 nm and 615 nm were used in two different channels to measure the fluorescence intensities. Exc, excitation; Em, emission. (**b**) The intensity ratio (660 nm/615 nm) was plotted against the intensity at 660 nm. Cells with highest ratio and sufficient intensity at 660 nm (highlighted in red) were collected in each round. Ratio sorting uses an internal standard to exclude complications from copy number, expression level and folding efficiency.

12| To quickly monitor the red-shifting of fluorescence emission in each round, after 12 h of induction, take some cells out of the culture dish and measure the emission spectrum using a plate reader.

TABLE 1 | Primers used for identifying the integration loci.

Primer name	Sequence
LW126	5′-CACCCTGGAAACATCTGATGGTTC-3′
LW127	5′-GCCAGTCCTCCGATTGACTGAGTC-3′
LW128	5′-CGAACAGAAGCGAGAAGCGAAC-3′
LW129	5′-CGCGCTTCTGCTCCCCGAGCTC-3′
LW130	5′-TCGCCCTTGCTCACCATGGTGGC-3′
LW131	5′-GACAGCTTCAAGTAGTCGGGGATG -3′
LW132	5′-CTTCCCCGAGGGCTTCAAGTGGG-3′

Isolate and characterize evolved mutants

13| Extract mRNA from $\sim 10^6$ sorted cells using the RNeasy kit (Qiagen). mRNA can be extracted from cells just sorted. When phenotypic changes are observed, one may want to check mutants in certain rounds while continuing the evolution process. If there are not enough sorted cells, they can be amplified in the absence of doxycycline, and then induced with 0.1 μg ml^{-1} of doxycycline for 10 h before mRNA extraction. A low concentration of doxycycline was used here to obtain sufficient mRNA for extraction while minimizing further mutation.

14| Clone the mRFP mutant genes using RT-PCR with the extracted total mRNA as template. Subclone the PCR-amplified mutant mRFP genes into the pBAD vector for protein purification through Ni-NTA chromatography. For mRFP and mPlum spectroscopic and photostability measurements using a fluorometer and a fluorescence microscope, please refer to the published procedures[12,19].

15| Sequence the pBAD clones with desired protein properties.

Identify integration loci of the mutant genes

16| Optional: Follow the inverse PCR protocol to determine the integration loci of provirus in the Ramos genome[20]. The only modification is that the secondary PCR products are directly sequenced without further cloning after agarose gel electrophoresis and purification. Specifically, use *Bam*HI and *Sau*3AI to digest genomic DNA separately. For *Bam*HI cloning use primers LW131 and LW132 for the primary PCR, and use primers LW130 and LW127 for the secondary PCR (see **Table 1** for primer sequences). For *Sau*3AI cloning use primers LW128 and LW129 for the primary PCR and primers LW126 and LW127 for the secondary PCR. In the early rounds such as round 2, we found the target gene integrated in different chromosomes at multiple sites that do not contain Ig genes. By contrast, we found the mPlum gene integrated at the Ig heavy-chain locus of chromosome 14 only.

? TROUBLESHOOTING

See **Table 2** for troubleshooting information.

TABLE 2 | Troubleshooting table.

Problem	Possible reason	Solution
Step 6: After double infection and induction, no positive cells can be found.	Viral titer is too low.	Make sure polybrene is added during infection. Re-prepare retrovirus in Steps 2–5 to improve the titer. If the titer is not too low, reinfection of the same cells can help.
Step 9: Sorted cells become unhealthy and cannot be propagated.	Too few cells are collected.	Sort more cells and grow them in a small dish to increase cell density. Use conditioned medium for cell propagation.
	Cells are contaminated during sorting.	Follow strict sterile standards during cell sorting and transferral.

● TIMING

The approximate time of each step is as follows:
Steps 2–4: 36 h (determination of viral titer will take ~11 d).
Step 6: 12–24 h.
Steps 7–9: ~18 d.
Steps 10 and 11: 3–7 days per round of evolution depending on how many cells are collected and the availability of the flow cytometer.
Step 12: 10 min.
Steps 13–15: 2 d.
Step 16: 2 d.

ANTICIPATED RESULTS

One can expect SHM to introduce point mutations into the target gene at a rate of 10^{-3}–10^{-4} per base pair per generation. These mutations will be distributed throughout the target gene with a bias for G/C base pairs in RGYW/WRCY hot spots.

Mutations at nucleotides not favored by SHM can also be found. However, mutations in the final evolved mutants will not show strong bias, because the mutants are evolved under selective pressure and thus are not representative of the general mutation spectrum of SHM. It is typical for the phenotype to be unchanged for several rounds of SHM and evolution before a significant improvement appears. There are two possible reasons: the improvement may require the accumulation of multiple reinforcing mutations, or it may require a critical mutation at nucleotides that are not favored by SHM and thus take longer to obtain.

Aside from the procedure described in Steps 1–9, other methods for establishing a stable Ramos cell population with the target gene integrated in the genome can also be used. For instance, linearized DNA can be introduced into Ramos cells through electroporation, and stable cells can be selected with antibiotics. We expect that other B-cell lines that hypermutate the Ig genes should also work for this protocol. However, it should be noted that different hypermutating cell lines have different mutation spectra, and subclones of a particular cell line can differ in overall mutation rate. Indeed, we found that foreign genes expressed in chicken ΔXRCC2-DT40 cells also show mutations with features of SHM. One advantage of using DT40 cells is that they grow much faster than Ramos cells (with an approximately fivefold shorter doubling time) and thus should significantly shorten the directed evolution time. Moreover, the SHM pattern of DT40 cells involves more transversion mutations and thus may complement that of Ramos cells[21].

In the process of evolving mRFP1.2, we found in early rounds of evolution that SHM could mutate the target gene integrated into a chromosome containing no Ig loci. However, mPlum was isolated from the Ig heavy-chain locus by round 23. These results show that SHM can mutate exogenous genes integrated at many loci in the genome[11], but suggest that the mutation rate may be higher at the Ig loci. Therefore, desired properties requiring multiple mutations are more likely generated from a clone with the target gene integrated at Ig loci. Amplifying such a clone might have taken up multiple early rounds of selection. In future applications, it would be more efficient to direct the target gene to such loci using gene-replacement vectors to jump-start the evolution[22].

ACKNOWLEDGMENTS This work was supported by the Howard Hughes Medical Institute, National Institutes of Health Grant NS27177, and US Department of Energy Grant DE-FG03-01ER63276 to R.Y.T. L.W. was supported by the Damon Runyon Cancer Research Foundation (Merck Fellow, Grant DRG-1767-03), the Searle Scholars Program, and the Beckman Young Investigators Program.

COMPETING INTERESTS STATEMENT The authors declare that they have no competing financial interests.

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

1. Lin, H. & Cornish, V.W. Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Edn. Engl.* **41**, 4402–4425 (2002).
2. Bloom, J.D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 447–452 (2005).
3. Minshull, J. & Stemmer, W.P. Protein evolution by molecular breeding. *Curr. Opin. Chem. Biol.* **3**, 284–290 (1999).
4. Papavasiliou, F.N. & Schatz, D.G. Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. *Cell* **109** (Suppl.): S35–S44 (2002).
5. Martin, A. & Scharff, M.D. AID and mismatch repair in antibody diversification. *Nat. Rev. Immunol.* **2**, 605–614 (2002).
6. Maizels, N. Immunoglobulin gene diversification. *Annu. Rev. Genet.* **39**, 23–46 (2005).
7. Neuberger, M.S. *et al.* Somatic hypermutation at A.T pairs: polymerase error versus dUTP incorporation. *Nat. Rev. Immunol.* **5**, 171–178 (2005).
8. Rajewsky, K., Forster, I. & Cumano, A. Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* **238**, 1088–1094 (1987).
9. Cumbers, S.J. *et al.* Generation and iterative affinity maturation of antibodies in vitro using hypermutating B-cell lines. *Nat. Biotechnol.* **20**, 1129–1134 (2002).
10. Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M. & Olsson, C. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol.* **166**, 5051–5057 (2001).
11. Wang, C.L., Harper, R.A. & Wabl, M. Genome-wide somatic hypermutation. *Proc. Natl. Acad. Sci. USA* **101**, 7352–7356 (2004).
12. Wang, L., Jackson, W.C., Steinbach, P.A. & Tsien, R.Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. USA* **101**, 16745–16749 (2004).
13. Sale, J.E. & Neuberger, M.S. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* **9**, 859–869 (1998).
14. Campbell, R.E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. USA* **99**, 7877–7882 (2002).
15. Camps, M., Naukkarinen, J., Johnson, B.P. & Loeb, L.A. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. USA* **100**, 9727–9732 (2003).
16. Larijani, M., Frieder, D., Basit, W. & Martin, A. The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in ung$^{-/-}$msh2$^{-/-}$ mice. *Immunogenetics* **56**, 840–845 (2005).
17. Naviaux, R.K., Costanzi, E., Haas, M. & Verma, I.M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J. Virol.* **70**, 5701–5705 (1996).
18. Go, W.Y. & Ho, S.N. Optimization and direct comparison of the dimerizer and reverse tet transcriptional control systems. *J. Gene Med.* **4**, 258–270 (2002).
19. Baird, G.S., Zacharias, D.A. & Tsien, R.Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. USA* **97**, 11984–11989 (2000).
20. Li, J. *et al.* Leukaemia disease genes: large-scale cloning and pathway predictions. *Nat. Genet.* **23**, 348–353 (1999).
21. Sale, J.E., Calandrini, D.M., Takata, M., Takeda, S. & Neuberger, M.S. Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature* **412**, 921–926 (2001).
22. Kardinal, C., Selmayr, M. & Mocikat, R. Genetic stability of gene targeted immunoglobulin loci. I. Heavy chain isotype exchange induced by a universal gene replacement vector. *Immunology* **89**, 309–315 (1996).