Immuno-Detection by sequencing (ID-seq) enables large-scale high-dimensional phenotyping in cells.

Jessie A.G. van Buggenum[1], Jan P. Gerlach[1], Sabine E.J. Tanis[1], Mark Hogeweg[1,2], Pascal W.T.C. Jansen[3], Jesse Middelwijk[4], Ruud van der Steen[4], Michiel Vermeulen[3], Cornelis A. Albers[1,2] and Klaas W. Mulder[1,*]

Affiliations:
Department of Molecular Developmental Biology, Radboud Institute for Molecular Life Sciences, Radboud University, PO Box 9101, 6500 HB, Nijmegen, The Netherlands[1]. Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, PO Box 9101, 6500 HB, Nijmegen, The Netherlands[2]. Department of Molecular Biology, Radboud Institute for Molecular Life Sciences, Radboud University, PO Box 9101, 6500 HB, Nijmegen, The Netherlands[3]. Biolegio BV, PO Box 91, 6500 AB Nijmegen, the Netherlands[4].

* To whom correspondence should be addressed (k.mulder@science.ru.nl)

Abstract

Cell-based small molecule screening is an effective strategy leading to new medicines. Scientists in the pharmaceutical industry as well as in academia have made tremendous progress in developing both large-scale and smaller-scale screening assays. However, an accessible and universal technology for measuring large numbers of molecular and cellular phenotypes in many samples in parallel is not available. Here, we present the Immuno-Detection by sequencing (ID-seq) technology that combines antibody-based protein detection and DNA-sequencing through the use of DNA-tagged antibodies. We used ID-seq to simultaneously measure 84 (phospho-)proteins in hundreds of samples and characterise the role of 225 kinases in primary human epidermal stem cells. Our analysis revealed a previously unappreciated downregulation of mTOR signalling associated with differentiation and identified 13 kinases regulating epidermal renewal through several distinct mechanisms.

Introduction

Quantification of protein levels and phospho-rylation events is central to investigating the cellular response to perturbations such as drug treatment or genetic defects. This is particularly important for cell-based phenotypic screens to discover novel drug leads in the pharmaceutical industry. However, the complexity of biological and disease processes is not easily captured by changes in individual markers. Currently, a major limitation is the trade-off between the number of samples and the number of (phospho-)proteins that can be measured in a single experiment. For instance, immunohistochemistry[1], immuno-fluorescence[2] allow high-throughput protein measurements using fluorescently labelled antibodies. However, these methods are severely limited by the number of (phospho-)proteins measured in each sample due to spectral overlap of the fluorescent reporter dyes. Current approaches that use antibody-DNA conjugates detect only a few epitopes at low sample throughput[3–7] . The more recently developed mass cytometry technology[8–11] presently allows multiplexing of around 30 antibodies per sample, yet is limited in sample throughput and requires expensive dedicated equipment and reagents. Although mass spectrometry-based shot-gun proteomics has recently become more readily available, this approach is still not cost-efficient to routinely apply on hundreds to thousands of samples and generally does not detect phosphorylated proteins effectively. Here, we present Immuno-Detection by sequencing (ID-seq) as a streamlined universal technology for measuring large numbers of molecular and cellular phenotypes in many samples in parallel. We show that high-throughput sequencing of antibody coupled DNA-barcodes allows accurate and reproducible quantification of 84 (phospho-)proteins in hundreds of samples simultaneously. We applied ID-seq in conjunction with the Published Kinase Inhibitor Set (PKIS) to characterize the role of >200 kinases in primary human epidermal stem cell renewal and differentiation. This demonstrated a previously unrecognized downregulation of mTOR signaling during differentiation and uncovered 13 kinases regulating epidermal renewal through distinct mechanisms.

Results

Precise and sensitive multiplexed (phospho-) protein detection via sequencing antibody-coupled DNA-tags.

The Immuno-Detection by sequencing (ID-seq) technology is designed to simultaneously measure many proteins and post-translational modifications in high throughput (Figure 1a). At the basis of ID-seq lie antibodies that are labelled with a double-stranded DNA-tag[12] containing a 10 nucleotide antibody-dedicated barcode and a 15 nucleotide Unique Molecular Identifier (UMI, Figure S1). Each antibody signal is now digitized and non-overlapping, allowing many antibodies to be combined and measured simultaneously. Following immunostaining and washing, DNA-barcodes are released from the antibodies through reduction of a chemically cleavable linker[12] and a sample specific barcode is added through PCR (see materials and methods for details). Finally, samples are pooled to prepare an indexed sequencing library (Figures 1a, S1, S2 and Supplementary note 1). This triple barcoding strategy facilitates incorporation of 100s-1000s of samples per experiment and achieves count-based quantification over four orders of magnitude (Figures S3 and Supplementary note 2). Furthermore, measurements on 17 antibody-DNA conjugates using singleplex immuno-PCRs or multiplexed ID-seq show high correspondence ($R = 0.98 \pm 0.046$), demonstrating that multiplexing does not interfere with antibody detection (Figure 1b and Figure S4). Moreover, the ID-seq library preparation procedure is reproducible ($R = 0.98$, Figure 1c) and accurate, as determined using multiple distinct DNA-tag sequences per antibody (Figure S5). Thus, the ID-seq technology allows precise and sensitive multiplexed protein quantification through sequencing antibody-coupled DNA-tags.

Several existing technologies, including quantitative mass spectrometry (MS) based shotgun (phospho-)proteomics and mass cytometry, allow (relative) quantification of protein and signalling dynamics. ID-seq may complement these approaches as a sensitive and accurate targeted (phospho-)proteomics approach with potential for high sample throughput. We constructed a panel of 84 antibody-DNA conjugates covering a wide range of biological processes. These include cell cycle, apoptosis, DNA damage, epidermal self-renewal and differentiation, as well as intracellular signalling status for the EGF, G-protein coupled receptors, calcium signalling, TNFα, TGFβ, Notch, BMP and Wnt pathways. We selected antibodies that showed concentration and epitope dependent signals over 2-3 orders of magnitude (Figure S6 and S7). Moreover, these antibodies displayed high signal-to-background ratios (average ~75x, Figure S8) and high precision among biological replicates in ID-seq experiments ($CV < 0.2$, $n = 14$, Figure 1d), meeting FDA standards for immunoassays. We used this antibody panel (which includes 45 phospho-specific antibodies) to compare ID-seq and quantitative MS shotgun proteomics ($n = 3$, >7,500 detected proteins per sample, ID-seq $n = 6$) of proliferating and differentiated human epidermal keratinocytes. Of the 35 non-phosphorylated proteins covered in our ID-seq experiment, 16 were confidently quantified in our proteomics dataset (≥2 unique peptides, Figure S9a). ID-seq and MS showed comparable coefficients of variation among replicates ($CV < 0.2$, Figure S9b) and trends in protein dynamics during differentiation were generally conserved between the two methods (Figure S9c). Moreover, MS derived absolute protein copy numbers per cell (estimated using an external standard protein mix, see online Methods for details) correlated with the antibody-DNA tag counts obtained from ID-seq (Figure S9d). It is important to note that the ID-seq procedure involves cell fixation and is influenced by antibody affinity and thus differs from absolute protein quantification via MS on unfixed cell lysates. Overall, ID-seq offers a complementary, flexible and streamlined high-throughput approach for (phospho-)proteomics on selected targets.

Exploring (phospho-)protein dynamics with ID-seq.

Primary human epidermal stem cells (keratinocytes) depend on active epidermal growth factor receptor (EGFR) signalling for self-renewal *in vitro* and *in vivo*[13]. We inhibited this pathway using the potent and selective inhibitor AG1478 (10 μM, 48 hrs) to determine whether ID-Seq recapitulates keratinocyte biology. In these experiments we would expect to observe dynamic changes in EGFR signaling activity, as well as in expression of differentiation associated proteins. To analyse the effects of AG1478 treatment on each antibody signal, we developed a generalised linear mixed (glm) model that takes into account the negative binomial distribution of ID-seq count

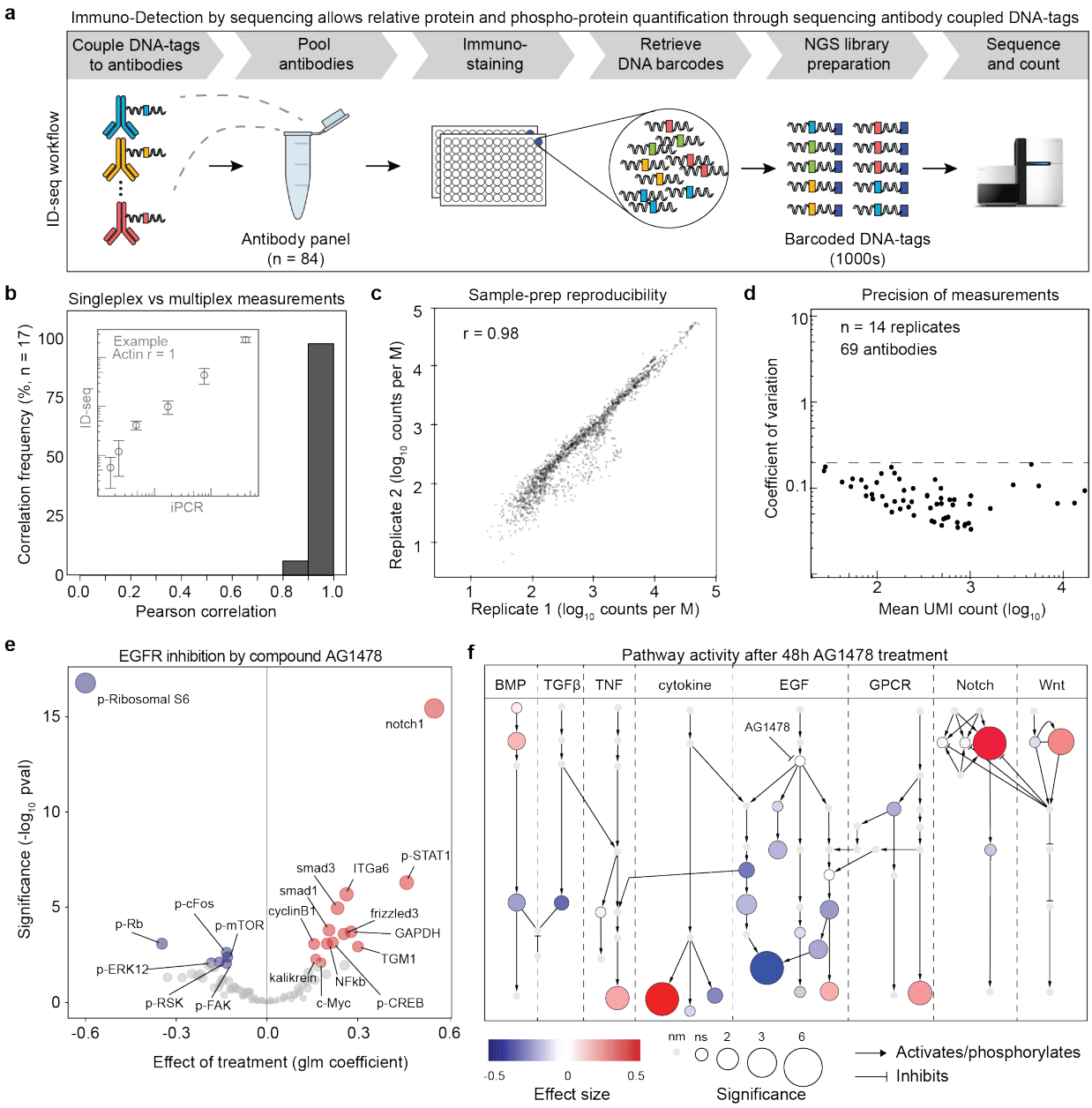

Figure 1. Immuno-Detection by Sequencing (ID-seq) technology development. (**a**) Concept of the ID-seq technology. First, pool DNA-tagged antibodies. Second, perform multiplexed immunostaining on fixed cell populations, and release DNA-tags. Third, barcode the released DNA-tags through a 2-step PCR protocol. Finally, sequence the barcoded DNA-tags via next generation sequencing (NGS) and count barcodes. (**b**) Signals from singleplex epitope detection (via an immuno-PCR measurement) were compared with multiplexed epitope detection using ID-seq. The histogram summarises correlations between 17 immuno-PCR and corresponding ID-seq measurements. Insert panel illustrates an example from the Actin antibody showing signal mean and sd (n = 4). Underlying data for all 17 antibodies can be found in Figure S4. (**c**) Scatter plot indicates the high reproducibility of PCR-based ID-seq library preparation (r = Pearson correlation). Libraries from the same released material were prepared on separate occasions and analysed in different sequencing runs. (**d**) The scatterplot shows the counts (mean) and coefficient of variation from 69 antibody-DNA conjugates (n = 14 biological replicates). Dashed line indicates 20% variation. (**e**) Volcano plot shows the effect (estimate) and significance (-$\log_{10}$ pval) of AG1478 treatment (n = 6), based on the model analysis of ID-seq counts (Supplementary Note 3). Significance (-$\log_{10}$ pval) determines node size. Red nodes show significantly increased ($p < 0.01$) and blue nodes show significantly decreased ($p < 0.01$) (phosphor-)protein levels. (**f**) Pathway overview of ID-seq measurements after 48 hours of AG1478 treatment. Colour indicates the effect size, and node size represents the significance of effect (-$\log_{10}$ pval). Light grey nodes without border indicate not measured (nm) proteins (Supplementary Figure 7 includes (phospho-)protein names).

data and incorporates potential sources of variation (e.g., replicates, batches, sequencing depth). This model derives the effect ('estimate') of treatment on each antibody, followed by a likelihood ratio-test to determine the significance of the effect (Supplementary note 3). We identified 13 increased and 7 decreased (phospho-)proteins upon AG1478 treatment ($p < 0.01$, Figure 1e). Up-regulation of the known differentiation markers transglutaminase 1 (TGM1) and Notch1 confirmed

successful differentiation. Next, we projected the estimates and significance levels of our ID-seq results onto a literature derived signalling network (Figure 1f, see Figure S10a for node identities). As expected, EGFR pathway activity was down regulated as a whole upon AG1478 treatment. We also identified effects on the activity of several other pathways, including the bone morphogenetic protein (BMP) and Notch cascades, which are known players in epidermal biology[14–17]. RT-qPCR analysis revealed that these effects arose from changes in mRNA expression of BMP ligands and Notch receptors (Figure S10b). Activation of the BMP and Notch pathways was confirmed by the induction of the classical downstream target genes Id2 and Hes2, respectively (Figure S10b). Finally, ID-seq faithfully recapitulated known EGFR pathway phosphorylation dynamics[18] during a 1-hour EGF stimulation time-course (Figure S11, n = 8, 5 minute intervals, 96 samples in total). These results illustrate the utility and power of ID-seq and our glm model to uncover (phospho-)protein dynamics among treatment conditions.

Identification of novel kinases involved in epidermal stem cell maintenance.

The extracellular signals involved in epidermal renewal and differentiation include EGF, TGFβ, BMP, Notch ligands and Wnts[14,19–23]. However, the contributions of intracellular effector kinases on renewal and differentiation are not well documented. We applied ID-seq to human epidermal keratinocytes treated with the Published Kinase Inhibitor Set (PKIS), an open-source chemical probe library[24–26] containing ~300 small molecules targeting 225 kinases across all major kinase families in the human proteome (Figure 2a). To determine the effects of kinase inhibition at the molecular level, cells were seeded in 384-well plates, treated with the 300 PKIS compounds for 24 hours and subjected to ID-seq with a panel of 70 antibodies. Replicate screens were highly correlated (R = 0.98) and had low UMI duplicate rates (1.2%), indicating high data quality (Figure S12a-d).

Using our glm model and likelihood ratio statistics we annotated the effect and significance of each measured molecular phenotype (as measured by our antibody conjugates) for each kinase probe. Supervised k-means clustering showed 7 groups of compounds that resulted in distinct patterns across the interrogated antibodies (Figure S13), suggesting that these chemical probes affect the same or related processes in epidermal cells. To some extent this is represented by the hierarchical clustering of the antibodies (left-to-right axis of the heatmap, Figure S13). However, this approach does not intuitively consider overlapping or complementary information captured by the individual read-outs. To incorporate this, we performed principal component analysis (PCA) using the signed log-transformed p-values from our glm model (where the direction of the effect was used as a sign) as input. Essentially, this joins together the phenotypes that explain independent fractions of variation within the data into a single score. We then combined the results of this analysis with the ID-seq data to identify the underlying phenotypes, and thereby the cellular processes, captured by the top 5 principal components through second-order clustering (Figure 2b). This revealed that proteins that are significantly up regulated upon differentiation, including TGM1 and Notch1, correlate strongly to PC2 (Figures 2b,d and S14a,b), indicating that the PC2 score most likely represents cellular differentiation. The same conclusion was reached using the glm model estimates as input for the PCA (data not shown). Keratinocyte differentiation is associated with a G2/M cell cycle arrest *in vivo* and *in vitro*[27–29], which is reflected by an accumulation of the G2 markers cyclin B1 and phosphorylated cdc2 along the PC2 axis (Figure 2e). Thus, the distinct advantage of using this aggregate measure, is that the biological effect of a given treatment is more robust to fluctuations of individual read-outs and therefore represents the underlying process more comprehensively. To validate this, we selected 18 probes that showed high PC2-scores from the PKIS library for colony formation experiments, the gold standard *in vitro* assay for epidermal stem cell activity (Figure S15a, [30]). Automated image analysis was used to quantify colony numbers, as well as the size and TGM1 level of each colony per treatment (n = 3 replicates). Indeed, the results of 14 out of 18 probes were consistent with decreased epidermal stem cell renewal (Figure S15b), authenticating the PC2-score as a *bona fide* measure of differentiation.

We gathered that those molecular phenotypes that display a significant in- or decrease ($p < 0.01$, 1% FDR) between the top and bottom 10% of PC2-ranked probes are likely to distinguish the renewing and differentiated epidermal cell states. This revealed involvement of the Wnt pathway (measured by Fzd3 and phosphorylated-LRP6), MAPK signalling (phospho-p38, phospho-SRC,

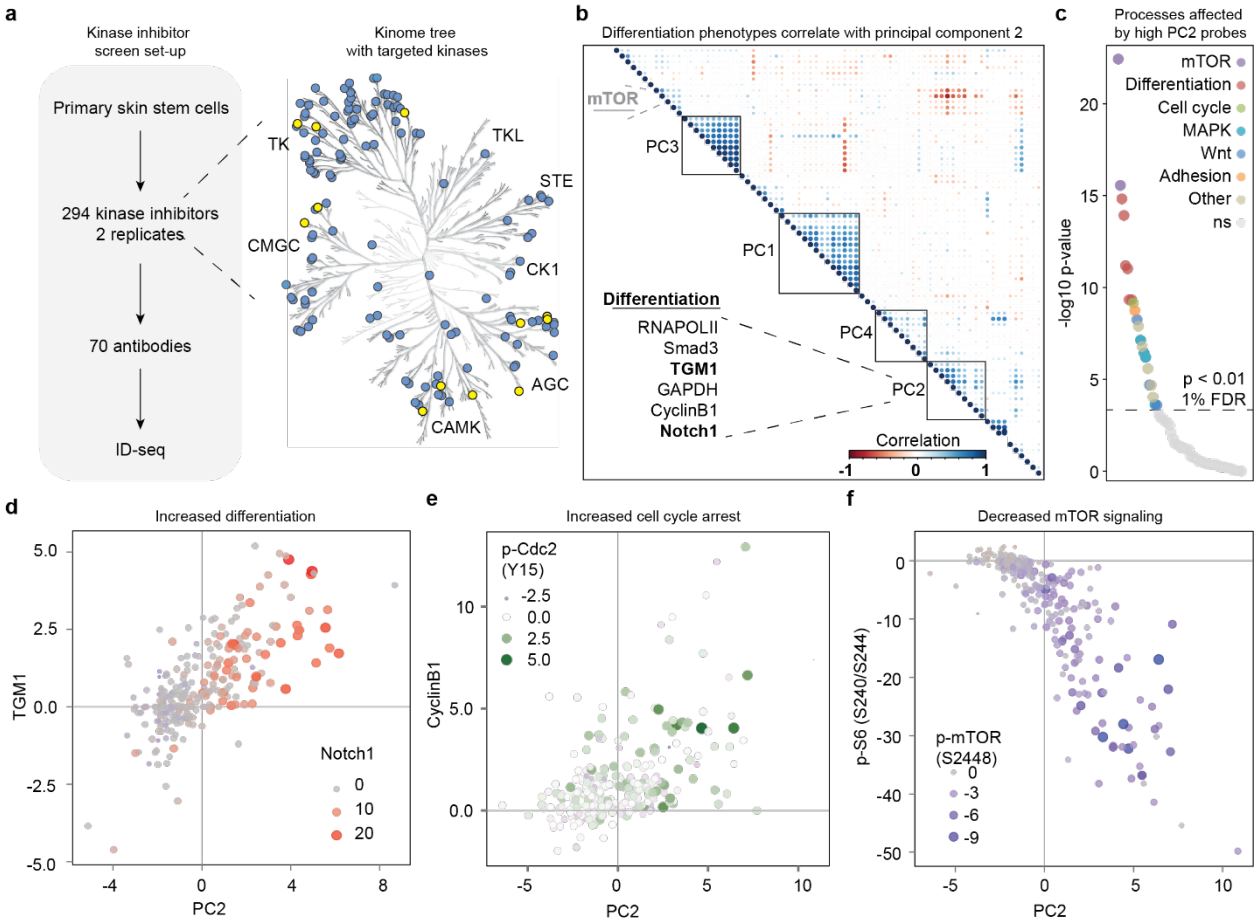

Figure 2. ID-seq screen of Published Kinase Inhibitor Set (PKIS) identifies probes affecting differentiation, cell cycle and mTOR signalling of primary human keratinocytes. (**a**) Schematic overview of screen set-up. Kinase-tree shows kinases targeted by the inhibitory probes (blue) and significantly enriched kinases (yellow, Figure 3a). The probes target all major kinase families (TKL, STE, CK, AGC, CAMK, CMGC and TK). (**b**) To combine probe effects on multiple phenotypes to one measure, we performed principal component analysis using the signed $\log_{10}$ p-values of the ID-seq analysis. Then, we clustered all phenotypes and the top 5 PCs to identify the PC summarising differentiation of the skin stem cells (in bold phenotypes TGM1 and Notch1). (**c**) Significantly ($p < 0.01$, FDR 1%, t-test) affected phenotypes from probes with high PC2 (top 10%) compared to low PC2 (bottom 10%). (**d-f**) The scatterplots illustrate the correlation of the PC2 score with the indicated ID-seq phenotype (signed $\log_{10}$ p-values) of differentiation markers TGM1 and Notch1 (left panel), cell-cycle markers Cyclin B1 and phosphorylated Cdc2 (middle panel) and phosphorylated S6 and mTOR (right panel).

phospho-cFOS and phospho-RSK), integrin-mediated adhesion (phospho-FAK) and the mammalian target-of-rapamycin (mTOR) pathway (phospho-mTOR and phospho-S6) in epidermal renewal and differentiation (Figure 2c and S16). The down regulation of the mTOR pathway activity is particularly striking, and its strong continuous negative association with PC2 suggests an integral role in epidermal biology (Figure 2f). It is important to note that the PKIS probes were tested a low concentration (10 nM) at which the EGFR inhibitor AG1478 did not result in differentiation. This indicates that the compounds resulting in high PC2-scores are potent inducers of differentiation.

We reasoned that kinases that strongly associate with PC2 are likely involved in epidermal stem cell renewal, as their inhibition leads to differentiation. For each of the chemical probes in the PKIS library, the biochemical selectivity and potency towards 225 individual kinases have been determined.[26] We applied outlier statistics to assign a set of inhibitory probes to each of the 225 kinases ($p < 0.01$, supplementary table 1). Subsequent Gene Set Enrichment Analysis (GSEA) identified 13 kinase sets enriched ($p < 0.01$, 1%FDR) in probes leading to an increased PC2-score (i.e. cell differentiation) and of which the corresponding kinase is expressed in keratinocytes (Figures 3a and S17). The potency of the probes to inhibit these 13 kinases scaled with their ability to induce keratinocyte differentiation, implying a role of these kinases in keratinocyte renewal (Figure S18). Indeed, this analysis returned the EGFR as the top hit, reflecting its key importance in epidermal stem cell renewal *in vitro* and *in vivo*. This list also included PRKD3 and FYN, two intracellular

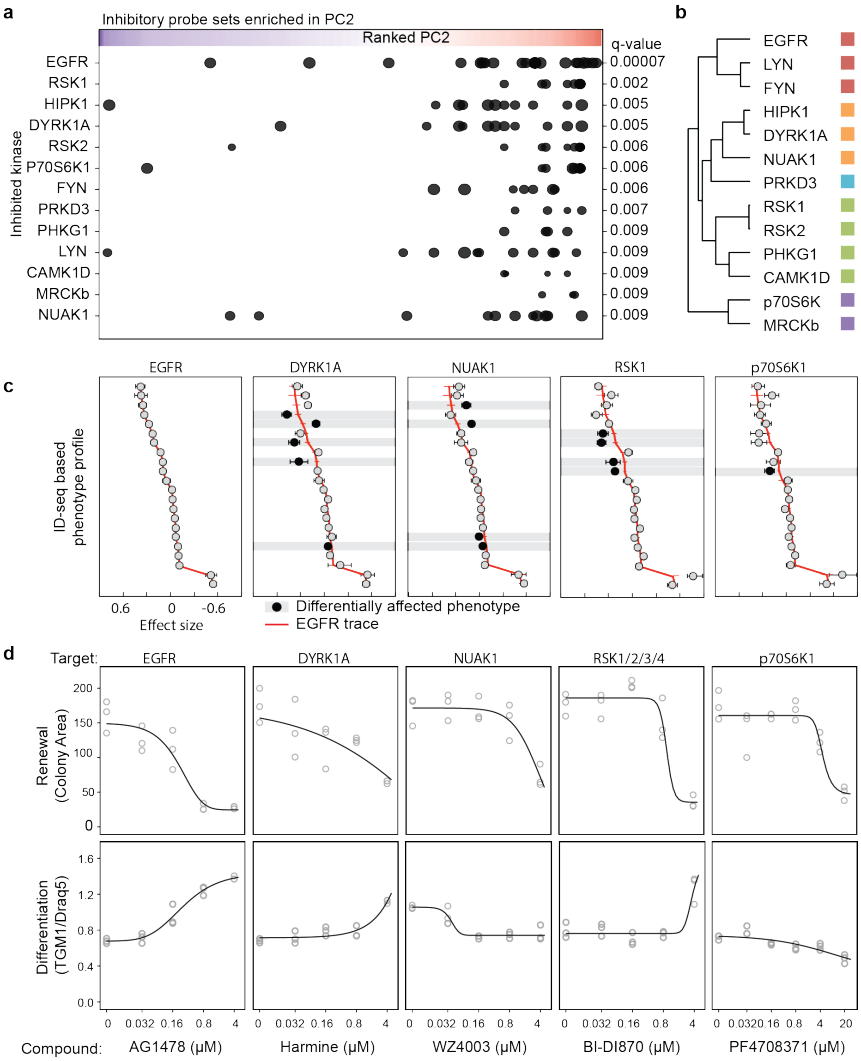

Figure 3. Enrichment analysis of probe-sets in PC2 identifies specific kinases expressed in the proliferative basal layer of the skin and that are needed for keratinocyte renewal or inhibited differentiation. (**a**) Summary plot of gene set enrichment analysis of probes in PC2 shows expressed and enriched kinases ordered according to the significance of enrichment (p-values (FDR) < 0.01, Figure S17). Probes (points in the graph) are ranked according to PC2 (x-axis), and point size shows % of inhibition for the indicated kinase by the probe. (**b**) K-means clustering of mean probe effect on phenotypes per kinase. (**c**) The mean and standard error of phenotypic change of probes inhibiting DYRK1A, NUAK1, RSK1/2/3/4, p70-Ribosomal S6 kinase. Compared to EGFR the other four kinases have a comparable phenotypic profile with few different phenotypes (black nodes, $p < 0.05$, t-test). (**d**) Mean colony area (upper panels) and Integrated Intensity of TGM1/Draq5 immunostaining (bottom panels) with modelled dose-response curves (n=3) of keratinocytes after nine days of growth in the presence of specific kinase inhibitors AG1478, Harmine, WZ4003, BI-DI870 and PF4708371 targeting EGFR, DYRK1A, NUAK1, RSK1/2/3/4, and p70-Ribosomal S6 kinase respectively.

kinases shown to impact epidermal biology[31–35]. Immunohistochemically staining of human skin samples showed that expression of the EGFR, RSK1 and PHKG1 is restricted to cells residing in the epidermal stem cell niche, whereas NUAK1 is expressed throughout the epidermis (Figure S19), consistent with a role in epidermal biology. Taken together, ID-seq identified both known and previously unrecognized kinase effectors of epidermal renewal across 4 major kinase families (Figure 2a, yellow nodes).

The PKIS probe library was designed as a platform for lead discovery and to provide chemical scaffolds for further medicinal chemistry[24–26]. Consequently, each probe may inhibit more than one kinase (Figure S20a). Importantly, the probes that inhibit the EGFR are distinct from those inhibiting the other kinases, indicating that the identification of these 12 kinases did not result from cross-reactivity of the probes for the EGFR (Figure S20b). For each of these kinases, we calculated an average score for each phenotype based on the effect of the probes inhibiting the kinase. Hierarchical clustering and PCA on these profiles identified 5 subgroups among the enriched kinases (Figure 3b and S21a,b). The observation

that the EGFR and its downstream LYN and FYN kinases form a tight cluster indicates that this grouping reflects direct mechanistic relationships. In turn, this predicts that the kinases in the other subgroups may function through mechanisms that are different and potentially independent from the EGFR. If this were indeed the case, we would expect inhibition of these kinases to result in distinctive effects on (subsets of) the interrogated molecular phenotypes. We compared the effects (average model derived estimate +/- SEM) on 20 dynamic phenotypes (as defined in Figure 2c) for these 13 identified kinases. Ranking these phenotypes based on the effect of EGFR inhibition and plotting the data of the other kinases in the same order showed that the overall trend in the measured phenotypes was conserved among the kinases, representing their role in differentiation. Importantly, most of the kinases showed significant deviations ($p < 0.05$) in distinct subsets of phenotypes compared to the EGFR, indicating that they indeed function through discrete mechanisms to regulate epidermal renewal (Figure 3c and S21c).

To verify the importance of the kinases in these distinct subgroups in epidermal self-renewal, we obtained highly selective inhibitors of the EGFR, RSK1-4, p70S6K, NUAK1 and DYRK1A and tested them in *in vitro* colony formation assays (Figures 3d and S22a-d). Quantification of the number, size and differentiation marker levels of colonies at a broad range of inhibitor concentrations revealed that inhibition of the EGFR, RSK1-4 and DYRK1A decreased the self-renewal capacity of the epidermal stem cells (as determined by colony size) and stimulated the expression of the late differentiation marker TGM1. In contrast, p70S6K and NUAK1 inhibition resulted in decreased renewal, but did not increase TGM1 expression. Together, these results highlight the importance of these kinases in epidermal renewal.

Discussion

Cell based phenotypic screens are frequently used in academia and the pharmaceutical industry to identify leads for drug development[36,37]. However, obtaining insight into the molecular mechanism of action of the selected compound is generally time consuming and expensive[38–40]. We developed the Immuno-Detection by sequencing (ID-seq) technology as an approach to facilitate high throughput highly multiplexed molecular phenotyping. We showed that ID-seq is precise, sensitive and applicable to large numbers of samples. We applied ID-seq on primary human epidermal keratinocytes in conjunction with the Published Kinase Inhibitor Set (PKIS) and identified 13 kinases that are important for epidermal stem cell function. Since many of the targeted therapies that are approved or in clinical trials are directed against kinases[40,41], our findings serve as an important resource to identify potential drugs that could lead to severe skin toxicity. The straightforward ID-seq workflow was designed to be compatible with automation for applications in industry and will enable many potential mechanisms of action to be assayed for 100s to 1000s of compounds in a single experiment. In principle, ID-seq can be applied to any cell system, any perturbation and with any validated high-quality antibody, making it a flexible solution for large-scale high-dimensional phenotyping.

Acknowledgements. We would like to thank E. Janssen-Megens and S. van Genesen for technical support and discussions, and Henk Stunnenberg and Ana Pombo for discussions and critical reading of the manuscript. K.M., J.v.B, J.G., M.H and S.T are financially supported by by the Radboud University, the Dutch Organisation for Scientific Research (NWO-VIDI) and the European Union (Marie-Curie Career Integration Grant).

Author contributions. J.v.B. designed the experiments, designed ID-seq barcodes, performed the ID-seq experiments and colony formation assays, wrote the code and R-package, analysed the data. J.G. and M.H. performed the iPCR and RNA-seq experiments, S.T. and P.J. and M.V performed proteomics experiment, J.M. and R.v.d.S. produced all oligo DNA nucleotides, C.A developed the model for ID-seq count data analysis, K.M. conceived and oversaw the study and analysed the data, K.M. and J.v.B. wrote the manuscript with input from all other authors.

Data availability. Used antibodies and oligo sequences are available as Supplementary Table 2 and Supplementary Table 3, respectively. R-package for ID-seq data (pre-)processing is available from https://github.com/jessievb/IDseq. Automated image analysis scripts for colony formation assays are available from https://github.com/jessievb/automated_CFA. Sequencing data from ID-seq experiments is available through GEO[42] Series accession number GSE100135

(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100135)

Competing financial interests. J.M. and R.v.d.S. are employees of Biolegio BV, an SME producing and selling oligo DNA nucleotides. The other authors declare no competing financial interests.

Materials and Methods

Cell culture. Keratinocytes (pooled foreskin strain KNP, Lonza) were expanded as described[29] supplemented with Rock inhibitor (Y-27632, 10 µM). After expansion of the keratinocytes on feeders, the cells were grown for 1-3 days on keratinocyte serum-free medium (KSFM) with supplements (bovine pituitary extract (30 µg/mL) and EGF (0.2 ng/mL, Gibco) in a 96 or 384 wells plate at ~10.000 cells per well. Before 48 hours AG1478 treatment (10 µM), the cells were cultured for 48 hours in a 96 wells plate for the ID-seq analysis, or in 15 cm dish for the proteomics experiment. Before EGF stimulation (100 ng/ml), keratinocytes were grown for three days on KSFM with and one day without supplements. After starvation, every 5 minutes EGF was added (n = 8). For the PKIS screen, cells were seeded on a 384 wells plate and grown for 24 hours in KSFM with supplements, followed by 24-hour treatment with PKIS compounds (10 nM) or AG1478 (10 nM) or DMSO.

Conjugation of antibodies to dsDNA. Antibodies and dsDNA were functionalised and conjugated as described[43]. See Supplementary Table 2 for a list of antibodies. In short, antibodies were functionalized with NHS-s-s-PEG4-tetrazine in a ratio of 1:10 in 50 mM borate buffered Saline pH 8.4 (150 mM NaCl). Then, N3-dsDNA was produced and functionalized with DBCO-PEG12-TCO (Jena Bioscience). See Supplementary Table 2 for a list of oligo sequences. Finally, purified functionalized antibodies were conjugated to purified functionalized DNA by 4-hour incubation at room temperature in borate buffered saline pH 8.4 in a ratio of 4:1 respectively. The reaction was quenched with an excess of 3,6-diphenyl tetrazine. After pooling, the conjugates were incubated with ProtA/G beads in BBS overnight. After thorough washes with PBS, the conjugates were eluted from the beads with 0.1 M citrate pH2.3 and immediately neutralized with TrisHCl pH 8.8. Subsequently, a buffer exchange into PBS (pH7.4) was performed using two Zeba-spin desalting columns. The size of the eluted DNA was checked on an agarose gel, confirming that all unconjugated DNA was removed using this purification approach.

Immuno-staining with Antibody-DNA conjugates and release of DNA tags. Keratinocytes were fixed with 4% paraformaldehyde in PBS for 15 minutes at RT, washed three times with PBS and stored at 4 °C (up to 3 - 4 days before further use). Then, cells were permeabilized and blocked for 30 minutes using 0.5x protein free blocking buffer (Thermo Fisher) in PBS with 0.1 % Triton and 200 ng/ml single strand Salmon Sperm DNA (sssDNA). Blocking the cells and wells with sssDNA is crucial to suppress background binding of the Ab-DNA conjugates[43]. Then, cells were incubated with conjugated antibodies in the same buffer at 0.1 µg/mL antibody, at 4 °C overnight. After immunostaining with conjugates, the cells were thoroughly washed with PBS (3x short, 3x 15 min, 3x short). Then, release buffer was freshly prepared (10 mM DTT in borate buffered saline pH 8.4). Cells were incubated with 20 to 50 µl of release buffer depending on the plate type and well-size and incubated for 90 minutes at RT, with careful mixing (on a vortex) every 30 minutes. Released DNA barcodes were collected and stored at -20 °C.

Sample barcoding and sequencing library preparation. To barcode the released DNA-tags from each cell population (See Supplementary Note 1 for sequence design), a 25 µl PCR was performed per sample containing 8-15 µl sample with released DNA-tags, 0.2 mM dNTPs, 1 ul PFU polymerase, 1x PFU buffer (20mM Tris-HCl pH8.8, 2mM $MgSO_4$, 10mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% triton, 0.1 mg/ml BSA), spike-in DNA barcodes, forward primer (AATGATACGGCGACCACCG, Biolegio) and a well specific reverse primer (Supplementary Table 3, Supplementary Scheme 1b). In a 96 wells PCR machine (T100 Thermal Cycler, Biorad) the following program was used: 1) 3 min at 95 °C, 2) 30 sec at 95 °C, 3) 30 sec at 60 or 54 °C, 4) 30 sec at 72 °C, 5) repeat 3&4 nine times, 6) 5 min at 72 °C, 7) ∞ 12 °C. Then, all well-specific labelled DNA barcodes from one plate were pooled to 1 sample. This sample was purified using a PCR purification column (Qiagen) according to manufacturer's protocol. Samples were eluted with 30 µl nuclease-free water. To remove any residual primers, samples were treated with Exonuclease I in 1x PFU buffer for 30 min at 37

°C. After inactivation for 20 min at 80 °C, another 25 µl PCR reaction was prepared with 15 - 17 µl sample, 0.2 mM dNTPs, x 1 µl PFU polymerase, 1x PFU buffer, forward primer (AATGATACGGCGACCACCG, Biolegio) and a sample specific reverse primer with Illumina index barcode and adapter sequence (Supplementary Table 3, Supplementary Scheme 1d). The same program as PCR reaction (I) was used, and reactions were purified over a PCR purification column (Qiagen). All PCR reactions were then incubated for 45 minutes with one µl Exonuclease I to remove residual primers. The PKIS screen samples were further size-selected using size selection columns (Zymo, according to manufacturer's protocol) for fragments > 150 bp. Finally, all samples were purified over PCR purification mini-elute column (Qiagen) and eluted in 10 µl elution buffer. Final sequencing samples were run on a 2% agarose gel (0.5x TBE) with 10× SYBR Green I (Life Technologies) and scanned on a Typhoon Trio+ machine (GE Healthcare), or analysed with the 2100 Bioanalyzer (Agilent) to confirm the size of the DNA fragments (expected size around 185 bp).

ID-seq data analysis. Sequence data from the NextSeq500 (Illumina) was demultiplexed using bcl2fastq software (Illumina). The quality of the sequencing data was evaluated using a FastQC tool (version 0.11.4 and 0.11.5, Babraham Bioinformatics). Then, all reads were processed using our dedicated R-package (IDSeq, Supplementary Note 2). In short, the sequencing reads were split using a common "anchor sequence" identifying the position of the UMI sequence, Barcode 1 (antibody specific) and Barcode 2 (well specific) sequence. After removing all duplicate reads, the number of UMI sequences were counted per barcode 1 and 2. Finally, barcode 1 ("antibody)" and barcode 2 ('well') sequences were matched to the corresponding antibody and well information.

Using R-package *DESeq2*[44], we calculated normalisation factors (Estimated Size factor) to account for differences in sequencing depth per sample. Using *lme4,* we analysed the effect of a specific condition using a linear mixed effect model (Supplementary Note 3).

Immuno-PCR experiments. The immuno-PCR experiments were performed as described previously in our paper on antibody-DNA conjugates[43]. In short, each antibody was conjugated to dsDNA, and used in an immunostaining as described. DNA was released using 10 mM DTT in BBS pH8.4 and measured by quantitative PCR using iQTM SYBR Green Supermix on CFX 96 machine. The 2^-Ct values were used to calculate the mean signal and standard deviation from 4 biological replicates. The Pearson correlation between these immuno-PCR and the multiplexed ID-seq signal was calculated using the mean.

Proteomics. Cells were harvested, washed, snap frozen and stored at -80 °C until lysis and mass spectrometry analysis. Induction of differentiation was validated by qPCR (data not shown) before lysis. Cells were lysed using lysis buffer (4% SDS, 100 mM TrisHCl pH 7.6, 100mM DTT) and by boiling for 3 minutes at 95 °C. DNA was sheared using sonication, 5 cycles 30 sec ON/30 sec OFF (high). The samples were centrifuged for 5 minutes at 16.000 x *g*, 4 °C and the supernatant was taken for protein quantification with the PierceTM BCA Protein Assay Kit (Thermo Scientific).

For the generation of tryptic peptides, we applied filter-aided sample preparation (FASP)[45]. To be able to absolutely quantify the proteins in the samples we used a standard range of proteins (UPS2-1SET, Sigma) which we spiked into one of the samples (3,3 µg in sample equivalent to 100.000 cells)[46]. To obtain deep-proteome samples were fractionated using strong anion exchange (SAX), collecting fractions of the flow through and elutions at pH 11, 8, 5 and 2 of Britton & Robinson (BR) buffer. Samples were desalted and concentrated using C18 stage-tips.

The peptide samples were separated on an Easy nLC 1000 (Thermo Scientific) connected online to a Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer. A 240 min acetonitrile gradient (5-23%, 8-27%, 9-30%, 11-32% and 14-32% for FT, pH 11, 8, 5 and 2, respectively) was applied to the five fractions. MS and MS/MS spectra were recorded in a Top speed modus with a run cycle of 3s. MS/MS spectra were recorded in the Ion trap using Higher-energy Collision Dissociation (HCD) fragmentation.

To analyse the raw mass spectrometry data we used MaxQuant (version 1.5.1.0, database: Uniprot_201512\HUMAN)[47] with default setting and the "match between runs" and "iBAQ" algorithms enabled. We filtered out reverse hits and imputed missing values using Perseus (default settings, MaxQuant software package).

RNA expression levels of kinases using CEL-seq2 mRNA quantification.

mRNA sequencing was performed according to the CELseq2 protocol[48] with adaptations. Reverse transcription was performed in 2 µl reactions overlaid with 7 µl Vapor-Lock (Qiagen) using Maxima H minus reverse transcriptase (ThermoFisher) and 100 pg purified RNA per sample. Primer sequences were adapted to allow sequencing of 63 nucleotides of mRNA in read 1 and 14 nucleotides in read 2, comprising the sample barcode and UMI. Reverse transcription primer: 5'GCCGGTAATACGACTCACT-ATAGGGGTTCAGACGTGTGCTCTTCCGATC TNNNNNNNN[6ntsamplebarcode]TTTTTTTTTT TTTTTTTTTTTTTTV3', random-octamer-primer for reverse transcription of amplified RNA: 5'CACGACGCTCTTCCGATCT-NNNNNNNN3', library PCR Primers: 5'AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACG-ACGCTCTTCCGATCT3' and 5'CAAGCAGAAGACGGCATACGAGAT[6ntindex]GTGACTGGAGTTCAGACGTGTG-CTCTTCCGATC3'. Sequencing was performed using the NextSeq500 from Illumina.

Colony formation assay. In 6 wells plate, 200.000 feeders (J2-3T3) cells were seeded in DMEM (with 10% BS and 1% pen/strep). After one day, cells were inactivated by 3-hour treatment with mitomycin C. After thorough washes with DMEM, 1000 keratinocytes were seeded into each well[29]. The following day, treatment was started (day 0) by refreshing medium and addition of 0.01 µM compound (or as indicated) or vehicle DMSO. Cells were grown in the presence of compounds for eight more days, and the medium was refreshed on days 2 and 5. Rocki was present until day 2 of the treatment. Cells were fixed, stained with TGM1 specific antibodies and scanned as described before [49]. Raw images from the LiCor Odessey system were processed with CC Photoshop and CellProfiler with consistent settings. Data obtained via automatic counting and imaging analysis via CellProfiler was analysed and visualized in the R programming language.

References

1. Kabiraj, A., Gupta, J., Khaitan, T. & Bhattacharya, P. T. Principle and Techniques of Immunohistochemistry – a Review. *Int J Biol Med Res* **6,** 5204–5210 (2015).
2. Giepmans, B. N. G. The Fluorescent Toolbox for Assessing Protein Location and Function. *Science (80-.).* **312,** 217–224 (2006).
3. Ryazantsev, D. Y., Voronina, D. V. & Zavriev, S. K. Immuno-PCR: achievements and perspectives. *Biochem.* **81,** 1754–1770 (2016).
4. Nong, R. Y., Gu, J., Darmanis, S., Kamali-Moghaddam, M. & Landegren, U. DNA-assisted protein detection technologies. *Expert Rev. Proteomics* **9,** 21–32 (2012).
5. Shahi, P., Kim, S. C., Haliburton, J. R., Gartner, Z. J. & Abate, A. R. Abseq : Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Nat. Publ. Gr.* 1–12 (2017). doi:10.1038/srep44447
6. Dezfouli, M., Vickovic, S., Iglesias, M. J., Schwenk, J. M. & Ahmadian, A. Parallel barcoding of antibodies for DNA-assisted proteomics. *Proteomics* **14,** 2432–2436 (2014).
7. Darmanis, S. *et al.* ProteinSeq: high-performance proteomic analyses by proximity ligation and next generation sequencing. *PLoS One* **6,** e25583 (2011).
8. Bandura, D. R. *et al.* Mass cytometry: Technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* **81,** 6813–6822 (2009).
9. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332,** 687–96 (2011).
10. Bodenmiller, B. *et al.* Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat. Biotechnol.* **30,** 858–67 (2012).
11. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11,** 417–22 (2014).
12. Buggenum, J. A. G. L. van *et al.* A covalent and cleavable antibody-DNA conjugation strategy for sensitive protein detection via immuno-PCR. *Sci. Rep.* **6,** 22675 (2016).
13. Nanba, D., Toki, F., Barrandon, Y. & Higashiyama, S. Recent advances in the epidermal growth factor receptor/ligand system biology on skin homeostasis and keratinocyte stem cell regulation. *J. Dermatol. Sci.* **72,** 81–6 (2013).
14. Kolev, V. *et al.* EGFR signalling as a negative regulator of Notch1 gene transcription and function in proliferating keratinocytes and cancer. *Nat Cell Biol* **10,** 902–911 (2008).
15. Lewis, C. J. *et al.* Bone Morphogenetic Protein Signaling Suppresses Wound-Induced Skin Repair by Inhibiting Keratinocyte Proliferation and Migration. *J. Invest. Dermatol.* **134,** 827–837 (2013).
16. Park, G. T. Bone morphogenetic protein-2 (BMP-2) transactivates Dlx3 through Smad1 and Smad4: alternative mode for Dlx3 induction in mouse keratinocytes. *Nucleic Acids Res.* **30,** 515–522 (2002).
17. Phillips, M. a, Qin, Q., Hu, Q., Zhao, B. & Rice, R. H.

Arsenite suppression of BMP signaling in human keratinocytes. *Toxicol. Appl. Pharmacol.* **269,** 290–6 (2013).

18. Olsen, J. V *et al.* Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **127,** 635–648 (2006).

19. Watt, F. M., Estrach, S. & Ambler, C. A. Epidermal Notch signalling: differentiation, cancer and adhesion. *Curr. Opin. Cell Biol.* **20,** 171–179 (2008).

20. He, W., Cao, T., Smith, D. a, Myers, T. E. & Wang, X. J. Smads mediate signaling of the TGFbeta superfamily in normal keratinocytes but are lost during skin chemical carcinogenesis. *Oncogene* **20,** 471–83 (2001).

21. McDonnell, M. A., Law, B. K., Serra, R. & Moses, H. L. Antagonistic effects of TGFbeta1 and BMP-6 on skin keratinocyte differentiation. *Exp. Cell Res.* **263,** 265–73 (2001).

22. Shi, Y. *et al.* Wnt and Notch signaling pathway involved in wound healing by targeting c-Myc and Hes1 separately. *Stem Cell Res. Ther.* **6,** 120 (2015).

23. Lim, X. & Nusse, R. Wnt Signaling in Skin Development, Homeostasis, and Disease. *Cold Spring Harb. Perspect. Biol.* **5,** a008029–a008029 (2013).

24. Knapp, S. *et al.* A public-private partnership to unlock the untargeted kinome. *Nat. Chem. Biol.* **9,** 3–6 (2012).

25. Dranchak, P. *et al.* Profile of the GSK Published Protein Kinase Inhibitor Set Across ATP-Dependent and-Independent Luciferases: Implications for Reporter-Gene Assays. *PLoS One* **8,** e57888 (2013).

26. Elkins, J. M. *et al.* Comprehensive characterization of the Published Kinase Inhibitor Set. *Nat. Biotechnol.* **34,** 95–103 (2016).

27. Gandarillas, a, Davies, D. & Blanchard, J. M. Normal and c-Myc-promoted human keratinocyte differentiation both occur via a novel cell cycle involving cellular growth and endoreplication. *Oncogene* **19,** 3278–3289 (2000).

28. Zanet, J. *et al.* A mitosis block links active cell cycle with human epidermal differentiation and results in endoreplication. *PLoS One* **5,** (2010).

29. Gandarillas, A. & Watt, F. M. c-Myc promotes differentiation of human epidermal stem cells. *Genes Dev.* **11,** 2869–2882 (1997).

30. Barrandon, Y. & Green, H. Three clonal types of keratinocyte with different capacities for multiplication. *Proc. Natl. Acad. Sci.* **84,** 2302–2306 (1987).

31. Calautti, E., Missero, C., Stein, P. L., Ezzell, R. M. & Paolo, G. D. Fyn Ty . Rosine Kinase Is Involved in Keratmocyte Differentiation Control. 2279–2291 (1995). doi:10.1101/gad.9.18.2279

32. Calautti, E. *et al.* Fyn tyrosine kinase is a downstream mediator of Rho/PRK2 function in keratinocyte cell-cell adhesion. *J. Cell Biol.* **156,** 137–148 (2002).

33. Fenton, S. E. & Denning, M. F. FYNagling divergent adhesive functions for Fyn in keratinocytes. *Exp. Dermatol.* **24,** 81–85 (2015).

34. Efimova, T. Novel Protein Kinase C Isoforms Regulate Human Keratinocyte Differentiation by Activating a p38delta Mitogen-activated Protein Kinase Cascade That Targets CCAAT/Enhancer-binding Protein alpha. *J. Biol. Chem.* **277,** 31753–31760 (2002).

35. Matsui, M. S., Chew, S. L. & DeLeo, V. A. Protein Kinase C in Normal Human Epidermal Keratinocytes During Proliferation and Calcium-Induced Differentiation. *J. Invest. Dermatol.* **99,** 565–571 (1992).

36. Michelini, E., Cevenini, L., Mezzanotte, L., Coppa, A. & Roda, A. Cell-based assays: fuelling drug discovery. *Anal. Bioanal. Chem.* **398,** 227–238 (2010).

37. Swinney, D. C. The contribution of mechanistic understanding to phenotypic screening for first-in-class medicines. *J. Biomol. Screen.* **18,** 1186–1192 (2013).

38. Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10,** 188–195 (2011).

39. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11,** 191–200 (2012).

40. Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10,** 507–519 (2011).

41. Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9,** 28–39 (2009).

42. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–210 (2002).

43. Buggenum, J. A. G. L. van *et al.* A covalent and cleavable antibody-DNA conjugation strategy for sensitive protein detection via immuno-PCR. *Sci. Rep.* **6,** 22675 (2016).

44. Love, M. I., Anders, S. & Huber, W. *Differential analysis of count data - the DESeq2 package*. *Genome Biology* **15,** (2014).

45. Wiśniewski, J. R., Ostasiewicz, P. & Mann, M. High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. *J. Proteome Res.* **10,** 3040–3049 (2011).

46. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2,** 1896–1906 (2007).

47. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* **26,** 1367–1372 (2008).

48. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17,** 77 (2016).

49. Mulder, K. W. *et al.* Diverse epigenetic strategies interact to control epidermal differentiation. *Nat. Cell*

Biol. **14,** 753–63 (2012).