Published in final edited form as:

Anal Biochem. 2016 February 1; 494: 55–67. doi:10.1016/j.ab.2015.11.002.

Definitive Localization of Intracellular Proteins: Novel Approach Using CRISPR-Cas9 Genome Editing, with Glucose 6-Phosphate Dehydrogenase as a Model

Netanya Y. Spencer^{2,3,*}, Ziyang Yan¹, Le Cong⁴, Yulong Zhang¹, John F. Engelhardt¹, and Robert C. Stanton^{2,3,5}

¹Department of Anatomy and Cell Biology, The University of Iowa, Iowa City, IA

²Joslin Diabetes Center, Boston, MA

³Harvard Medical School, Boston, MA

⁴Broad Institute of MIT and Harvard, Cambridge, MA

⁵Beth Israel Deaconess Medical Center, Boston, MA

Abstract

Studies to determine subcellular localization and translocation of proteins are important because subcellular localization of proteins affects every aspect of cellular function. Such studies frequently utilize mutagenesis to alter amino acid sequences hypothesized to constitute subcellular localization signals. These studies often utilize fluorescent protein tags to facilitate live cell imaging. These methods are excellent for studies of monomeric proteins, but for multimeric proteins, these methods are unable to rule out artifacts from native protein subunits already present in the cells. That is, native monomers might direct the localization of fluorescent proteins with their localization signals obliterated. We have developed a method to rule out such artifacts, and we use glucose 6-phosphate dehydrogenase (G6PD) as a model to demonstrate the method's utility. Because G6PD is capable of homodimerization, we employed a novel approach to remove interference from native G6PD. We produced a G6PD knockout somatic (hepatic) cell line using CRISPR-Cas9 mediated genome engineering. Transfection of G6PD knockout cells with G6PD fluorescent mutant proteins demonstrated that the major subcellular localization sequences of G6PD are within the N-terminal portion of the protein. This approach sets a new gold standard for similar studies of subcellular localization signals in all homodimerization-capable proteins.

*Correspondence to: netanya.spencer@joslin.harvard.edu.

Author Contributions: NYS designed and performed all the experiments and wrote the manuscript. ZY and LC assisted with design of experiments. YZ assisted with performance of experiments. JFE, ZY, LC, and RCS assisted with writing and editing the manuscript. The authors declare that they have no competing interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Glucose 6-Phosphate Dehydrogenase; CRISPR-Cas9 Genome Engineering; Nuclear Localization Signal; Knockout Cell Lines; Dimer; Multimer

Introduction

Protein subcellular localization affects every aspect of cell biology research. For monomeric proteins, identification of subcellular localization signals is fairly direct and usually involves mutating specific sequences followed by techniques such as fractionation, immunofluorescence staining, and tagging mutant proteins with fluorescent protein tags [1]. For multimeric proteins, identification of subcellular localization signals may not be reliable as follows. If the protein of interest already exists in the cells and is capable of forming dimers or larger multimers, individual subunits of the native protein may dimerize (or form larger multimers) with the subunit containing the mutant localization signal and thus may direct localization of the multimer to the native site even if the mutation has in fact obliterated the subcellular localization signal in one subunit (see Figure 1). The best way this potential artifact has been somewhat avoided in the past was by mutating the amino acids known to cause dimerization simultaneously with mutating a hypothesized subcellular localization signal [2]. That approach is practical only if the protein under study has very few amino acids at the dimer interface, as mutations of a large number of amino acids in the interface may inadvertently affect subcellular localization signals, thus confounding results. Additionally, for many proteins the structure is not available and thus the identities of the amino acids at the dimer interface are unknown. Another approach would be to introduce the mutant fluorescently tagged protein into a cell line that naturally does not express the native protein. This approach has the obvious disadvantage of not allowing observation of the protein in cell types of real physiological interest.

We have now developed a method to avoid this artifact, using CRISPR-Cas9 genome editing. Our novel approach is to knock out the protein of interest completely before reintroducing it with a mutated subcellular localization signal and a fluorescent tag. (Figure 2). In the past, this approach has not been possible because rapid, reliable, complete knockout of genes in cell lines was not achievable. Thus, our study shows for the first time that this issue has been resolved in a definitive manner. For studying subcellular localization signals in multimeric universally expressed or essential proteins this approach has been impossible until now because cell lines without the protein of interest do not naturally exist.

One such universally expressed protein is glucose 6-phosphate dehydrogenase (G6PD). G6PD forms a homodimer in the active state and sometimes is also found as a tetramer (Figure 1). The dimer interface is complex, involving many amino acids. Thus G6PD is useful as a model protein for addressing the question of how to identify subcellular localization signals in ubiquitous, complex multimeric proteins. Additionally, G6PD is a primary hub controlling cell biology, and is essential to life [3]. Therefore the regulation of its subcellular localization is crucial to understand. In the nucleus in particular, G6PD appears to be a key player in signaling; we have recently observed in hepatocytes that there

is nuclear localization of a low level of G6PD in addition to the high level which has long been known to be in the cytoplasm [4]. This study is the first ever to search for subcellular localization signals within G6PD.

In this study we demonstrate three main points. First, we show that complete knockout (*via* clustered regularly interspaced short palindromic repeats (CRISPR-Cas9) genome engineering) of any multimeric protein is essential to complete a study of its subcellular localization in the context of fluorescent protein tags in live cells. Thus, CRISPR-Cas9-based technologies are the preferred approach for such studies. Second, we show that G6PD's N-terminal segment contains this protein's primary subcellular localization signals. Third, we demonstrate cell-specific variability in response to complete G6PD knockout; this has both the potential to lead to new approaches to discovery of cell specific therapies. Our approach has applications for all intracellular localization studies of multimeric proteins.

Materials and Methods

Cell culture

Tib73 cells [5] (an immortalized mouse liver cell line, from ATCC) and HeLa cells (a human cervical cancer cell line) were grown in 37 °C incubators with 5% CO₂ in DMEM (Sigma D5796), 10% fetal bovine serum, plus 1% penicillin-streptomycin.

Cloning of p.G6PD-mCherry

B6SJLF1 mice (from Jackson Laboratories) were housed and euthanized with carbon dioxide in accordance with protocols approved by the University of Iowa Institutional Animal Care and Use Committee. Liver tissue was dissected from mice immediately after euthanasia. For total RNA isolation, small samples of liver (~50 ug each) were immediately snap-frozen in liquid nitrogen. Frozen samples were rapidly homogenized in a solution of 25 mM sodium citrate, 5 M guanidine thiocyanate, 10 mM EDTA, 0.86% beta-mercaptoethanol. Homogenates were centrifuged for 5 minutes at 12,000 RPM at 4°C. Supernatants were layered over a solution of 5.7 M CsCl, 100 mM EDTA. Samples were centrifuged at 35,000 rpm in a Beckman SW-55 rotor overnight at 20 °C. The total RNA pellet was washed in H₂O and ethanol precipitated. Life Technologies's Micro-FastTrack 2.0 kit was used to isolate mRNA from total RNA. The mRNA was reverse transcribed to cDNA using ABI's High Capacity RT kit. PCR to amplify G6PD was performed (see below). All PCR was done using Life Technologies's Accuprime PFX Supermix following the manufacturer's instructions. All restriction enzymes and their appropriate buffers were from New England Biolabs.

To amplify G6PD from cDNA, the following primers and strategies were used. The 5' primer (Primer A in Table 1) added an XhoI site and a Kozak site immediately upstream of the G6PD coding sequence. The 3' primer (Primer B, Table 1) eliminated the stop codon and added a HindIII site immediately downstream of G6PD. Then, to produce the fusion protein of G6PD with mCherry, mCherry was PCR amplified from pmCherry-N1 (Clontech). The 5' primer in this reaction (Primer C, Table 1) added both a HindIII site and a linker sequence, and eliminated the mCherry start codon to avoid any translation of mCherry alone without

the fused upstream G6PD. Because of this primer's design, the amino acid sequence of the linker for in-frame fusion of G6PD and mCherry is EASGGSGA. Meanwhile, the 3' primer for this reaction (Primer D, Table 1) included mCherry's stop codon and added a NotI site. The restriction sites produced by the respective PCR reactions were used to ligate the G6PD and mCherry sequences into the XhoI and NotI sites of pacAd5CMV K-NpA (obtained from the Vector Core of the University of Iowa). The resultant plasmid was named p.G6PD-mCherry (pGmCh).

Cloning of G6PD-mCherry truncation mutants

Using a similar strategy as outlined above, mouse G6PD amino acids 1–361 and amino acids 362–515 were individually cloned as fusion proteins with mCherry, including the same linker sequence as for full-length G6PD-mCherry. For PCR, the template was the p.G6PD-mCherry plasmid. To amplify the sequence encoding AA1-361, the forward PCR primer was the same forward primer used above for amplifying G6PD from mouse cDNA (Primer A, Table 1). The reverse primer (Primer E, Table 1) places a HindIII site immediately after the codon for AA361 in the G6PD sequence. To amplify the sequence encoding AA362-515 of G6PD, the forward PCR primer (Primer F, Table 1) places an XhoI site, and a Kozak site including ATG, just before AA362 of G6PD. The reverse primer was the same reverse primer used to amplify full-length G6PD from mouse cDNA (Primer B, Table 1). The full-length G6PD sequence was cut out of p.G6PD-mCherry with XhoI and HindIII, and was replaced by standard cloning procedures either with the amplified PCR product encoding AA1-361 (to produce p.G1-361-mCherry) or with the amplified PCR product encoding AA362-515 (to produce p.G362-515-mCherry).

Cloning of pHM830-G6PD and its truncation mutants

Plasmid pHM830 was obtained from Addgene (Addgene plasmid 20702), where it has been deposited by Dr. Thomas Stamminger, whose kind contribution of this plasmid we appreciate greatly [1]. We amplified this plasmid as directed by Addgene. This plasmid contains the gene for green fluorescent protein (GFP) without a stop codon, followed by a multiple cloning site (MCS), followed by the LacZ gene. Cloning a protein-coding sequence in-frame into the MCS (while not introducing any stop codons) results in a fusion protein with GFP at the N-terminus and β -galactosidase (β gal) at the C-terminus, with the newly introduced protein in between. We used the following primers to amplify either all or part of the G6PD gene to insert into this MCS, using pGmCh as the PCR template. To amplify the entire G6PD gene the forward primer was Primer G in Table 1 and the reverse primer was Primer H in Table 1. To amplify the region of the G6PD gene coding for amino acids (aa's) 1–361, the same forward primer (Primer G) was used but the reverse primer was Primer I in Table 1. To amplify the region of the G6PD gene coding for aa's 362–515, the forward primer was Primer J in Table 1, and the reverse primer was again Primer H. PCR was performed using Accuprime PFX Supermix (Life Technologies) following the manufacturer's cycling protocols. PCR products were purified with a Qiagen QIAquick PCR Purification Kit. These PCR products and pHM830 were digested with AflII and NheI (both from New England Biolabs). Enzyme-digested PCR products and plasmid were gel-purified, ligated (T4 DNA Ligase from New England Biolabs), and amplified in DH5 α *E.coli* (Life Technologies) to produce plasmids containing genes coding for the following three fusion

proteins: GFP-G6PD- β gal, GFP-(G6PD AA1-361)- β gal, and GFP-(G6PD AA362-515)- β gal. These plasmids were named “pHM-G6PD”, “pHM-G1-361”, and “pHM-G362-515” respectively. DNA sequencing confirmed that these clones were obtained successfully.

Cloning of Plasmids for Genome Engineering

Plasmid pX330 was obtained from Addgene (Addgene plasmid 42230), where it has been deposited by Dr. Feng Zhang, whose kind contribution of this plasmid we appreciate greatly [1]. We amplified this plasmid as directed by Addgene. This plasmid codes for all necessary elements of a system of clustered regularly interspaced short palindromic repeats (CRISPR), including the Cas9 gene and two closely-packed BbsI enzymatic restriction sites for cloning a short sequence to code for a guide RNA. Guide sequences were designed using the CRISPR design tool provided online by the Zhang laboratory at <http://crispr.genome-engineering.org/> and <http://crispr.mit.edu/>. Two target sequences rather than just one for each species (mouse and human) were chosen to try to double the chance of knocking out the G6PD gene. Since the genomic sequences are not identical between mouse and human at the targeted regions, two targeting guide sequences were designed for mouse G6PD and two other sequences were designed for human G6PD. These sequences are listed in Table 2. These oligonucleotides were designed according to the following five principles [7, 10, 13, 14]: (1) The oligonucleotides do not contain the protospacer adjacent motif (PAM) sequence, in order to avoid Cas9 cleaving the plasmid after transfection. (2) In the forward oligonucleotide, following the CACC (which is present for cloning into pX330's double BbsI site), an extra “G” is introduced which did not exist at the end of any of the targets in the human or mouse G6PD sequences. This additional “G” is required in order to allow guide RNA to be transcribed in mammalian cells. (3) The target sequence follows this added “G”. (4) The reverse oligonucleotide complements the forward oligonucleotide, with an additional sticky end for cloning into pX330's double BbsI site. (5) All target sequences were selected as far upstream (toward the 5' end) as possible within the first protein coding exon of each species (Exon 2 in both species) in order to try to introduce mutations that (because of frame-shift, non-sense mediated decay of mRNA, and other cellular mechanisms) would not allow any endogenous protein expression at all, preferably not even a short N-terminal segment of the protein.

Forward and reverse oligonucleotides with 5' phosphate modifications were ordered from Integrated DNA Technologies. Each set of forward and reverse oligonucleotides were annealed by heating to 95°C and then slowly cooling. Plasmid pX330 was digested with BbsI (NEB), gel-purified (Qiagen QIAquick Gel Extraction kit), and ligated with the annealed oligonucleotides using T4 DNA Ligase (NEB). DH5 α *E. coli* were transformed with the ligated samples following the manufacturer's directions and plated on ampicillin agar plates. After overnight growth, clones were picked and amplified in ampicillin LB broth. Clones were mini-prepped (Qiagen QIAprep Spin Miniprep Kit) and digested with appropriate enzymes to determine which clones contained insert. Following this selection process plasmids were sequenced by the University of Iowa DNA Sequencing Core and plasmids with correct sequences were kept for further amplification and purification by large-scale cesium chloride gradient centrifugation following standard protocols [15]. The

plasmids for targeting human G6PD were named pHu1 and pHu2. The plasmids for targeting mouse G6PD were named pMs1 and pMs1A.

Transfection

For transfection experiments, cells were transfected with the plasmids of interest using Lipofectamine LTX with Plus reagent. TIB73 cells grew equally well on both plastic and glass surfaces. Therefore, TIB73 cells were continuously propagated on plastic dishes, but for high-resolution imaging experiments they were split at 1:4 to 1:10 ratios and plated into glass coverslip-bottom CultureWells (Grace Bio-Labs 112358/112359).

Transfection of CRISPR-Cas9 plasmids

We transfected 70% confluent Tib73 and HeLa cells with CRISPR plasmids (pHu1, pHu2, pMs1, or pMs1A) using Lipofectamine LTX with Plus reagent in 6-well tissue-culture plates. For selection and screening purposes, we simultaneously co-transfected pTRE2-pur (Clontech) into the same wells at a ratio of 1:4 (i.e., 1ug pTRE2-pur per 4 ug CRISPR plasmid). In cases where we co-transfected both species-specific CRISPR plasmids simultaneously (e.g., both Hu1 and Hu2 sequences, or else both Ms1 and Ms1A sequences) we used a ratio of 2:2:1 (2 ug of each CRISPR plasmid per 1ug pTRE2-pur). Approximately 24 hours after transfection, we added puromycin (Gibco) to each well (5 ug/mL for both HeLa cells and Tib73 cells as determined by dose-response experiments). The following day, untransfected control cells of either cell type treated with the same dose of puromycin were >99% dead and floating off the plates. In transfected wells of HeLa cells, >95% confluent cells were observed (indicating that ~95% of the HeLa cells in each well were expressing puromycin resistance protein and presumably Cas9; the other ~5% of the HeLa cells were killed by the puromycin, presumably because they remained untransfected). In transfected wells of Tib73 cells, <20% confluent Tib73 cells were observed (indicating that <20% of the Tib73 cells were expressing puromycin resistance protein and presumably Cas9; the other ~80% of the Tib73 cells were killed by puromycin, again presumably because they remained untransfected). These results were exactly as expected based on previous experiments on transfection efficiencies of HeLa and Tib73 cells. The cells were trypsinized at approximately 24 hours post-puromycin treatment (48 hours post-transfection), pelleted, resuspended, diluted in growth medium, and counted. Cells were plated into 96-well plates at a density of approximately one cell per well. Such limiting dilutions were performed several times to ensure that genuinely clonal populations were obtained. Control cells that had not been transfected or treated with puromycin were taken through all cell passages and dilutions concurrently with the CRISPR-treated cells. No puromycin was included for the remainder of the experiments because stable puromycin resistance and stable Cas9 expression were not needed to achieve the goal of a cell line with G6PD entirely knocked out.

Growing CRISPR-Cas9-treated clonal cells

Clonal populations of CRISPR-Cas9-treated HeLa cells and Tib73 cells were amplified by the following cell culture protocol. Cells forming confluent patches covering at least 30% of the surface of a single well in a 96 well plate were trypsinized for 5 minutes at 37 °C with 30 uL trypsin (Gibco 0.25%). Then growth medium (DMEM, 10% FBS, 1% penicillin-

streptomycin) was added to a total volume of 200 μ L, well contents were pipetted 3–4 times to break up cell clumps, and then cells from 1 well were divided amongst 2–3 wells of fresh 96 well plates. After cells attached to the new plates overnight, the medium was removed and replaced with fresh medium. Cells were continuously split this way every time they reached confluence, in order to allow growth of larger volumes, with cells eventually being grown in 6-well and larger plates. However, CRISPR-treated Tib73 cells always grew faster and with a more healthy appearance in smaller wells; plates with diameters larger than the diameter of 6-well plate wells were almost completely unable to sustain growth of CRISPR-treated Tib73 cells. The reasons for this were outside the scope of this study, but in general, to keep CRISPR-treated Tib73 cells alive it was necessary to use small wells, no matter the seeding density or split ratio of the cells. At high plating densities these cells could be coaxed to grow in 60 mm and 100 mm dishes, but they appeared unhealthy (largely rounding up and dying) in such plates, and grew at a much slower rate than in smaller wells. This was in contrast to control Tib73 cells that were not CRISPR-treated; such wild-type cells grew readily in all sizes of tissue-culture plates.

Screening for G6PD knockout cells

Three independent methods of screening for knockouts were used: (a) 96-well plate format G6PD enzyme activity assays, (b) sequencing of amplified genomic DNA, and (c) Western blotting.

(a) Screening for KO cells by 96-well plate format G6PD enzyme activity

assays—96-well plate format G6PD enzyme activity assays were performed using the Life Technologies Vybrant Cytotoxicity Assay Kit (V-23111). This kit is designed to determine if chemical treatment of cells results in cytotoxicity and cell lysis. However, the kit works by detecting G6PD released due to lysis of cells that have been subjected to such toxic treatments. The kit conveniently contains a positive control lysis buffer which lyses all cells, releasing all the cells' G6PD. The released G6PD produces NADPH from the substrates NADP⁺ and G6P (included with the kit reagents); the NADPH then undergoes further chemistry with the kit reagents to produce the red fluorescent compound resorufin. Therefore, we simply used the kit's lysis buffer in all the wells, releasing all G6PD that might have been present in the cells, followed the kit protocol otherwise, and scanned for red fluorescence to determine total G6PD activity levels in each well. Red fluorescence in the wells of 96-well plates was detected with a Typhoon 8610 scanner using these settings: Band Pass filter 580, Green Laser 532, PMT voltage 400.

(b) Screening for KO cells by sequencing of amplified genomic DNA—

Sequencing of amplified genomic DNA was accomplished as follows: Several wells of CRISPR-treated or control cells grown in 96 well plates were lysed as follows. Medium was removed from wells. Adherent cells in the wells were treated with 10 μ L of PCR Lysis Buffer (50mM KCl, 1.5 mM MgCl₂, 10 mM Tris pH 8.5, 0.5% NP40, 0.5% Tween-20, 400 μ g/mL proteinase K); then plates were tightly sealed and incubated for 45 minutes at 56 °C and then 15 minutes at 98 °C. The samples were used as templates in PCR reactions with Accuprime PFX Supermix (Life Technologies). For HeLa samples, the forward primer was Primer K in Table 1, the reverse primer was Primer L in Table 1.; for Tib73 samples, the

forward primer was Primer M in Table 1, and the reverse primer was Primer N in Table 1. PCR products were analyzed on a 1.5% agarose gel and then cloned into a TOPO vector using the Zero Blunt® TOPO® PCR Cloning Kit (Life Technologies catalog # K2875-20). Cloned samples were amplified in *E.coli*, plasmid minipreps were prepared (Qiagen QIAprep Spin Miniprep Kit), and samples were sequenced. Resulting control and CRISPR-treated DNA sequences were compared to determine the presence of mutations.

(c) Screening for KO cells by Western Blotting—For Western blotting, samples were scraped from 6-well plates in 50 uL RIPA buffer ((0.15 M NaCl, 50 mM Tris pH 7.4, 1% deoxycholic acid, 1% Triton-X100, 0.1% SDS, 1mM EDTA, supplemented with 1× Proteoblock protease inhibitor cocktail). Samples (50ug protein) were run on standard denaturing SDS-PAGE, and proteins were transferred to nitrocellulose (NC) at 400 mA for 80 minutes at room temperature or 12 V overnight at 4 °C. NC membranes were blocked in Odyssey Blocking Buffer (Licor, catalog # 927-40100) and then the primary antibody, rabbit anti-G6PD (Bethyl catalog # A300-404A), was used at 1:1000. Goat anti-actin (Santa Cruz sc-1616) was used as a loading control at 1:100. Secondary infrared-dye conjugated antibodies were used at 1:10,000–20,000 dilutions. NC membranes were scanned on an Odyssey infrared scanner. For NEMO (IKK γ) Western blotting, we used Abcam rabbit monoclonal antibody # ab178872 at a 1:5000 dilution.

Microscopy

For live cell microscopy experiments, cells were cultured in Gibco DMEM without Phenol Red (Life Technologies catalog #21063), with 10% fetal bovine serum and 1% penicillin/streptomycin. All cells were cultured in a 37 °C humidified incubator (5% CO₂). Cells were grown in coverslip-bottom Culturewells (Grace BioLabs). Confocal fluorescence microscopy was performed on a Zeiss confocal laser scanning microscope and a Leica confocal spinning disk microscope, both with oil-immersion 63× objectives. Data analysis was performed with Metamorph software. Both microscopes were used for both live and fixed-cell imaging. For red fluorescence imaging of mCherry fusion proteins, cells were imaged live at 24–48 hours post-transfection, or fixed in methanol and counterstained with Vectashield DAPI prior to imaging. For nuclear counterstain in live cells, Hoechst 33342 (Life Technologies) was used according to the manufacturer's instructions. Identical results were obtained whether live or methanol-fixed cells were imaged, both in terms of red fluorescence intensity as well as in terms of subcellular localization. For the quantitation experiments in Figure 9, red protein-expressing cells were manually counted at 63× with an oil immersion objective on the Zeiss laser scanning microscope while moving the cell culture dishes around the microscope stage to find red cells. Red cells were categorized as expressing red protein in both nucleus and cytosol as opposed to cytosol alone (no cells expressed red in nucleus alone). More than 100 fluorescing cells in each group were observed.

Results

A. G6PD as a Model for Studies of Subcellular Localization Signals

Subcellular Localization of G6PD: Rationale and Background for the Study—

Many intracellular proteins form homodimers. To address the issue of native monomers dimerizing with mutant fluorescent monomers (see Figure 1), we chose to use G6PD, which forms homodimers as well as tetramers, as a model for studying subcellular localization. G6PD has long been considered almost exclusively a cytoplasmic protein [16–19]. In our previous study [4], we showed G6PD localization in the nucleus based on antibody studies and, indirectly, by enzymatic activity (i.e., NOX4 activity dependent on G6PD activity in nuclei). These results were strongly suggestive of G6PD nuclear localization but might be criticized by some investigators as artifacts. Thus, nuclear localization of G6PD should be confirmed by alternate methods as well. In this study we therefore also show G6PD nuclear localization by fluorescent protein tagging in live cells. Additionally, our earlier findings [4] led us to hypothesize that G6PD nuclear localization is highly regulated and may occur in certain cells only under certain physiological conditions. If this is correct, G6PD location should be observed at different sites within different cells.

G6PD tagged with mCherry (GmCh) localizes either in cytoplasm only, or in both nucleus and cytoplasm—

The subcellular localization of G6PD-mCherry (GmCh) was investigated following transfection of pGmCh into TIB73 cells by confocal fluorescence microscopy. As shown in Figure 3A, GmCh expressed by Lipofectamine-mediated transfection of pGmCh localizes almost exclusively to the cytoplasm of some cells, but to both the nucleus and cytoplasm of other cells in the same culture dish. Thus, GmCh can localize to the cytoplasm, or to both cytoplasm and nucleus. The cells are therefore not uniform; this may be due to varying physiological states of the cells. The precise physiological states resulting in altered GmCh locations are currently being investigated in our laboratory. The locations observed for GmCh are highly likely to be reflective of wild-type native G6PD localization because (1) fluorescent tags only rarely change a protein's native subcellular localization [20], and (2) the large majority of cells demonstrated at least some cytoplasmic localization of GmCh as predicted based on previous literature describing G6PD localization in the cytoplasm [16, 21].

Nuclear translocation of G6PD may be due to nuclear localization sequences

—Transit of a protein molecule into the nucleus is regulated in some cases by a nuclear localization signal (NLS) within the protein's amino acid sequence, and in other cases by binding of the protein to another protein containing such an NLS [22, 23]. Similarly, transit out of the nucleus can depend on a nuclear export signal (NES) or binding to another protein containing an NES. Because G6PD-mCherry was found in both the nucleus and cytoplasm of many cells, we hypothesized that G6PD would have a nuclear localization signal, a nuclear export signal, or both.

The G6PD amino acid (AA) sequence contains many candidate NLS and NES motifs. Therefore, the plasmids p.G1-361-mCherry and p.G362-515-mCherry were produced to allow expression of fluorescently-tagged N-terminal or C-terminal portions of G6PD. This

would narrow down the possibilities for where functional NLS and/or NES sequence(s) might be found within G6PD. These plasmids were transfected into Tib73 cells and confocal fluorescence microscopy was performed to determine subcellular localization of fluorescence. As shown in Figure 3B, G1-361-mCherry (the protein that contains AA's 1–361 of G6PD fused to mCherry) localized in cells in a pattern essentially indistinguishable from full-length GmCh. However, G362-515-mCherry (which contains only G6PD AA's 362–515 fused to mCherry) frequently localized outside the nucleus as large clumps or aggregates although in other cells it demonstrated localization similar to GmCh (Figure 3C). These results showed that the nuclear localization signals, nuclear export signals, or other cellular localization signals in the G6PD sequence were likely within the first 361 amino acids of the protein.

However, with these data alone, this conclusion could not be definitively proven because of the presence of native G6PD protein in the cells (see Figure 1). Wild-type G6PD forms a homodimer and sometimes a tetramer (in which two dimers dimerize together) [24]. Thus, a mutant G6PD molecule (such as the fluorescently-tagged truncation mutants described above) might be able to dimerize with a native G6PD monomer already present in a cell (Figure 1). If such a heterodimer could form, the native wild-type half of this heterodimer might direct the subcellular localization of the entire dimer. Similarly, if a tetramer were to form in a cell containing both native G6PD and mutant G6PD, if even one of the four subunits of the tetramer were the native sequence, the one native molecule might direct the subcellular localization of the entire tetramer. Then, fluorescence microscopy analyses would fail to indicate sites where the mutants would go on their own but would only reflect the native wild-type G6PD localization.

An alternative G6PD fluorescent fusion protein gives results identical to G6PD-mCherry—In addition to possible confounding by native G6PD, the nuclear localization of the mCherry-tagged proteins could conceivably be artifactual due to the mCherry tag itself [20, 25]. To rule out this possibility, we cloned another plasmid expressing the fusion protein, GFP-G6PD- β -galactosidase. Transfection with this plasmid (named pHM-G6PD) showed that this protein also localized to both cytoplasm and nucleus in some cells, but only to cytoplasm in other cells (Figure 4). Indeed, co-transfection of plasmid expressing G6PD-mCherry (pGmCh) with pHM-G6PD showed 100% overlap of green and red fluorescence (Figure 4). Because the sequences of β -galactosidase, mCherry, and GFP are very different from each other, the probability that all these tags would direct the tagged G6PD molecules to the same non-native sites is very low. Thus, this result indicates that the tagged G6PD molecules were not directed to non-native sites by the tags and instead are very likely localized to the authentic sites of native G6PD for cell-specific physiologic reasons.

B. CRISPR Genome Engineering: An Essential Tool to Validate Subcellular Localization Studies

Confounding effects of dimerization cannot be ruled out without knocking out native protein—As described above, the results in Figure 1 did not conclusively prove that the N-terminal segment of G6PD contains any subcellular localization signals. It was

equally conceivable that native G6PD monomers present in the cells were causing confounding localization (see Figure 1). To rule out this possibility, it was necessary to use cells that contained no native G6PD and then transfect the fluorescent mutants into these cells.

Some G6PD knockout cell lines can be produced by genome engineering—

Genome engineering technology (Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 methodology) was chosen to produce G6PD knockout cells. In addition to Tib73 cells, HeLa cells (a human cervical cancer cell line) were also simultaneously used. CRISPR/Cas9 methods were used to produce double-strand breaks within the first protein-coding exon (Exon 2) of the mouse G6PD gene in Tib73 cells (Figure 5A) or the human G6PD gene in HeLa cells (Figure 5B). G6PD was readily knocked out of Tib73 cells but could not be knocked out of HeLa cells. This result will be discussed further below.

To utilize our method to investigate subcellular localization of dimerizing proteins, it is absolutely essential to verify that complete knockout of the protein of interest has been successful. Cells might repair Cas9-induced double-stranded breaks with in-frame mutations. Thus, slightly shortened (internally truncated) versions of the protein of interest might be produced, instead of real knockouts. G6PD knockout (from Tib73 cells only) was therefore confirmed by a series of G6PD activity assays, limiting dilutions for DNA sequence analyses, and Western blots for G6PD as described below.

G6PD activity assays identify G6PD knockout cells—As an initial screen for knockout cells, if the protein of interest is an enzyme whose activity is amenable to measurement, activity assays can be used. As shown in Figure 3C, wells of 96-well plates containing different populations of CRISPR-treated Tib73 or HeLa cells were assayed for G6PD activity. No wells of CRISPR-treated HeLa cells with lower G6PD activity compared to untreated cells were found. However, in non-clonal populations of Tib73 cells that had been CRISPR-treated, more than 50% of the wells showed greatly decreased G6PD activity compared to untreated cells (Figure 5C). Therefore these populations of cells were chosen for further limiting dilution cloning and amplification in order to obtain monoclonal cell populations.

Limiting Dilutions and DNA sequencing identify G6PD knockout cells—As mentioned above, for our approach to be valid, complete knockout must be confirmed. Thus, although enzymatic activity assays give strong indication of decreased protein levels, sequencing must be performed. However, far fewer samples need to be sequenced after enzymatic activity screening. Thus, after activity assays, our second method for screening for G6PD knockout cells was the technique of limiting dilutions followed by DNA sequencing. Tib73 cells from populations in which G6PD activity was decreased were diluted and replated in 96-well plates to achieve less than one cell per well, and then were grown until about 70% confluent. These numerous clones were split to continue culturing, and aliquots of these cells were taken for DNA sequence analysis. The goal of the CRISPR treatment had been to introduce a frame shift and thus an early stop codon in the G6PD gene, as close as possible to the start codon of the gene. This would ensure that no G6PD

mRNA or protein could be synthesized, because the transcription machinery would encounter the stop codon almost immediately after the start codon. Additionally, remaining transcripts with premature stop codons would also be subjected to degradation by nonsense-mediated decay or other cellular surveillance pathways, ensuring there would be little if any production of functional protein. Sanger sequencing was therefore performed on DNA samples isolated from numerous HeLa and Tib73 clones.

In some of the surviving CRISPR-treated Tib73 cells, DNA sequencing demonstrated that the G6PD gene indeed had frame shifts and early stop codon formation (Figure 6A, B). Two clonal populations with such stop codons were expanded in culture and used for further experiments as described below. These clones were named F4 and 1A-5 (sequences shown in Figure 6A, B). As expected, untransfected control Tib73 cells retained native G6PD sequence, enzymatic activity, and immunodetectability by Western blots.

Western blotting confirms that G6PD knockout cells contain no G6PD—A confirmatory step to demonstrate complete knockout, although perhaps not absolutely necessary, is to demonstrate total loss of protein after gene knockout. Thus, after identifying G6PD knockout cells by DNA sequencing, our third confirmatory screen for the knockout of G6PD was performing Western blotting for G6PD protein (Figure 7A). Western blotting demonstrated that G6PD knockout cells retained no G6PD protein expression, providing protein-level confirmation that complete knockout of G6PD had been achieved (lanes marked F4 and 1A-5 in Figure 7A). On the other hand, control cells that had not undergone CRISPR transfection demonstrated easily detectable G6PD levels (lane marked T in Figure 7A, indicating wild-type Tib73 cells). Many nonclonal cell populations were found (lanes marked N in Figure 7A). In some of these nonclonal populations, G6PD protein was undetectable by Western blot (lanes marked N₁ and N₄ in Figure 7A). Thus, even without limiting dilution to isolate a clonal population, these populations of cells had completely abolished G6PD expression. Other non-clonal populations (lanes marked N₂ and N₃ in Figure 7A) demonstrated varying levels of G6PD still present, but not as high as in wild-type cells. This indicated that CRISPR worked in at least some of the cells in every population. It is therefore likely that by further diluting, a much larger number of clonal G6PD knockout cell lines could have been isolated, but as this was not necessary for the purposes of this study, clonal dilutions were completed only in the case of the cell lines named F4 and 1A-5. Cell line 1A-5 has been retained as a backup. Further experiments were done almost entirely with F4 because it was the first knockout cell line isolated.

G6PD knockout cells have near-normal levels of a protein encoded by an overlapping gene—To utilize our method to investigate subcellular localization of dimerizing proteins, it is absolutely essential to make sure that major changes have not occurred in other closely associated genes as a result of the CRISPR-Cas9 treatment. In the case of G6PD, if genes closely associated with G6PD were adversely affected along with G6PD during the genetic engineering process, any biological effects that might be observed in these cells could be attributed to effects on those genes rather than to the loss of G6PD. In particular, there is one very closely associated gene, named NFκB Essential Modulator (NEMO). As shown in Figure 7B, in both mice and humans, the G6PD gene, located on the

X-chromosome, partially overlaps the NEMO gene, also known as IKK γ or IKBKG [26, 27]. This unusual overlap of two genes could indicate that regulation of these two genes occurs in a co-dependent manner [28]. Therefore a major concern was that the knockout of G6PD might abrogate or severely impair NEMO expression, not because of loss of G6PD protein's potential regulatory effects on NEMO expression, but because the CRISPR-Cas9-induced mutations themselves might prevent NEMO transcription. To investigate this possibility, Western blotting was used to assay for NEMO expression (Figure 7C), and this demonstrated that G6PD knockout cells still contain NEMO. The NEMO levels trended toward decreased levels in G6PD knockout cells. Whether this decrease in expression was the result of decreased transcription directly due to the CRISPR-induced mutation (for example, the mutation might adversely affect binding of transcription factors to the NEMO promoter), or due to the complete loss of G6PD and potential direct or indirect effects of G6PD on NEMO levels, cannot be ascertained from this result. However, it is clear that biological effects observed in these cells are more likely to be due to the complete knockout of G6PD than to these cells' trend toward decreased NEMO expression.

C. Somatic cells vary in their ability to survive CRISPR-Cas9-mediated gene knockout

Comparison of HeLa and Tib73 cells—A possible pitfall of our method is that CRISPR-Cas9-treated cells might not survive knockout of essential genes. Thus it may sometimes be necessary to use more than one cell line to achieve the goal of gene knockout. As mentioned above, HeLa cells were simultaneously CRISPR-treated along with Tib73 cells, and DNA from resulting clones was sequenced. Although many CRISPR-induced mutations in the G6PD gene were observed in HeLa cells, none of these consisted of frame-shifting (and thus knockout) mutations. In fact, every clonal HeLa cell line that manifested a CRISPR-mutated G6PD DNA sequence had a multiple of three bases deleted. The most common mutations in HeLa cells were a deletion of either 12 or 18 bases. In every screened HeLa cell population, the HeLa G6PD gene, although mutable by CRISPR, remained in frame in all surviving HeLa cells; there were no frame-shifts (i.e., knockouts) found in HeLa cells. Thus, mutant forms of the protein were produced in these cells, missing a few amino acids near the N-terminus, but probably still retaining some enzymatic activity, allowing the cells to survive. It is therefore likely that complete removal of G6PD from HeLa cells is cytotoxic. If any HeLa cells underwent frame-shift (total knockout) mutations upon CRISPR treatment, they must have died quickly, as they did not survive long enough to be identified in the limiting dilution process.

However, unlike HeLa cells, Tib73 cells proved hardier and more able to withstand G6PD knockout. Importantly, the fact that Tib73 cells survived G6PD knockout while HeLa cells did not may indicate that Tib73 cells were able to upregulate another NADPH source or other compensatory pathway(s) that certain aggressive cancer cells (e.g., HeLa) cannot upregulate. Such compensatory proteins and pathways represent possible future therapeutic targets, for instance in cancer cells that do not survive knockout of G6PD.

E. G6PD's N-terminal segment contains its primary subcellular localization signals

Transfection of mutants into knockout cells reveals that the N-terminal segment is crucial—Once a knockout cell line has been produced, we demonstrated that

we could use these cells for transfection of fluorescently-tagged G6PD truncation mutants in the complete absence of native G6PD. WT and G6PD KO cells were transfected with the plasmids pGmCh, pG1-361-mCherry, or pG362-515-mCherry (Figure 8). After transfection with pGmCh (which has the full length of G6PD fused to mCherry), approximately the same distribution pattern of red fluorescence was found in KO as in WT cells. That is to say, the majority of cells expressed red fluorescence primarily or exclusively in the cytoplasm but a few expressed red fluorescence also in the nucleus (Figure 8A). Similarly, after transfection with pG1-361-mCherry (which expresses only the N-terminal 361 amino acids of G6PD followed by mCherry) approximately the same fluorescence patterns were again observed (Figure 8B); that is, a few cells had some degree of red fluorescence in the nucleus but most of the red fluorescence was in the cytosol. This was observed equally in both WT and KO cells. However, after transfection with pG362-515-mCherry (which expresses only the C-terminal amino acids from 362-515 followed by mCherry) the red fluorescence was aggregated entirely outside the nucleus (Figure 8C). The fluorescent protein could not distribute normally. The fluorescent aggregates were not specifically identifiable as being in the cytosol, endoplasmic reticulum, mitochondria, golgi, or other sites.

These experiments demonstrated that removing the N-terminal end of the G6PD protein adversely affects its localization to the nucleus, whereas removing the C-terminal end of the protein does not. Indeed, counts of cells with red fluorescence in cytoplasm, nucleus, or both showed more cells containing at least some red nuclear fluorescence after transfection with G1-361-mCh than with full length GmCh (Figure 9). Therefore the most likely part of the protein to contain subcellular (and in particular, nuclear) localization signals is within the first 361 amino acids (aa's).

As mentioned above, protein translocation into the nucleus can depend on NLS motifs within a protein, or on binding to another protein containing an NLS. Although we have not fully answered the question of which of these possibilities is the case with G6PD, our data definitively lead to the conclusion that within the first 361 aa's is the sequence that either itself serves as an NLS, or that allows G6PD to bind to another protein that enters the nucleus. Thus G6PD translocation to the nucleus is not a random event but occurs due to specific regulation exerted by functional signals in the protein, which will be investigated further in future studies.

Discussion

Novelty of our approach

We have demonstrated a new method to avoid a potential artifact in subcellular localization studies. Without this method, subcellular localization signal identification in multimeric proteins could potentially be confounded when the native protein still exists in the cells under study (Figure 1). Past studies have not been able to address this question definitively. Now, this problem can be absolutely ruled out by first using CRISPR-Cas9-mediated genome editing to knock out the protein of interest before using mutant fluorescently tagged proteins to search for subcellular localization signals (Figure 2). To use this approach, it is important to show that the gene is totally knocked out (not just internally truncated or partly

mutated), and, if relevant to the gene of interest, it is also important to show that any closely associated or overlapping genes are not severely affected.

Knockout cells are necessary for confirmatory subcellular localization studies in live cells

Initially, we transfected fluorescently tagged wild-type and mutant G6PD into wild-type cells. With this approach, we observed very similar patterns of localization of these mutants as compared to the full-length G6PD-mCherry protein (Figure 3). This observation indicated that even with very large mutations in the transfected gene, the resultant fluorescent protein still localized to some extent in native-like subcellular locations when transfected into mammalian cells.

One possible explanation for this observation was that our mutants of G6PD, expressed in cells already containing native G6PD, might be forced to localize to the native site(s) by heterodimerization [24] with the native G6PD in the cells (Figure 1). The only way to rule out such heterodimerization would be to start with complete G6PD knockout cells and transfect them with our mutant fluorescent proteins.

If any heterodimerization of wild-type (native) and mutant fluorescent G6PD took place in the experiments shown in Figure 1, the localization did not change greatly compared to the experiments shown in Figure 8 (KO cells), as demonstrated by the quantitation in Figure 9. This observation could have several interpretations. For example, perhaps heterodimerization did not occur to any appreciable extent in WT cells. Another explanation could be that heterodimers are not forced to the native site by the WT protein subunit of the dimer. Alternatively, localization may depend on dimerization (perhaps an NLS-like signal forms due to dimerization). Another possibility might be that heterodimers cannot effectively enter the nucleus. Although these questions about WT-mutant heterodimers remain unanswered, the experiments with knockout cells in Figures 8–9 remove all possibility of heterodimers. Thus the localization seen in these cells can be due only to the sequences of the mutant transfected proteins, not to any native proteins.

Our method is also applicable to situations where one protein is known to form a heterodimer with another protein but does not homodimerize. That is, if one of the two proteins is hypothesized to contain a subcellular localization signal, the other protein (which may very well contain many known or unknown subcellular localization sequences itself) could be knocked out by CRISPR-Cas9 techniques. Meanwhile in a separate plasmid, the gene for the protein that has not been knocked out, (the one with the hypothesized localization sequence) can be cloned, mutated or truncated as desired, fluorescently tagged, and introduced into the cells. In such a case, even with the wild-type protein present, the fluorescently-tagged mutant protein could still be used to verify the identity of the amino acid sequence that directs localization.

Applicability to other cell lines: potential pitfalls and solutions

These studies with G6PD are a new gold-standard approach to intracellular localization studies of any proteins that dimerize, of which there are many [29]. However, each protein must be considered individually in the context of the cell type being studied. Indeed, as we demonstrated, HeLa cells could not survive G6PD knockout. The discovery that some

somatic cell lines (i.e., Tib73) can tolerate complete G6PD knockout whereas others (e.g., HeLa) cannot is very important. In our study, HeLa cells were initially used as a so-called “positive control.” HeLa cells were chosen for this because they are readily transfected (usually >70% transfection efficiency). On the other hand, Tib73 cells usually display only around 5% transfection efficiency. Thus, we hypothesized that HeLa cells would more readily be transformed to a G6PD knockout cell line than would Tib73 cells. Notably however, the opposite conclusion was eventually reached, as G6PD was successfully knocked out of Tib73 cells, while HeLa cells did not survive knockout of G6PD. This demonstrated that this method (CRISPR-mediated gene knockout for subcellular localization studies), although in one sense a new gold standard, cannot be applied to all cell lines of physiological importance. However, we recommend that this method be applied as often as possible in studies of this kind, to obtain confirmatory results of subcellular localization sequence identities.

Although a limitation of our knockout approach in some cases is the inability of some cell lines to tolerate specific gene knockout, similar CRISPR-Cas9 techniques could in theory still in many such cases be used as follows. First, specific mutations in genes could be introduced to delete or alter specific amino acid sequences suspected of being nuclear localization signals (or other subcellular localization signals). Additionally, in the same experimental cells, a fluorescent tag could be added to the protein *via* CRISPR and homologous recombination. Although this approach would require significant effort, particularly in the case of genes on somatic chromosomes, it could conceivably overcome the cell death issues encountered with knocking out essential genes in such studies. Such an approach would essentially be the equivalent of our approach of knocking out the gene and then transfecting the cell with the mutant, fluorescently tagged version.

Closely associated/overlapping genes: another potential pitfall and solution

Expression of proteins encoded by genes that are closely associated with (i.e., overlapping) the gene of interest is another possible concern, as demonstrated in our case by the fact that the NEMO gene overlaps the G6PD gene. For the method we propose to be useful in subcellular localization studies, every gene should be examined for the presence of such other genes, and expression of those other proteins should be investigated. In our case, NEMO levels only trended toward decreased expression when G6PD was completely knocked out (Figure 7C). If NEMO had been completely knocked out, a different section of G6PD could have been targeted by CRISPR, because CRISPR is very precise. Similarly, other genes targeted in this way for future subcellular localization studies of other dimerizing proteins can be targeted at a variety of points if necessary, to avoid major effects on overlapping genes.

N-terminal region of G6PD contains main subcellular localization signals

Any protein that enters the nucleus in a regulated fashion must either itself contain an NLS or must bind another protein that has such a sequence. Although our study does not answer the question of which of these is the case with G6PD, it does demonstrate that either the NLS or the major site involved in binding the other unidentified protein is within the first 361 amino acids of G6PD.

The localization of fusion proteins of fluorescent proteins with other proteins is widely used in the literature. However, fluorescent proteins may alter a protein's subcellular localization. Indeed, it has been recently demonstrated that fusion of mCherry, as well as numerous other monomeric fluorescent proteins, to bacterial proteins may affect bacterial subcellular localization by causing formation of foci within bacteria [30]. In our mammalian cells, however, we found that different tags on G6PD protein molecules all co-localized, indicating that G6PD, and not the tag, was responsible for the protein's localization (Figure 4).

The C-terminal 154 amino acids (AA's) of G6PD molecule are mostly dispensable for normal G6PD protein localization. We conclude this based on the fact that p.G1-361-mCherry localized mostly normally. Therefore future studies will be needed to dissect the first 361 AA's of G6PD to determine the precise AA's that constitute its localization signals or binding sites. Furthermore, these studies do not absolutely rule out the possibility of weaker subcellular localization signals within the C-terminal 154 AA's, and more detailed studies of this portion of G6PD are also being pursued. Additionally, it is conceivable that G6PD subcellular localization sequences may be affected by posttranslational modifications. Such modifications are also under study. In future studies, fluorescent tagging will allow quantitation of nuclear G6PD translocation and accumulation under different physiological stimuli (although this is outside the scope of the present study).

Overall Significance

This study is the first ever, studying any multimeric protein, in which any investigators have ever eliminated the possibility of their mutant tagged protein molecules dimerizing with native protein molecules and being directed to the native localization site rather than the mutation-permitted site. Such studies were essentially impossible before the advent of genome engineering, but with CRISPR the cost has come down to the point where making targeted knockout cell lines for such purposes is realistic and achievable.

This paper is therefore the first to demonstrate that CRISPR technology can be used for the purpose of studying subcellular localization of homodimerizing proteins. Many such proteins' subcellular transport and localization can more readily be accurately studied now by use of fluorescently tagged mutant proteins introduced into genetically-targeted knockout cells.

Additionally, G6PD is an essential enzyme and plays major roles in health and disease. In this study a novel G6PD knockout liver cell line has been produced and used to demonstrate that G6PD subcellular localization signals, and in particular, nuclear localization signals, are more likely to reside in the N-terminal 361 amino acids of the protein than the C-terminal portion of the protein. These results will therefore also inform future studies of G6PD regulation of cellular processes.

Conclusions in Brief

In this study, we have demonstrated three major points. First, we have found that CRISPR-Cas9 technology can be used as an essential and helpful new tool for subcellular and nuclear localization studies of homodimerizing proteins, which further opens up the possibility of

applying this methodology in various similar scenarios. Second, we have found that the N-terminal segment of G6PD is the site of its potential subcellular localization signals. Third, we demonstrate that different somatic cell types vary in their response to complete loss of G6PD. Taken together, these studies advance both our knowledge of techniques for investigating subcellular translocation of G6PD and many other homodimerizing proteins as well as open new avenues for exploration of therapeutic approaches to G6PD-influenced diseases such as diabetes, cancer, and many others.

Acknowledgments

Although we did not make use of HeLa-specific genome data, we did use a HeLa cell line in this project, and we wish to acknowledge our sincere gratitude to Henrietta Lacks, now deceased, and her surviving family members for their contributions to biomedical research.

Funding Statement: This work was supported by the University of Iowa Roy J. Carver Chair in Molecular Medicine, and the National Institute of Diabetes and Digestive and Kidney Diseases Grants R01 DK051315 (to JFE) and K01 DK083429 (to NYS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abbreviations

Cas9	CRISPR-associated protein 9
CRISPR	clustered regularly interspaced palindromic repeats
G6PD	glucose 6-phosphate dehydrogenase
GmCh	G6PD-mCherry fusion protein
KO	knockout
LMW	low molecular weight
MCS	multiple cloning site
NEMO	NF κ B essential modulator
NES	nuclear export signal
NLS	nuclear localization signal
WT	wild-type

References

1. Sorg G, Stamminger T. Mapping of nuclear localization signals by simultaneous fusion to green fluorescent protein and to beta-galactosidase. *BioTechniques*. 1999; 26:858–862. [PubMed: 10337476]
2. Koike M, Ikuta T, Miyasaka T, Shiomi T. Ku80 can translocate to the nucleus independent of the translocation of Ku70 using its own nuclear localization signal. *Oncogene*. 1999; 18:7495–7505. [PubMed: 10602508]
3. Stanton RC. Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *Iubmb Life*. 2012; 64:362–369. [PubMed: 22431005]
4. Spencer NY, Yan ZY, Boudreau RL, Zhang YL, Luo MH, Li QA, Tian X, Shah AM, Davisson RL, Davidson B, Banfi B, Engelhardt JF. Control of Hepatic Nuclear Superoxide Production by Glucose 6-Phosphate Dehydrogenase and NADPH Oxidase-4. *J Biol Chem*. 2011; 286:8977–8987. [PubMed: 21212270]

5. Patek PQ, Collins JL, Cohn M. Transformed cell lines susceptible or resistant to in vivo surveillance against tumorigenesis. *Nature*. 1978; 276:510–511. [PubMed: 723934]
6. Hsu PD, Lander ES, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*. 2014; 157:1262–1278. [PubMed: 24906146]
7. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*. 2013; 31:827–832.
8. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013; 154:1380–1389. [PubMed: 23992846]
9. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*. 2013; 8:2281–2308. [PubMed: 24157548]
10. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014; 343:84–87. [PubMed: 24336571]
11. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S, Jaenisch R, Zhang F, Sharp PA. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology*. 2014
12. Pyzocha NK, Ran FA, Hsu PD, Zhang F. RNA-guided genome editing of mammalian cells. *Methods in molecular biology*. 2014; 1114:269–277. [PubMed: 24557909]
13. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; 339:819–823. [PubMed: 23287718]
14. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*. 2013; 153:910–918. [PubMed: 23643243]
15. Sambrook, J.; Fritsch, EF.; Maniatis, T. *Molecular cloning : a laboratory manual*. 2. Cold Spring Harbor Laboratory; Cold Spring Harbor, N.Y: 1989.
16. Matsubara S, Kato T, Oshikawa K, Yamada T, Takayama T, Koike T, Watanabe T, Izumi A, Sato I. Glucose-6-phosphate dehydrogenase in rat lung alveolar epithelial cells. An ultrastructural enzyme-cytochemical study. *European journal of histochemistry : EJH*. 2002; 46:243–248. [PubMed: 12472120]
17. Matsubara S, Matsubara D, Ishibashi T, Takizawa T. Demonstration of glucose-6-phosphate dehydrogenase in rat Kupffer cells by a newly-developed ultrastructural enzyme-cytochemistry. *European journal of histochemistry : EJH*. 2003; 47:173–176. [PubMed: 12777215]
18. Lizotte E, Tremblay A, Allen BG, Fiset C. Isolation and characterization of subcellular protein fractions from mouse heart. *Anal Biochem*. 2005; 345:47–54. [PubMed: 16125124]
19. Mulligan-Kehoe MJ, Russo A. Inhibition of cytoplasmic antigen, glucose- 6-phosphate dehydrogenase, by VH-CH1, an intracellular Fd fragment antibody derived from a semisynthetic Fd fragment phage display library. *Journal of molecular biology*. 1999; 289:41–55. [PubMed: 10339404]
20. Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, Uhlen M, Simpson JC, Lundberg E. Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. *Nat Meth*. 2013; 10:315–323.
21. Matsubara S. Simultaneous demonstration of acid phosphatase and glucose-6-phosphate dehydrogenase in mouse hepatocytes. A novel electron-microscopic dual staining enzyme-cytochemistry. *European journal of histochemistry : EJH*. 2002; 46:237–242. [PubMed: 12472119]
22. Miyawaki A. Proteins on the move: insights gained from fluorescent protein technologies. *Nat Rev Mol Cell Biol*. 2011; 12:656–668. [PubMed: 21941275]
23. Wang R, Brattain MG. The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. *FEBS letters*. 2007; 581:3164–3170. [PubMed: 17588566]

24. Au SW, Gover S, Lam VM, Adams MJ. Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency, Structure (London, England : 1993). 2000; 8:293–303.
25. Seibel NM, Eljouni J, Nalaskowski MM, Hampe W. Nuclear localization of enhanced green fluorescent protein homomultimers. Analytical Biochemistry. 2007; 368:95–99. [PubMed: 17586454]
26. Courtois, G.; Israël, A. NF- κ B Defects in Humans: The NEMO/Incontinentia Pigmenti Connection. 2000.
27. Smahi A, Courtois G, Vabres P, Yamaoka S, Heuertz S, Munnich A, Israel A, Heiss NS, Klauck SM, Kioschis P, Wiemann S, Poustka A, Esposito T, Bardaro T, Gianfrancesco F, Ciccociola A, D'Urso M, Woffendin H, Jakins T, Donnai D, Stewart H, Kenwrick SJ, Aradhya S, Yamagata T, Levy M, Lewis RA, Nelson DL. Genomic rearrangement in NEMO impairs NF-kappaB activation and is a cause of incontinentia pigmenti. The International Incontinentia Pigmenti (IP) Consortium. Nature. 2000; 405:466–472. [PubMed: 10839543]
28. Galgoczy P, Rosenthal A, Platzer M. Human-mouse comparative sequence analysis of the NEMO gene reveals an alternative promoter within the neighboring G6PD gene. Gene. 2001; 271:93–98. [PubMed: 11410370]
29. Ispolatov I, Yuryev A, Mazo I, Maslov S. Binding properties and evolution of homodimers in protein–protein interaction networks. Nucleic Acids Research. 2005; 33:3629–3635. [PubMed: 15983135]
30. Landgraf D, Okumus B, Chien P, Baker TA, Paulsson J. Segregation of molecules at cell division reveals native protein localization. Nature methods. 2012; 9:480–482. [PubMed: 22484850]

Figure 1. Homodimerization may affect subcellular localization studies

This diagram shows three possible scenarios. The top section shows what is observed when wild-type proteins are fluorescently tagged with mCherry and introduced into wild-type cells. Abbreviations: GmCh, G6PD-mCherry; NLS, Nuclear Localization Sequence. The middle section of this figure shows the confounding effect that might conceivably happen if fluorescently tagged mutant proteins are introduced into wild-type cells. In this scenario, it is impossible to rule out dimerization of fluorescent mutant monomers with wild-type monomers, the latter of which might or might not direct the localization of the dimer. The bottom section of this figure shows the solution to this problem: the use of knockout cells in which no native non-fluorescent wild-type monomers exist. When fluorescent mutants are introduced to such cells they may remain monomers (left) or may dimerize (right) but in either case the localization will not be confounded by native wild-type protein. Thus the localization of the fluorescent mutants can help determine accurately the identity of the amino acids that make up the NLS.

Figure 2. Overview of method to rule out dimerization artifacts *via* complete knockout of the gene of interest

CRISPR-Cas9 is used to produce knockout cells. Meanwhile, plasmids containing constructs of fluorescent proteins fused to the wild-type or a mutant gene of interest are cloned. Fluorescent proteins are expressed in knockout cells. Finally, protein localization is observed by fluorescent confocal microscopy.

Figure 3. G6PD-mCherry fusion protein has strong fluorescence in some nuclei but not in others, demonstrating that individual cells in the same culture dish regulate G6PD nuclear localization differently

Tib73 cells transfected with (A) pGmCh, (B) pG1-361-mCherry or (C) pG362-515-mCherry were imaged live by laser scanning confocal microscopy. Blue: Hoechst 33342, Red: G6PD-mCherry and mutant constructs.

Figure 4. GFP-G6PD-β-galactosidase and G6PD-mCherry colocalize, showing that tags do not affect localization

Both constructs were simultaneously transfected into Tib73 cells. After 48 hours live cells were stained with Hoechst and imaged on a Zeiss confocal microscope.

Figure 6. DNA Sequencing of clonal G6PD knockout Tib73 cells demonstrates complete G6PD knockout

(A, B) *Two mouse G6PD sequences with frameshift mutations leading to early stop codons.*
 (A) shows sequencing of G6PD from clonal cell line 1A-5. (B) shows sequencing from clonal cell line F4.

Figure 7. G6PD knockout cells have no G6PD protein expression but still contain NEMO, demonstrating that this overlapping gene was largely unaffected by G6PD knockout

(A) Western blotting for G6PD in control, non-clonal mixed populations, and clonal populations. N, non-clonal mixed populations. Populations N₁ and N₄ demonstrate that some nonclonal populations of CRISPR-treated Tib73 cells express no detectable G6PD. Populations N₂ and N₃ demonstrate that other nonclonal cell populations still express detectable although somewhat decreased levels of G6PD compared to control Tib73 cells (T) that were not transfected with CRISPR plasmids. F4, clonal cell line named F4. 1A-5, clonal cell line named 1A-5. These clonal populations of cells demonstrate no G6PD expression. (B) *Overlap of G6PD and NEMO (IKBKG) genes*, modified from http://useast.ensembl.org/Homo_sapiens/Location/View?db=core;g=ENSG00000160211;r=X:153742697-153792696. (C) Western blotting for NEMO/IKKγ/IKBKG. NEMO, also known as IKKγ or IKBKG, tended to be decreased after G6PD was knocked out but was still easy to detect, and not entirely abolished.

Figure 8. Localization of G6PD-mCherry expressed in G6PD KO cells demonstrates that the N-terminal region of G6PD contains subcellular localization-determining sequences

(A) *Full length G6PD-mCherry*. Representative cells demonstrating expression localized only in the cytosol (left) or in both cytosol and nucleus (right) of KO cells. (B) *G6PD amino acids 1–361-mCherry*. Representative cells demonstrating expression localized mainly but not absolutely exclusively in the cytosol (left) or somewhat more evenly in both cytosol and nucleus (right). (C) *G6PD amino acids 362–515-mCherry*. Representative cells demonstrating nearly exclusive retention of the fluorescent mutant protein outside of the nucleus. These results show that the nuclear localization signal is in G6PD amino acids 1–361.

Figure 9. Quantitation of fluorescent locations confirms and validates results obtained previously with WT cells

This graph shows the percentage of cells showing both nuclear and cytosolic fluorescence, demonstrating that the fluorescent protein containing G6PD amino acids 362–515 has decreased expression in the nucleus in both WT and KO cells, verifying that the nuclear-directing amino acids are located within amino acids 1–361. The KO cell study proves that the study done in the WT cells was an accurate finding.

Table 1

Primers other than those used for pX330 cloning

Primers used to clone fluorescent fusions of G6PD or truncation mutants

Primer A

5'-AAA-CTC-GAG-TCG-CCA-CCA-TGG-CAG-AGC-AGG-TGG-CCC-TGA-GCC-GGA-3'

Primer B

5'-CCG-GAA-GCT-TCG-AGC-TTG-TGA-GGG-TTC-ACC-CAC-TTG-T-3'

Primer C

5'-CGA-AGC-TTC-CGG-TGG-TTC-AGG-TGC-CGT-GAG-CAA-GGG-CGA-GGA-GGA-T-3'

Primer D

5'-TGG-CGG-CCG-CTA-CTT-GTA-CAG-CTC-GTC-CAT-GCC-GCC-GGT-3'

Primer E

5'-ATC-AAG-CTT-CGG-CTT-TGC-CAC-AGC-GCA-GGA-TGA-A-3'

Primer F

5'-TTT-CTC-GAG-TCG-CCA-CCA-TGC-TGA-ATG-AAC-GCA-AAG-CTG-AAG-TG-3'

Primer G

5'-CTG-GCC-CGC-TAG-CGC-AGA-GCA-GGT-GGC-CCT-GAG-CCG-GA-3'

Primer H

5'-CGG-GCC-CTC-TAG-AGA-GCT-TGT-GAG-GGT-TCA-CCC-ACT-TGT-3'

Primer I

5'-CGG-GCC-CCT-TAA-GGG-CTT-TGC-CAC-AGC-GCA-GGA-TGA-A-3'

Primer J

5'-CGG-GCC-CGC-TAG-CCT-GAA-TGA-ACG-CAA-AGC-TGA-AGT-G-3'

Primers used for PCR for screening knockout cells

Primer K

5'-GGG-AGG-AGG-AGC-TCA-ACT-TA-3'

Primer L

5'-GGG-CTA-ACT-TCT-CAA-TGC-TCT-C-3'

Primer M

5'-ATG-CTC-TTT-CGT-TCT-CCT-ACC-3'

Primer N

5'-GGA-CAC-ACA-GGT-GGA-GTA-TAT-G-3'

Table 2

Oligonucleotide sequences cloned into pX330

Human1-fw
5'-CAC-CGA-TAT-GTG-TGT-ATC-CGA-CTG-A-3'
Human1-rv
5'-AAA-CTC-AGT-CGG-ATA-CAC-ACA-TAT-C-3'
Human2-fw
5'-CA-CCG-CCG-CAC-ACC-TGG-GTC-CGG-CT-3'
Human2-rv
5'-AA-ACA-GCC-GGA-CCC-AGG-TGT-GCG-GC-3'
Mouse1-fw
5'-CA-CCG-CCT-GGT-ACA-ACT-CTT-CCC-TC-3'
Mouse1-rv
5'-AA-ACG-AGG-GAA-GAG-TTG-TAC-CAG-GC-3'
Mouse1A-fw
5'-CAC-CGC-ACA-CAC-CTG-GGT-CCG-GCT-C-3'
Mouse1A-rv
5'-AAA-CGA-GCC-GGA-CCC-AGG-TGT-GTG-C-3'