A systematic approach to quantitative Western blot analysis

Lakshmi Pillai-Kastoori*, Amy R. Schutz-Geschwender, Jeff A. Harford

LI-COR Biosciences, 4647 Superior Street, Lincoln, NE, 68504, USA

ABSTRACT

Attaining true quantitative data from WB requires that all the players involved in the procedure are quality controlled including the user. Appropriate protein extraction method, electrophoresis, and transfer of proteins, immunodetection of blotted protein by antibodies, and the ultimate step of imaging and analyzing the data is nothing short of a symphony. Like with any other technology in life-sciences research, Western blotting can produce erroneous and irreproducible data. We provide a systematic approach to generate quantitative data from Western blot experiments that incorporates critical validation steps to identify and minimize sources of error and variability throughout the Western blot process.

1. Introduction

Western blotting is a simple yet powerful procedure to investigate the presence, relative abundance, relative mass, presence of post-translational modifications (PTM) as well as to study protein-protein interactions. These applications of Western blot provide valuable information in both academic research, diagnostic and therapeutic testing. This workhorse method, first described by Towbin et al. [1] and Burnette [2], relies on the specific interaction of antibodies with target antigens present in the sample mixture. After sample proteins are separated on a protein gel and transferred to the membrane, primary and secondary antibodies are used to bind and visualize the target protein. The Western blot was originally intended to provide a yes/no answer about the presence of the target protein in a protein sample. This qualitative method confirms the presence of target bands by simple visual assessment [3–8]. With the surge in the field of systems biology and the need to understand complex biological systems, Western Blot is no longer limited to generating qualitative data.

The reproducibility of Western blot analysis and other immunosassays is an ongoing source of concern in the scientific community [6,9–11]. The immunoblotting process involves a complex series of interdependent steps that are influenced by subjective choices and user expertise [6,11–14]. Variations in experimental design, methodology, and technique can be substantial sources of error. This variability is particularly troubling for quantitative analysis of Western blot data, where an error could lead to misinterpretation of data [6,14,15]. Seemingly minor or insignificant differences in the reagents and parameters used for each experiment may have a surprisingly strong influence on the results [6,9,13,14,16–18]. Careful experimental design and well-characterized methods are essential to avoid common pitfalls and determine which assay parameters are most critical for reproducible results [6,11–13].

However, reproducibility does not guarantee accuracy; an assay may produce precise data that are inaccurate [13,16,19]. Accuracy and precision are important but distinct aspects of reproducibility. Accuracy indicates how closely the measured values represent the true value of the target or analyte in a sample. Do these values reflect a genuine change or difference in the experimental sample? Precision describes the repeatability, variation, and error of assay measurements. If the test is repeated many times, how similar are the results? Quantitative analysis of Western blot data should strive for both precision and accuracy, to help ensure that the reported observations convey meaningful information about the experimental samples [14,19–22].

This review aims at providing a step-by-step breakdown of the approach to perform and gather quantifiable data from Western blots (Fig. 1).

2. Quantitative Western Blot analysis: core concepts

At the heart of protein detection via Western Blotting is the use of quality reagents and correct methodology. While there are several resources available to overcome the technical limitations of Western Blotting, there is limited information available on how data can be quantified and the correlation between the combined linear range of detection and normalization strategy.

2.1. Break it right: “Fit for purpose”

The quality of the data gathered from a Western blot is as good as the quality of the sample used. Protein extracts need to be prepared, purified and quantified using reagents that best suit the sample type. How would you choose from the numerous extractions, fractionation,
Quantitative analysis of Western blot data

![Diagram showing the steps of Western blot analysis: Isolate and quantify protein extracts, Validate the antibodies, Validate the Internal Loading Controls (ILC), Identify the Combined Linear Range, and Determine change in protein abundance.]

**Fig. 1. Quantitative approach to Western Blotting.** Understanding of the physiology and cellular location of the target protein will allow for optimal protein extraction outcomes from various sample source. Quantification of protein extracts prior to Western blotting is crucial first step in generation of accurate data. Validation of antibodies is required both by the vendor as well as the end-user in the context of the experimental system under study. Internal loading controls (ILC) needs to identify and validated within the context of the experimental study. Both target and internal loading control needs to detect and analyzed within the combined linear range (CLR) of detection. Normalization of the target protein to appropriate ILC will provide an accurate representation of how target protein abundance is affected in the experimental study. Finally, biological and technical replicates not only provide statistical power but also rule out user bias and accounts for biological variability.

and purification methods? Several comprehensive reports have been published by researchers including Bass et al. [23], Pandey et al. [24], Wingfield et al. [25], Peach et al. [26], Gavini et al. [27], and Miskiewiz et al. [28], describe in great details various considerations for preparing quality protein extracts. Work published by Murphy and the group also describes the implication of sample fractionation on downstream protein detection [29]. Users’ experimental workflow influences the strategy for extraction and solubilization of target protein. It is critical, however, to standardize the sample preparation method in alignment with the experimental design and using consistent protocols throughout the duration of the study. In summary, the user needs to consider the type of target protein (nuclear, transmembrane, mitochondrial, etc.) as well as the type of tissue/cell that houses the target in order to choose the extraction, purification, and quantification strategy.

### 2.2. Sample: loaded or overloaded?

The next important question is; how much sample needs to be loaded for separation using two-dimensional electrophoresis. Sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) matrix, electroblotting (transfer), and protein binding capabilities of membranes influence together with the type of target protein influence the overall performance of Western Blotting technique. Overloading of samples is a widespread problem that often goes unrecognized and may compromise the accuracy of quantitative analysis [3,4,6,7,30,31]. Murphy and others have successfully shown that in fact, loading low amounts of protein sample can improve outcomes in Western Blotting [15,29,32]. For accurate quantitative analysis, sample loading must be evaluated and calibrated experimentally by the user. How do you determine the correct loading concentration for the sample of interest?

For accurate quantitative data analysis, the relationship between sample loading and band intensity must be evaluated and calibrated to determine the linear range of detection for the assay [3,6,7,29]. The linear range of detection is the range of sample loading where band intensity increases in proportion to sample loading or target abundance. Within this range, a change in sample input will produce a linear and proportional response in signal output which is determined coefficient of determination $R^2$. Closer the $R^2$ value is to more the linear regression reflects the data. For example, a two-fold or five-fold increase in sample loading should theoretically induce an equivalent increase in relative fluorescence units which would be recorded as band intensity (Fig. 2).

Analyzing data outside of the linear range may result in inaccurate and non-reproducible data. All quantification must be performed within the range where an increase in sample input ($x$, $2x$, and $5x$) produces a linear and proportional response in signal output ($y$, $2y$, and $5y$). Above the linear range of detection (shoulder), strong bands exceed the capacity of the assay and become saturated; the expected, proportional increase in band intensity does not occur and strong signals will be underestimated. Below the linear range (tail end), faint bands are difficult to reliably distinguish from membrane background and do not reflect actual differences in band intensity.

We strongly emphasize the dual importance of linearity and proportionality in the quantitative analysis of Western blot data. Linear regression can be used to fit a data set to a straight line, but this does not imply or guarantee proportionality. An $R^2$ value equal to 1 indicates that the regression line represents the data perfectly, but this line may or may not represent a proportional signal response. A linear and proportional relationship is represented by the equation $y = mx$, which describes a straight line passing through the origin. However, it may be possible to identify regions or subsets of the data that do display a proportional response.

### 2.3. Combined linear range (CLR) of detection

The linear range of detection is central to quantitative analysis of
Western blot data; the accuracy and validity of analysis are based on this foundation. Because it is influenced by numerous experimental factors, the linear range should be determined empirically for each Western Blot assay. A combined linear range denotes the linear range for both the target protein as well as a loading protein (also called Internal Loading Control (ILC)). Internal loading control as the name suggests is present internal to the test sample itself and could be a single Housekeeping protein (HKP) or total cellular protein. Serial dilutions of biological samples can be used to identify the appropriate linear range of sample loading for both the target protein(s) and internal loading control. It is important to note that this linear range will likely be different for each target protein and loading control combinations and may be affected by the detection chemistry and/or imaging platform. The relationship between sample input and band intensity must be both linear and proportional for both target protein and internal loading control to enable quantitative analysis. Choosing an appropriate ILC may seem simple, but it is an important and sometimes obscure aspect of experimental design for accurate quantitative analysis of Western blot data [6,10].

2.3.1. Internal loading controls

The ILC must meet the two requirements when evaluated in the specific conditions, treatments, cell or tissue types, and experimental context where it will be used. Because experimental manipulations may influence the expression of the ILC, biological stability should be validated in all samples that represent the intended conditions and treatments. Biological variability in expression of the ILC is a source of error that may undermine the accuracy of the analysis; particularly for analysis of small differences in target abundance [5,11,33-35]. The variability introduced by technical limitations (gel matrix, transfer conditions, buffers, etc.) of the Western blot method should also be considered when choosing and validating an ILC. Either type of variability may make it more difficult to reliably detect small differences and effects, leading to false-negative results [4,15,36,37].

Normalization of a target protein is most accurate when the target protein and ILC are detected in the same lane on a single blot [6,8,38,39]. To achieve this goal, detection must be performed in the combined linear range – the range of sample loading that produces a linear, proportional response in band intensity for both the target and the ILC [3,6,7,29,33]. If they cannot be accurately detected in the same range of sample loading, the results will not be meaningful; a different ILC should be selected for the experiment. This may occur for a variety of reasons, such as the use of an abundant HKP as an ILC for a low-abundance target, a pan/PTM analysis experiment where an abundance of unmodified and modified forms varies dramatically, if primary antibody specificity and affinity are poor, or if transfer methods require further optimization.

2.4. Internal loading control: housekeeping protein

A single endogenous reference protein, such as a housekeeping protein, is often used as an ILC. Members of Actin, tubulin families and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) are common examples. This method is referred to in the literature as a single-protein loading control [3]. Although this method is widely used and can be quite accurate, it is uniquely vulnerable to biological variability [40]. An HKP or other single-protein loading control alters and reformulates the experimental hypothesis, as described by Aldridge et al. [3]. Rather than evaluating the abundance of the target protein relative to total sample protein or cell number, the experiment now examines target abundance relative to that particular single-protein loading control [3]. If the expression of that single protein is stable and unaffected by the relevant experimental conditions and samples, this ILC meets the first requirement and the hypothesis is valid. But if the expression of that single-protein loading control is variable, the reformulated hypothesis will fail.

Adoption of HKPs as ILCs for immunoblotting was inspired by housekeeping genes, the stably expressed transcripts used as endogenous controls for analysis of gene expression analysis. Stable expression of actin, tubulin, and other housekeeping genes in many cell types and tissues initially indicated that their protein products should be similarly appropriate for Western blot normalization. However, the expression of some HKPs is now known to be altered in response to certain experimental conditions. Recent studies have demonstrated regulated expression of actin, tubulin, and other HKPs in response to conditions such as cell confluence, disease state, developmental stage, hormonal state, drug treatment, and cell or tissue type [33,35,37,41-49]. Manceau et al. [43], demonstrate the variability in β-tubulin and β-actin across different tissues extracts from mouse neonates. Because the variability of HKP expression in experimental samples will introduce error in quantitative analysis, validation is an important first step. Before an HKP or other single-protein loading control is used for quantitative analysis of Western blot data, its stable expression should be demonstrated and verified for the relevant experimental samples and manipulations [3,29,33-35,37,47].

2.5. Internal loading control: total cellular protein

In recent years, total cellular protein staining of the blotted membrane has emerged as the preferred method for normalization of sample loading in quantitative analysis of Western blot data [35,37,41,44,47,50-53]. Prior to the immunodetection of the blot, a fluorescent total protein stain is used to visualize and assess actual sample loading across the blot. The membrane is then processed for antibody incubations and detection. Total protein loading controls are much more resistant to biological variability than an HKP or other single-protein loading control [3,15,31,35,44,47,50,51] and may provide a wider linear range of detection than an HKP – and because of its resistance to biological variability, it requires less validation prior to use [3,4,7,15,35,41,44,50,51]. Because variation is often introduced during electrophoretic transfer from the gel to a membrane, total protein staining of the blot is preferable to staining of the gel prior to transfer or staining of a duplicate gel loaded with the same samples [3,35,47,51]. Any method used for the detection of total protein loading should not damage the bound sample proteins or interfere with subsequent antibody-based detection of the blot [6,31,35,54,55]. A total protein stain should meet a number of criteria, including an appropriate linear range of detection that is compatible with the target of interest; stable signals and low sample-to-sample variability; lack of interference with immunodetection and other downstream analysis; suitability for membrane staining, to correct for variability during transfer; and compatibility with multiple detection and quantification [3,6,35,41,47,51-53,55,56]. Several fluorescent membrane stains have recently been described in the literature, including Ponceau Stain [55], Coomassie Brilliant Blue, SYPRO Ruby stain, Blot FastStain [57], and Revert 700 Total Protein Stain [35,58-62].

2.6. Detection chemistry

Two main types of detection chemistry are used to visualize signals on Western blots. Enhanced chemiluminescence (ECL) is a popular enzymatic method that uses horseradish peroxidase (HRP) as an indirect reporter of secondary antibody binding. This HRP reporter produces photons of light as it consumes a luminol-based substrate. Fluorescence detection is a non-enzymatic method that uses fluorophore-labeled secondary antibodies as a direct reporter of antibody binding. Although chemiluminescence is widely used, its inherent enzyme/substrate kinetics are a source of variability and often fail to produce the linear, proportional signal response required for quantitative analysis of Western blot data [5,6,8,11,63]. The rate of the chemiluminescent reaction is dependent on the local concentration of enzyme and substrate. Substrate availability is a dynamic property, and
the reaction rate will vary continuously over time and across the surface of the blot. A low-intensity band will consume the available substrate much more slowly than a strong band with a high local concentration of HRP reporter that may rapidly deplete the available substrate. If the HRP concentration is too high, a signal may be lost in areas where the substrate is rapidly consumed and exhausted; this is the cause of ghost bands (reverse banding) and “burned-in” bands with brown or yellow precipitate [6,64]. Different chemiluminescent substrate formulation can be specifically used to detect femtograms (ultrasensitive), picograms or milligrams of target protein within a lysate [65,66]. As such simultaneous detection of high and low concentration of proteins in the same blot can be challenging when using one formulation. ECL substrate [67–69] thereby, resulting in a non-linear and narrow dynamic range of detection. However, titration of protein amount, the primary antibody can help in mitigating all the above-mentioned issues, provided the scientist is cognizant [52,70].

We and others developed fluorescent Western blot methods as a more convenient and quantitative alternative to indirect chemiluminescence [63,71]. Fluorophores are retained at the site of antibody binding and emit photons upon exposure to the appropriate wavelength of excitation light, typically in the near-infrared (NIR) spectrum [5,6,8,38,71–73]. Signals are very stable and much more reproducible because they are unaffected by timing, substrate availability, and other variables that limit the usefulness of chemiluminescent methods for quantitative analysis [5,6,8,11,74]. Fluorescence imaging also enables multiplex detection of a target protein and ILC in the same lane on the same blot, for more accurate correction of sample-to-sample and lane-to-lane variation [6,8,38,63]. Stripping of Western blots for re-probing can cause substantial loss of sample proteins from the membrane and is an avoidable source of error [6,8,30,31]. Fluorescence is generally recognized as the most accurate and reliable method for quantitative analysis of Western blot data [4–6,41,73,74]. Source of species in which antibodies were raised, the wavelength of secondary antibody tagged fluorophore, and imaging system needs are critical to performing fluorescence-based Western blot detection of proteins.

Saturation artifacts are not readily apparent in a Western blot image and may go undetected if serial dilutions of a sample are not examined to define the linear range of detection. Two types of saturation are frequently observed in quantitative analysis of Western blot data. Membrane saturation is caused by the overloading of the sample protein. During the transfer of an overloaded gel to a blotting membrane, abundant sample proteins bind in layers on the surface of the membrane [11,15,31,76]. If highly abundant proteins exceed the local binding capacity of the membrane, they will be washed away. The layering of abundant proteins also limits antibody access and may only allow antibodies to bind the top layer of protein. This layering effect causes underestimation of strong bands and may also interfere with the detection of low-abundance proteins on the blot. Membrane saturation is common may be difficult to identify without the evaluation of a dilution series of the sample. This type of saturation can affect any Western blot assay, regardless of the detection chemistry or imaging method [11,15,77].

Signal saturation occurs when the intensity of a strong band exceeds the capacity of the detection chemistry or imaging system. At this point, strong signals begin to plateau and increasing amounts of target no longer produce the expected increase in signal (Fig. 3). Chemiluminescent detection is highly susceptible to saturation; the generation of a signal is restricted by substrate availability, and the linear range is quite limited, even when detected by digital imaging [5–7,11,15,74]. Fluorescent Western blot methods are relatively resistant to saturation and provide a much broader linear range, particularly when imaged with a digital system that provides an appropriately wide dynamic range [5–8,73,74]. However, in heavily loaded samples, self-quenching of tightly packed fluorophores may be possible [6].

2.7. Normalized Western blot data

Normalization mathematically corrects for small, unavoidable variations in sample loading and protein transfer by comparing the target protein to the ILC in each lane. The target protein signal in each lane is divided (normalized) by the signal value for the ILC in that lane. When target signals are normalized to sample loading, relative levels of the target protein can be compared across the blot to determine if changes in band intensity represent biological differences between samples. Normalization is based on the fundamental assumption that both the target and ILC signals are dependent on sample concentration. For this assumption to hold true, as mentioned above the ILC must meet two requirements: it must be expressed at a generally constant level across all relevant experimental samples and conditions; and the resulting signal intensity must be proportional to abundance, without saturation effects or technical limitations. A loading control that does not meet both requirements is not an accurate indicator of sample loading and should not be used for normalization and quantitative analysis of Western blot data [3,31,47,51].

Without normalization, an apparent difference in target abundance on a Western blot cannot be accurately interpreted. This concept is illustrated in Fig. 4. Detection of the target protein is seen in Fig. 4A (green). Plotting the raw intensity values for the target may suggest that target abundance is variable among the different samples (Fig. 4C). However, the observed difference between target bands may be the result of inconsistent loading of total protein extract (Fig. 4B). Sample #4 appears to have reduced levels of target protein compared to rest of the samples, however, after normalization to correct variations in sample loading; it is evident that sample extracts were loaded unequally in each lane and therefore, any changes in target protein observed may be due to improper loading of protein extracts and not due to experimental intervention (Fig. 4D).

Normalization is appropriate for the correction of small and unavoidable differences in cell number, sample concentration, loading error, and position or edge effects from electrophoresis and transfer. However, it should not be relied on to correct for preventable error and variability in the quantitative analysis of the Western blot process [3,6]. Normalization should be used in addition to careful experimental design and sample preparation, not in place of it. Minimizing or eliminating sources of variation throughout the experiment is critical for reproducible quantitative analysis of Western blot data (see Refs. [6,7,29,78] for excellent information about protein extraction, sample handling, and other error-prone aspects of the Western blot method).

Total protein normalization is an aggregate method that combines the band intensities from many different sample proteins in each lane – essentially using multiple ILC proteins to provide a more accurate readout of sample loading. This multi-protein approach is also called normalization by sum [36]. This normalization method also offers a more direct readout of the cellular material loaded than antibody-based detection of a single endogenous protein [3,4,31,35,36,41,44]. Total protein normalization is now increasingly viewed as the “gold standard” for performing quantitative analysis of Western blot data [52,53]. The Journal of Biological Chemistry recommends normalization to total protein loading, clearly stating their preference “that signal intensities are normalized to total protein by staining membranes with Coomassie Blue, Ponceau S, or other protein stains” [52,53].

3. Background subtraction

Western blots not only contain the signal generated by the immunodetection of the target protein but also signal generated as a result of nonspecific bands, smears, and the inherent signal produced by the membrane itself. When a target band is quantified the background signal around the band is also included in the data analysis, and the goal is to reduce/remove the confounding error introduced by such background signal [23,79,80]. Currently, several Western blot analysis
software packages offer multiple options to subtract background signal and quantify target bands. Each Western blot has a unique profile with respect to artifacts, uniformity of background, positioning of lanes and bands and requires a background subtraction algorithm that can adapt to the said variations.

There are several methods to subtract the background from the Western Blot image [23,79,81–84] either using the Shape-Based methods for local and/or global background subtraction; Lane-Based methods such as Rolling Ball subtraction. A new patent-pending method called Adaptive Background Subtraction (ABS) removes user-bias by taking into consideration both shape as well as lane profiles to generate reproducible and accurate quantitative Western Blot data. ABS background subtraction algorithm present in Empiria Studio Software works well with real-life Western blots that have oddly-shaped misshapen bands and smears, and the data generated has greater accuracy than local and global subtraction methods (For more details read [85]; Fig. 5).

4. A systematic approach to accurate quantitative analysis of Western blot data

Here, we describe a five-step systematic strategy to increase the precision and reliability of the quantitative data generated from the Western blotting technique. This process incorporates critical validation steps that address common, but sometimes unrecognized, sources of error and variation. Because we and others strongly prefer and
recommend fluorescent Western blot methods for quantitative analysis [5-8,52,63,74], this strategy was specifically developed for use with fluorescence-based detection.

4.1. Validate the primary antibody

After sample proteins are separated on a protein gel and transferred to the membrane, primary and secondary antibodies are used to bind and visualize the target protein. This process relies on two key properties of the primary antibody: specificity, the antibody’s ability to recognize and bind to the target antigen; and selectivity, the antibody’s preference to bind the target antigen in the presence of a heterogeneous mixture of sample proteins. These characteristics should be verified and validated for each primary antibody. Because antibody performance is greatly influenced by the assay context and parameters, validation should be performed by the user in the intended assay and relevant experimental context [18,86-94]. The reproducibility of results produced with the primary antibody should also be verified [18,76,88,95].

4.1.1. Specificity

The specificity of antibody within the realm of a Western blot means that the antibody recognizes the target protein(s), either as a single
distinct band or a set of bands of the correct molecular mass. Detection of a single band at the expected molecular weight is an important first step but is not sufficient to prove antibody specificity. It is wise to remember that the presence of "one band" on a blot does not mean "one target" [7,53,86,87,96,97]. Antibodies can be validated by testing the performance of antibodies on either genetic knockout, knockdown, positive, and negative protein samples (see for more details [92,98-104]). The specificity of the antibodies against Pan and Phospho-EGFR was validated in a 2-pronged approach by using a peptide blocking strategy. A single blot (Fig. 6A and B) is divided into 2 halves (yellow line) and each half was then incubated with stained with a Total Protein Stain (Fig. 6A; Revert 700 stain) which was detected in the 700 nm channel and Anti-Phospho-EGF Receptor (EGFR) antibody (Fig. 6B) which was detected in the 800 nm channel. The portion of the blot with the blocked antibody cocktail (Anti-Phospho EGFR + Phospho-EGFR immunogen peptide) has Revert 700 signal but no visible signal in the 800 nm channel (Fig. 6B; Right of the yellow line). Staining with a total protein stain ensures that equal protein was loaded in all the lanes and it is not a lack of protein lystate that contributes to absent bands in the blocked portion of the membrane. This data suggests that antibody against Phospho-EGFR protein detects a band ~175 kDa which is absent in the blocked blot. A separate blot containing equal amounts of protein extracts was divided into 2 halves (Fig. 6C and D) and multi-color detection of Pan-EGF (Fig. 6C) and Phospho-EGF (Fig. 6D) was performed. The portion of the blot with the blocked antibody cocktail (Anti-Phospho EGFR + Phospho-EGFR immuno- gen peptide) has protein signal in the 700 nm channel from immuno-detection of Anti-Pan-EGFR (Fig. 6C) but no visible signal in the 800 nm channel (Fig. 6D; Right of the yellow line).

4.1.2. Selectivity
In a typical quantitative analysis of the Western blot data, target protein abundance may be lower relative to a large excess of unrelated sample proteins. The antibody must be able to overcome this imbalance and selectively bind the target antigen in a complex mixture, without interference from off-target binding [90,92]. For this reason, selectivity should be verified with endogenous levels of target expression in the complex sample [76,88]. Purified or overexpressed target protein alters the balance of protein abundance in the sample, creating an artificial context that may not reflect actual antibody selectivity and expected off-target binding [88,105,106].

Optimization of certain assay conditions may improve antibody selectivity. Insufficient antibody dilution and extended incubation times may promote off-target binding and detection of undesired, nonspecific bands [30,47,76,92]. Storage and reuse ("recycling") of diluted primary antibodies is not recommended; this practice results in inconsistent antibody quality, titer, and stability that may introduce error and make a quantitative analysis of Western blot data results less reproducible [76]. Blocking buffer can also have a dramatic impact on antibody selectivity; an inappropriate blocking agent may greatly increase off-target binding in Western blot analysis [107,108].

4.1.3. Publication guidelines
Detailed reporting of antibody details and experimental methods is an important aspect of quantitative analysis of Western blot data [53,87,109,110]. Recommendations for reporting of antibody information are addressed elsewhere (including [76,87,110]). The authors should describe how antibody specificity was validated, including the methods used, assay parameters, positive and negative controls, and sample type (cell or tissue type, source, lystate, overexpressed or purified target protein) The authors should be prepared to submit the raw validation data and unprocessed images during review [76,87,110].

4.2. Validate the internal loading control
An ILC must be present in all experimental samples at a stable level
Fig. 7. Identification and validation of an appropriate HKP in the MAPK/ERK pathway. Cell extracts prepared from untreated (UT) or Epidermal growth factor (EGF) (ET) treated A431 cells were loaded onto 10% 10% Bis-Tris gel (NP0303BOX) and run under the MOPS buffer system. All the blots were incubated with Intercept Blocking Buffer for one hour at room temperature. Technical replicates for both UT and ET categories are presented. Proteins belonging to MAPK/ERK pathway were selected to be tested via Western blot method. Total protein extract was visualized in each blot (a, a', a'') using Revert 700 Total Protein Stain (LI-COR P/N 926-11010). Blots; b, b', and b'' were incubated with Anti-Ras antibody (CST #3965; Lot #1); Anti-CDK4 (D9G3E) antibody (CST #12790; Lot#14); and Anti-CDK6 (DCS83) antibody (CST #3136; Lot #4). Merged images of the blots (c, c', c'') imaged under different acquisition channels; 700 nm for Revert 700 (a, a', a'') and 800 nm for Ras, CDK4, and CDK6 (b, b', b'') highlight uniform loading of sample in each lane. Ras, CDK4, and CDK6 levels appear to be unchanged between untreated and EGF treated A431 cell extracts (d, d', d''). F test to compare variance, P value > 0.05. Sample loaded: 5 micrograms; Blocking buffer: Intercept Blocking Buffer (TBS); Intercept T20 (TBS) Antibody Diluent; Protein ladder: Chameleon duo ladder (LI-COR P/N 928-70000; Lot# C70803-03). Imager: Odyssey® CLx; resolution: 169 μm; Intensity: auto mode.
to be used as a valid indicator or proxy of sample loading [3,29,33–35,37]. Stability should be demonstrated before an HKP or other single protein is used for quantitative analysis of Western blot data normalization, to ensure that expression of the ILC is not affected by experimental treatments or manipulations [29,33,34,47,53,111].

4.2.1. Stable or Unstable: ILC?

Expression of the HKP or other single-protein control should be carefully examined in samples that represent the desired experimental conditions. A comparison of HKP levels to total protein loading in these samples is a straightforward way to demonstrate the stability of expression. Alternatively, HKP levels may be compared to an unrelated, endogenous protein already shown to be stably expressed in the relevant experimental samples and conditions. If changes are made in experimental treatments, cell line, tissue type, cell density, or other relevant experimental parameters, stable HKP expression, as well as a linear range of detection, must be re-validated for the new conditions [33,34,47]. Several proteins belonging to the MAPK pathway were tested for stability in both vehicles treated and etoposide treated Jurkat cell extracts (Fig. 7c, c’, c”). Each blot was also stained to visualize total protein loaded in each lane (Fig. 7a, a’, a”). No significant changes in protein expression were observed for Ras, CDK4, and CDK6 proteins (Fig. 7b, b’, b”; d, d’, d”), and therefore, is suitable to be used as HKP for the samples under consideration.

4.3. Two-sides of the same coin: Post-Translationally modified (PTM) proteins

Western blots generated to detect PTM proteins utilize an ILC that accounts for the total presence of the unchanged protein (also known as Pan-protein). Here, a modification-specific primary antibody is multiplexed with a pan-specific primary antibody that recognizes the target protein in any modification state [6,38,71,72,112–114]. The pan- and modification-specific antibodies should be derived from different host species, so they can be discriminated by secondary antibodies labeled with spectrally distinct fluorophores [63]. Pan/Phospho and other types of pan/PTM analysis are widely used to monitor and compare relative changes in protein modification across a group of samples. Pan/Phospho analysis is specifically recommended by the Journal of Biological Chemistry, which stipulates that “phospho-specific antibody signals

Fig. 8. Determination of the combined linear range of detection for CPARP, β-ACTIN (ACTB), GAPDH, and COXIV. Jurkat cell extracts harvested from cells treated with 25 μM Etoposide for 5 h at 37 °C was loaded on 4–12% Bis-Tris NuPage gels and electrophoresed under the MOPS buffer system. The blots were blocked with Odyssey Blocking buffer: TBS for one hour at room temperature and subsequently incubated with primary antibodies against CPARP (CST #5625; Lot#13), ACTB (LI-COR P/N 926-42210; Lot #C80322-01), GAPDH (CST #5174; Lot #2), and COXIV (LI-COR P/N 926-42214; Lot #CS033324-02) (~1 μg/ml) at 4 °C overnight with gentle shaking. Subsequently, the primary antibody binding was visualized by using IRDye 800CW Goat anti-Rabbit IgG (H + L). Sample loaded: 1.25–40 micrograms; Blocking buffer: Odyssey Blocking Buffer (TBS); Protein ladder: Chameleon duo ladder (LI-COR P/N 928-70000; Lot# C70803-03). Imager: Odyssey® CLx; resolution: 169 μm; Intensity: auto mode. Linear range of detection for all the proteins are plotted in (A). Linear range of CPARP and COXIV (B) appears to have overlapping range of detection compared to CPARP and ACTB (C) and CPARP and GAPDH (D).
should be normalized to total levels of the target protein" [97,111]. Although pan/phospho analysis is the most commonly performed type of pan/PTM analysis, other protein modifications are also studied with this method (ubiquitination and palmitoylation analysis). Pan-protein normalization accounts for changes in expression of the target protein, which might confound analysis of the abundance of the modified form. It also eliminates the stripping and reprobing of blots, which introduces error by causing loss of blotted proteins from the membrane [6,31,38]. Because the unmodified and modified target protein is detected on the same blot and in the same lane, normalization can correct for transfer artifacts in the blot. Pan/PTM analysis should only be used for comparison of relative abundance; it does not provide information about the stoichiometry of the modification [77].

4.3.1. Publication guidelines

If an HKP is used for quantitative analysis of Western blot data normalization, authors should be prepared to provide evidence that its expression was not affected by the experimental treatments applied [53,111]. The editors of the Journal of Biological Chemistry have expressed a preference for total protein normalization in quantitative analysis of Western blot data [53].

4.4. Define the linear range of detection

Using the validated antibodies, it is important to determine the appropriate amount of sample to load to generate data that can be quantified. Normalization and analysis must be performed in the linear range of detection, where a linear and proportional response is observed between sample loading and band intensity.

4.4.1. Finding the combined linear range for a target protein and internal loading control

Careful examination of the linear range is particularly important for HKPs and other single-protein loading controls. The strong bands produced by highly abundant ILC are frequently affected by saturation. If the abundance of a target is substantially lower than the single-protein loading control, the combined linear range may be very narrow. Total protein normalization typically provides a wider combined linear range, without the rapid saturation observed for HKPs [35,44,50,51,55,56]. This wider combined linear range is very helpful for low-abundance target proteins, which often require the loading of large amounts of sample protein and cannot be used with a highly abundant HKP.

The experiment in Fig. 8 demonstrates the differences in a linear range that can be observed between different HKPs. Cleaved PARP (c-
PARP) and three different HKPs (β-actin, COX IV, and GAPDH) were detected in lysates from Jurkat cells treated with 25 μM Etoposide for 5 h at 37 °C (check material and method section for more details). By electrophoresis, all four proteins could be detected simultaneously on a single Western blot. This facilitates comparison by eliminating blot-to-blot inconsistencies in sample loading, electrophoresis, and transfer. Fig. 8A shows a combined linear range for poly (ADP-ribose) polymerase cleavage (CPARP), GAPDH, COXIV, and β-actin (ACTB). In Fig. 8B, the combined linear range for CPARP and COX IV is linear between ~7.5 and 30 micrograms. However, there is no useable combined linear range for CPARP, ACTB, and GAPDH (Fig. 8C and D). In these samples, band intensity for GAPDH and ACTB begins to saturate and plateau at ~15 μg of sample protein loading.

Working in the middle of the combined linear range helps to limit error and variability introduced by high- and low-intensity data points. With an HKP or other single-protein loading control, high-intensity data points introduce error and should not be used for normalization; these data points will increase the mean Coefficient of Variance (CV) of the normalized data and should be avoided [15,36]. From a statistical perspective, variability introduced by a single-protein loading control may increase the frequency of false-negative results (small, but statistically significant, differences between samples that are not identified during data analysis).

If validation experiments fail to identify a combined linear range of detection, as in Fig. 8C–D, a different ILC may be required. The use of total protein for normalization decreases the variability of high-intensity data points but does have the potential to increase variation for low-intensity measurements [36]. Fig. 9 demonstrates the combined linear ranges for OCT-4 protein in HEK293T cell extracts derived from control and experimental categories. For both lysates, 4.5 micrograms appear to be in the middle of the combined linear range of detection for Total protein extracts and OCT-4 protein. Optimization of transfer methods, sample loading, antibody choice, detection chemistry, or other assay parameters may help to improve the linear range of detection for the selected ILC. In general, increasing the number of internal control proteins will reduce the mean CV and improve the accuracy of normalization, making it easier to reliably detect small or subtle differences between samples [3,6,35,36,51].

4.4.2. Publication guidelines

When quantitative analysis of Western blot data results are reported, authors should disclose how signal intensity was quantified as well as how the linear relationship between sample loading and signal intensity was confirmed for each antigen [53,111].

4.5. Data analysis

Data analysis of Western blot data sets should be performed on biological as well as technical replicates generated using the validated antibodies, combined linear range of detection, and valid ILC established in the prior steps. After quantification of target and ILC signals, data analysis is performed.

Analysis typically begins with the designation of the experimental control sample that will be used for relative comparison. The lane normalization factor is then calculated for each lane, using the HKP band intensity values or combined total protein signal values for each lane (Fig. 10A). This factor is calculated by dividing the signal intensity value for each experimental sample by the signal intensity observed for the control. To apply the normalization factors, the signal intensity of the target band in each lane is divided by the lane normalization factor for that lane. This process generates the normalized signal intensity value for each sample.

Fold change = \[
\frac{\text{normalized signal}_{\text{Experimental sample}}}{\text{normalized signal}_{\text{Control sample}}}
\]

A ratiometric analysis is often performed to enable a relative comparison of target abundance across a group of samples. The normalized target signal for each sample is divided by the normalized target signal observed in the control sample. These ratios express the abundance of the target protein as a fold or percentage change, relative to the control. A comparison of all samples to the experimental control produces relative values that are unitless, proportional, and independent of the raw signal intensity. Fold change values > 1.0 indicate increased abundance relative to the control, and values < 1.0 denote decreased abundance. Percentage change expresses the same information as a percentage, with a positive value indicating increased relative abundance and a negative percentage indicating decreased abundance (Table 1).

Methods used to quantify and normalize signal intensities should be disclosed, along with data analysis methods and software tools used for quantification and analysis. Raw data showing image analysis and quantification, including original digital images of Western blots, may be requested during peer review [13,53,111,115].

4.6. Replicate and reproduce

Replication of quantitative analysis of Western blot data results confirms the validity and reproducibility of any observed changes [13,14,18,20,21]. Replicate measurements help to ensure that the experimental effects we report are reproducible – that they represent actual differences between samples, rather than artifacts of experimental variability or noise [12,18]. Replication is essential when analyzing small or subtle differences between samples, to characterize and understand the contribution of error and the limits of quantitative analysis [12,19,20]. Two types of replicates, technical and biological, are commonly used to address different questions in quantitative analysis of Western blot data.

4.6.1. Technical and Biological Replicates

Technical replicates are repeated measurements of the same sample, used to characterize the precision and variability of any assay or method [19,20,116]. Common examples include loading the same sample in multiple lanes on a Western blot, running replicate blots in parallel, or running the same kind of gel multiple times on different days. High variability between technical replicates makes it more difficult to separate an observed experimental effect from the inherent variation of the Q WB assay [12,19,20,53,117]. As a result, small but genuine differences between samples may not be detected.

Biological replicates are parallel measurements of independent and biologically distinct samples that are intended to control for random biological variation [9,19–21]. A biologically relevant effect should be observed reproducibly in independent samples. Analysis of samples derived from multiple individuals or from multiple, independent cell cultures are common examples. In some cases, a similar experimental effect may be demonstrated in a different biological system or context [13,20,86,117]. Biological variability is generally expected to be higher than technical variability [12,16,20]. Considerations for choosing appropriate replicates are described by others [19–21]. The replication strategy should be developed prior to the generation of quantitative
biological replicates and n = 3 technical replicate per biological replicate, using the normalized values (graph shows individual fold-change values and mean). (n = 3 technical replicates for each sample; n = 3 biological replicates). CLR for total protein and CPARP was identified at 5 micrograms for all three cell types. Blocking buffer: Odyssey Blocking Buffer (TBS); Protein ladder: Chameleon duo ladder (LI-COR P/N 928-70000). Image: Odyssey CLx; resolution: 169 μm; Intensity: auto mode. Image Studio software was used for protein band detection. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Abundance of target protein, expressed as fold or percentage change relative to the control sample.

<table>
<thead>
<tr>
<th>Fold Change</th>
<th>Percentage change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0%</td>
</tr>
<tr>
<td>1.25</td>
<td>+25%</td>
</tr>
<tr>
<td>1.5</td>
<td>+50%</td>
</tr>
<tr>
<td>2.0</td>
<td>+100%</td>
</tr>
<tr>
<td>0.75</td>
<td>-25%</td>
</tr>
<tr>
<td>0.5</td>
<td>-50%</td>
</tr>
<tr>
<td>0.1</td>
<td>-100%</td>
</tr>
</tbody>
</table>

4.6.2. Using replicates to characterize error and noise in quantitative analysis of Western blot data

The coefficient of variation (CV) of replicate measurements can be used to evaluate the precision of quantitative analysis of Western blot data results. A low CV indicates high precision of measurement and low signal variability in the assay; a larger CV denotes reduced precision and greater assay variability. The CV can be used to determine if the magnitude of a relative change in target abundance is large enough to be reliably detected above the assay noise. As a general rule of thumb, the magnitude of an observed effect should be at least twice as large as the mean CV of the replicate measurements. In other words, for an observed difference of 25% between samples (a 0.75-fold or 1.25-fold change), the mean CV of replicate samples should be less than 12%. To confidently report a small effect, it may be necessary to minimize sources of variation in the experimental protocol and increase the number of replicates to reduce the mean CV.

Normalization of Western blot data should always be performed before the analysis of individual replicate samples. Raw signal intensities should never be directly compared between blots. A multitude of experimental factors influence the raw signal intensities observed on each blot, and direct comparison of band intensity is meaningless. Raw band intensity is useful and informative only when analyzed by relative, ratiometric comparison to an appropriate control sample on the same blot. To minimize error introduced by position effects, we recommend loading replicate samples in random order for gel electrophoresis. Consistent placement of certain samples in the same position (for example, always loading the positive control in the farthest left lanes of the gel) may consistently produce edge effects that will be inadvertently propagated throughout the data analysis.

4.6.3. Publication guidelines

The number of replicates performed for each quantitative analysis of Western blot data measurement should be clearly indicated, with sufficient information to clearly distinguish technical replicates from independent biological replicates [13,19,53,111]. Figure legends should address experimental uncertainty and explicitly define n for each experiment. Some journals also request that authors present a quantitative analysis of Western blot data and other small datasets as scatter plots with error bars. This format more clearly depicts the spread and distribution of data points than the traditional bar-and-plunger presentation [53,109,118,119]. Fig. 10 illustrates how different data distributions can produce the same bar graph – a graph that may suggest...
different conclusions than the full dataset [118,119]. Weissgerber et al. [119] recently introduced free, online interactive tools designed to improve data visualization for small datasets. The “interactive dot plot” tool allows different graphs to be viewed as univariate scatterplots, box plots, and violin plots. It also enables visualization of clustered non-independent data, such as technical replicates. The “interactive repeated experiments” is designed for visualization and comparison of data from repeated, independent experiments.

5. Discussion

Accurate, reproducible quantitative analysis of Western blot data requires careful experimental design and execution. The core concepts described here should be considered during experimental design and planning. Before a sample is lysed for protein extraction, careful consideration must be given to habitually disregarded aspects such as the cellular location of the target protein, purification planning. Before a sample is lysed for protein extraction, careful consideration must be given to habitually disregarded aspects such as the cellular location of the target protein, purification strategy, as well as how the extract will be quantified. Loading of the quantified sample needs further attention into the gel and buffer composition, and electrophoresis and protein transfer methods. Validation of antibodies used for immunodetection of the target protein should be undertaken by the user and data shared with the readers. Identification of an appropriate housekeeping protein that is stable is crucial and operating within the combined linear range of the HKP and the target protein is even more essential.

In recent years, standards and checklists have been proposed to enhance reproducibility by improving the transparency of reporting, increasing recognition of common experimental pitfalls, and raising awareness about the importance of validation [13, 14, 109, 115, 120–123]. These measures have stimulated conversation about these important topics, which is critical for the development of the nomenclature and resources needed to reach a consensus on guidelines and best practice. Such guidelines must be a community effort and can be helpful to both new and established investigators in the field [13,14,121]. Standards and guidelines can also clarify and simplify the peer review process. Han et al. [120] recently examined the impact of a manuscript submission checklist on the transparency and quality of reporting in preclinical biomedical research. They evaluated the reporting of methodological and analytical information in two journals: Nature (which implemented a mandatory checklist in 2013) and Cell (which does not require a checklist). Comparison of 2013 (pre-checklist) and 2015 (post-checklist) data shows that a required submission checklist is associated with improved reporting of methodological details [120].

To be truly useful, guidelines cannot be overly rigid; in the laboratory, one size does not fit all [13]. But flexible and appropriate guidelines can call attention to common mistakes, raise awareness, and make researchers stop and think about the strengths and limitations of the methods they use. The systematic workflow we present here seeks to minimize the error introduced by common methodological limitations and improve the overall usefulness of quantitative analysis of Western blot data. We believe this flexible and adaptable approach has the potential to improve the quality, reproducibility, and precision of quantitative analysis of Western blot data. It will also help to ensure that the data generated in today’s quantitative analysis of Western blot data experiments will be able to meet increasingly rigorous publication standards in the future. The ultimate goal, of course, is to improve the overall accuracy of quantitative analysis of Western blot data by increasing the likelihood that the changes we observe, and report are reproducible and reflect real and meaningful differences between biological samples.

6. Materials and methods

6.1. Cell culture

A431 (ATCC® CRL-1555™), Jurkat (ATCC® TIB-152™) were cultured at 37 °C in a humidified atmosphere containing 5% CO2 as per ATCC guidelines. Cells were cultured up to 80% confluence before being treated or harvested to be stored at −80 °C.

6.2. Cell lysis

Cells were resuspended in radioimmunoprecipitation assay (RIPA) buffer (Pierce #89900) supplemented with Halt™ Protease and Phosphatase Inhibitor Cocktail (100×) (Thermo Scientific #878440) and incubated on ice for 20 min with intermittent vortexing. Lysates were sonicated twice at 1-min intervals until no pellet was visible and then clarified by centrifugation at 10,000 RCF for 15 min at 4 °C. The supernatant was taken and protein concentration determined by bicinchoninic acid (BCA) assay as previously described before being stored at −80 °C.

6.3. Commercial cell lysates

C32 Whole Cell Lysate: sc-2205, POU5F1 overexpression lysate (Origene # LY400950; Lot #01AF4D); HEK293T lysate (Origene #LY500001 Lot #01AF4D); were mixed with either 2× protein loading buffer (PLB) (LI-COR # 928-40004) OR 2× SDS buffer (OriGene) and denatured by boiling at 97 °C for 5 min. The aliquots were then immediately placed on ice for 5 min before being spun down.

6.4. Western Blot

Lysates were loaded onto NuPAGE 4–12% 15 well Bis-Tris gels (Invitrogen; NP0323BOX), NuPAGE™ 10% Bis-Tris, 1.0 mm, 10-well protein gels (ThermoFisher Scientific; NP0301BOX) and run at 200 V for 37 min. Gels were wet transferred using 20% methanol onto either Nitrocellulose membranes’ μm; 7 cm × 8.5 cm (P/N 926-31090). Membranes were blocked for one hour at room temperature in Odyssey Blocking Buffer (TBS) (LI-COR 927-50000). Primary antibodies were diluted with their respective blocking buffers (see figure legends) and incubated overnight at 4 °C. Washes were performed with TBS 0.1% Tween-20 (TBST) before the addition of secondary antibody for one hour at room temperature. Washes were performed with 1 × TBST before imaging on Odyssey CLx. Protein detection was performed using Image Studio Ver 5.2.

6.5. Antibodies

Primary antibodies include β-Actin (LI-COR 926-42212; Lot #CS03324-02), CPARP (CST #5625; Lot #13), ACTB (LI-COR P/N 926-42210; Lot #C80322-01), GAPDH (CST #5174; Lot #2), and COXIV (LI-COR P/N 926-42214; Lot #CS03324-02), Phospho EGFR (CST #3777; Lot #13); Pan EGFR (Thermo Fisher Scientific AHR5062; Lot # 737350137A), Ras (CST #3965; Lot #1), CDK4 (CST #12790; Lot #14), CDK6 (CST #3136; Lot #4), OCT-4 (CST #2750; Lot #4). Primary antibodies were detected using IRDye 800CW Goat (polyclonal) anti-Mouse IgG (H + L) highly cross-adsorbed (LI-COR #925-32210), IRDye 800CW Goat (polyclonal) anti-Rabbit IgG (H + L) highly cross-adsorbed (LI-COR # 926-32211).

6.6. Imaging

All blots were imaged wet on the Odyssey* CLx imaging system using 680 nm and 780 nm channels. Protein detection was performed using Image Studio Ver 5.2. Adobe Photoshop Elements 13 was used to prepare image panels and annotations.
Author contributions
A.R.S.G co-wrote the main manuscript text and prepared Figures. J.A.H supervised, reviewed, and proof-read the manuscript. LPK designed and conducted the experiments, wrote the main manuscript text and finalized the Figures. ARS and LPK are co-first authors of this manuscript.

Declaration of competing interest
L. Pillai-Kastoori, A. Schutz-Genschwender, and Jeff A. Harford are employees of LI-COR Biosciences.

Acknowledgments
We would like to sincerely thank Samuel Egel for critical review the manuscript; Jonathan Goodding for graphical contribution to the manuscript.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.abb.2020.113608.

References


N.O.I. Research, NIH Rigor and Reproducibility Training Module 3: Biological and Technical Replicates, NIH Office of Intramural Research, NIH.

A. Casadevall, L.M. Ellis, E.W. Davies, M. McFall-Ngai, P.C. Fang, A framework for improving the quality of research in the biological sciences, MBio 7 (2016).


Reproducibility: let's get it right from the start, Nat. Commun. 9 (2018) 3716.