Diverse epigenetic strategies interact to control epidermal differentiation

Diverse epigenetic strategies interact to control epidermal differentiation

Klaas W. Mulder[a], Xin Wang[b], Carles Escriu[a], Yoko Ito[c], Roland F. Schwarz[b], Jesse Gillis[d], Gabor Sirokmany[e], Giacomo Donati[a], Santiago Uribe-Lewis[c], Paul Pavlidis[d], Adele Murrell[c], Florian Markowetz[b] and Fiona M. Watt[a, e]

[a] Epithelial cell biology group, [b] Computational biology research group, [c] Genomic imprinting group, Cancer Research UK Cambridge Research Institute, Robinson Way, CB2 0RE, Cambridge, United Kingdom. [d] Centre for High-throughput Biology, 2125 East Mall, BC V6T 1Z4, Vancouver, Canada. [e] The Wellcome Trust Centre for Stem Cell Research, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN United Kingdom.

It is becoming clear that interconnected functional gene networks, rather than individual genes, govern stem cell self-renewal and differentiation. To identify epigenetic factors that impact on human epidermal stem cells we performed siRNA based genetic screens for 332 chromatin modifiers. We developed a Bayesian mixture model to predict putative functional interactions between epigenetic modifiers that regulated differentiation. We discovered a network of genetic interactions involving EZH2, UHRF1 (both known to regulate epidermal self-renewal), ING5 (a MORF complex component), BPTF and SMARCA5 (NURF complex components). Genome-wide localisation and global mRNA expression analysis revealed that these factors impact two distinct but functionally related gene sets, including integrin extracellular matrix receptors that mediate anchorage of epidermal stem cells to their niche. Using a competitive epidermal reconstitution assay we confirmed that ING5, BPTF, SMARCA5, EZH2 and UHRF1 control differentiation under physiological conditions. Thus, regulation of distinct gene expression programs through the interplay between diverse epigenetic strategies protects epidermal stem cells from differentiation.

Historically, reductionist approaches have been used to pinpoint the function of individual components in cellular systems[1]. Advances in genomics, including RNA interference screens, have enabled large-scale parallel interrogation of these elements[2]. However, focussing solely on individual 'hits' without reference to how they interact cannot reveal the complex nature of cellular decision-

making. This is especially relevant when factors that, individually, influence a particular phenotype turn out to cooperate even in the absence of a physical interaction. These functional or genetic interactions[3,4] may be of particular importance in epigenetic regulation of differentiation programs.

The epidermis is a multi-layered epithelium, maintained by stem cells residing in the basal-layer[5,6]. The onset of terminal differentiation involves cell cycle withdrawal and detachment from the basement membrane as a result of changes in cell-extracellular matrix interactions, cell-cell adhesion and the cytoskeleton[5,6]. The simple organisation of the interfollicular epidermis and the fact that its stem cell compartment can be maintained *in vitro* make it an ideal system for studying the complexity underlying cell fate choices[7].

Several transcription factors that regulate epidermal stem cells have already been identified. For instance, p63, mainly ΔNp63, regulates self-renewal and tissue assembly[8,9], whereas AP1 transcription factors are required for terminal differentiation[10]. Additional regulation takes place at the epigenetic level: polycomb group 1 and 2, DNA methylation and histone H4K20 monomethylation are involved in controlling stem cell renewal[11-17]. Nevertheless, a comprehensive survey of the epigenetic mechanisms involved, and how they cooperate, has not been carried out. Here, we report such an approach and identify chromatin-factor complexes that target distinct, yet functionally overlapping, gene sets to maintain the undifferentiated state.

RESULTS

SiRNA screens identify chromatin-associated factors regulating epidermal differentiation

To determine the function of 332 known and predicted chromatin-factors (Supplementary Table 1) in primary human epidermal stem cells (keratinocytes), we performed siRNA screens under five conditions in triplicate. Vehicle treated cells were compared with cells stimulated to differentiate with foetal calf serum, the EGF receptor inhibitor AG1478, bone morphogenetic proteins 2 and 7 (BMP2/7), or AG1478 and BMP2/7 in combination (Fig. 1a). Serum, AG1478 and BMP2/7 are known to stimulate differentiation via distinct cellular signalling pathways[18-20]. Differentiation was quantified by antibody-based detection of transglutaminase I (TG1), the key enzyme that catalyses assembly of the epidermal cornified envelope[21,22] (Fig. 1a and Supplementary Fig. 1a,b). TG1 antibody specificity was demonstrated using two independent siRNAs targeting TG1 (Supplementary Fig. 1a).

Several complementary approaches confirmed that the readout of differentiation in our system reflected the programme of terminal differentiation that occurs in human interfollicular epidermis. Three additional terminal differentiation markers, involucrin (upper spinous layer marker), periplakin and envoplakin (lower spinous layer markers), were upregulated, the kinetics of induction depending on the stimulus applied (Fig. 1b and Supplementary Fig. 1c). Genome-wide mRNA

profiling confirmed upregulation of genes associated with terminal differentiation and downregulation of genes associated with cell proliferation and adhesion, as expected (Fig. 1c and Supplementary Fig. 1d). We also demonstrated that treatment with AG1478 and BMP2/7, alone or in combination, reduced clonal growth, a quantitative readout of stem cell abundance[7] (Supplementary Fig. 1e). Thus, the treatments we chose induced *bona fide* differentiation responses.

Screen reproducibility within and between biological replicates was excellent (Pearson correlation 0.94 and 0.91, respectively) and no contribution of siRNA plate position was observed (Supplementary Fig. 1f-h). A Z-score transformation accurately represented the raw data (Supplementary Fig. 1f,h), allowing us to compile all experiments and treatment groups into a single high quality dataset. RT-qPCR and Western blotting confirmed that effective knock-down was achieved (Supplementary Fig. 2 a-d). SiRNA pool deconvolution experiments indicated that false-negative and false-positive rates were below 10% (Supplementary Fig. 2b-d).

Identification of protein complexes that regulate epidermal differentiation

We subjected our dataset to unsupervised two-way hierarchical clustering on the basis of differentiation stimuli that elicited similar responses (left to right) and chromatin factors that, upon knock-down, had similar effects on differentiation (top to bottom) (Fig. 2a). As expected, replicates of individual treatments clustered together and vehicle treated samples clustered separately from

samples in which differentiation had been induced. The clustering of differentiation stimuli reflects differences in the relative contribution of chromatin factors under specific conditions.

Examination of how individual chromatin factors clustered revealed groups of genes with similar effects, knock-down either stimulating (purple) or inhibiting (orange) differentiation. There was a differential requirement for some chromatin modifiers in responding to different differentiation stimuli, as exemplified by a cluster containing AIRE, Jmjd2A, HDAC8 and MBTD1 (Fig 2a). However, there were also groups of factors that had similar effects under all conditions. We found factors whose knock-down inhibited differentiation, such as *BRD4* and *CHD4* (Fig. 2a, orange cluster), and factors whose knock-down stimulated differentiation (Fig. 2a, purple cluster). Induction of differentiation after silencing in vehicle treated cells may indicate a role in stem cell renewal, as observed for *EZH2* and *UHRF1*[13,16,17] (Fig. 2a).

Gene set enrichment analysis (GSEA) of the whole dataset identified six protein complexes (as opposed to individual proteins) whose components, when silenced, consistently resulted in an increase (purple) or decrease (orange) of keratinocyte differentiation (Fig. 2b). This suggests that the NURF[23], MORF[24,25] and LSD1[25] complexes are important to keep epidermal stem cells in an undifferentiated state. The interaction between BPTF and SmarcA5 within the NURF complex[23] was confirmed by co-immunoprecipitating the endogenous

proteins (Fig. 2c). Silencing both BPTF and SmarcA5 expression in primary keratinocytes led to a greater reduction in clonal growth than silencing each gene individually, indicating that BPTF and SmarcA5 interact functionally as well as physically (Fig. 2d).

Disrupting components of the BRD4[26], NuRD[25] or SWI/SNF[25] complexes led to impaired differentiation (Fig. 2b). In line with these results, the NuRD and SWI/SNF complexes have previously been shown to function in epidermal differentiation in mouse models[27,28]. Thus, our siRNA screen recovers factors known to be important for epidermal biology *in vitro* and *in vivo* and identifies several new players.

Identification of a network of genetic interactions among ING5, SmarcA5, BPTF, EZH2 and UHRF1

In model organisms, such a yeast, genetically interacting genes often display similar phenotypes[29]. Therefore, we anticipated that chromatin-factors displaying highly similar effects in all tested conditions might be functionally connected. We developed a Bayesian mixture-modelling approach to predict such interactions among all 54,780 gene-pairs covered by our screen (X.W., M. Castro, K.W.M, and F.M., *submitted*, Fig. 3a and Materials and Methods). This yielded 837 statistically significant predicted positive interactions among 158 genes (*SNR*>10, Fig. 3a).

To identify groups of genes working in concert, we searched the full network for significantly connected modules. One module contained ING5, BRD1, BPTF and SmarcA5 (components of the MORF and NURF complexes, respectively), EZH2 and UHRF1 (Fig. 3a). Several of these genes had been identified independently by GSEA (Fig. 2b). We investigated this subnetwork further because it functionally connects two factors known to regulate keratinocyte self-renewal (EZH2 and UHRF1)[13,16,17] with factors that have not been implicated previously (ING5, SmarcA5 and BPTF).

We experimentally validated the role of the five factors in regulating differentiation. Pool deconvolution experiments confirmed our screen results with at least 2 independent siRNAs for each gene and thus argue against off-target effects (Supplementary Fig. 2b). Expression of these genes was down regulated upon differentiation, similar to the basal cell markers integrin α6 and β1, consistent with a role in stem cell renewal (Fig. 3b). Finally, silencing of each factor resulted in increased levels of IVL, TG1 and PPL mRNAs, indicating a *bona fide* differentiation response (Fig. 3c). The lack of IVL induction after Ezh2 silencing might be due to compensation by Ezh1 in regulating the epidermal differentiation complex (EDC), a cluster of genes, including IVL, on chromosome 1q21[12,13].

To determine whether components of the subnetwork display genetic interactions[3,30], we performed combinatorial knock-down experiments using two

independent sets of siRNAs targeting ING5, BPTF, SmarcA5, EZH2 and UHRF1. Knock-down efficiencies were systematically monitored and differentiation was quantified measuring TG1 mRNA levels by RT-qPCR (Supplementary Fig. 3a-c). Eight out of the ten possible combinations resulted in TG1 levels that significantly differed from the expected values (calculated from the effects of the individual knock-downs). This is represented by divergence from the diagonal in Fig. 3d and demonstrates true genetic interactions between these genes. Since SmarcA5 can reside in more than one protein complex, the genetic interactions of SmarcA5 and BPTF with the other network components are not identical, despite their physical interaction. In contrast, five unrelated control genes showed no, or marginally significant, interactions (Supplementary Fig. 4a-c). We conclude that our computational approach successfully identified a functional subnetwork enriched in strong, true genetic interactions and that ING5, SmarcA5, BPTF, EZH2 and UHRF1 jointly control epidermal stem cell differentiation.

Genome-wide identification of subnetwork target genes

The components of the self-renewal subnetwork we identified can be classified as exemplars of distinct underlying mechanisms, or epigenetic strategies. The MORF histone acetyl-transferase (HAT) complex, including ING5, represents a 'writer' of the histone code[24]. BPTF contains domains recognising both acetylated and methylated histones and is a modification 'reader'[31]. SmarcA5 belongs to the SNF2/RAD54 family of ATP-dependent chromatin remodelling factors responsible for physical movement and displacement of nucleosomes[32]. EZH2 is

involved in generating bivalent chromatin domains[13,17,33]. Finally, UHRF1 promotes keratinocyte self-renewal by regulating maintenance of DNA methylation[16].

To identify the genes regulated by subnetwork components, we first performed genome-wide analysis of ING5 binding by chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq) (Fig. 4a and Supplementary Table 2). We compared the data to localisation of a variety of individual histone modifications, RNA polymerase II, H3K27me3 and methylation of promoter regions (as a proxy for EZH2 and UHRF1 function, respectively) in self-renewing keratinocytes[16,34].

Specific and reproducible ING5 binding was primarily detected within 1 kb of the transcription start site (TSS) of transcribed genes marked by H3K4me3, H3K9ac, H3K36me3 and RNA polymerase II (Fig. 4a-d and Supplementary Fig. 5a-e). This is in agreement with biochemical studies showing binding of ING5 to H3K4me3 marked histone tails, and the capacity of the MORF complex to acetylate histones[24,35]. Moreover, the strength of the ING5 signal correlated with mRNA expression levels (Fig. 4c). Genes with high ING5 occupancy showed low levels of H3K27 tri-methylation, indicating that the MORF complex and EZH2 globally target distinct gene-sets (Fig. 4b,d). In contrast, there was no obvious correlation of ING5 with CpG methylation (Fig. 4b).

No ING5, SmarcA5 and BPTF ChIP signal was obtained on the transglutaminase I gene or genes in the EDC (Fig 4a and data not shown). Therefore control of differentiation by these chromatin factors is not through direct suppression of the EDC or TGM1, but likely involves regulation of genes important for maintaining the stem cell state.

Subnetwork components target distinct gene sets with overlapping functions

To identify the functionally relevant targets of ING5, DNA methylation and EZH2 dependent bivalent chromatin domains, we performed microarray profiling of proliferating and differentiated keratinocytes (Fig. 1c and Supplementary Fig. 1d). Genes differentially expressed during keratinocyte differentiation were significantly enriched in ING5 binding and CpG methylation ($P<10^{-15}$ and 10^{-4}, respectively, Fig. 5a, Supplementary Table 2), whereas bivalent chromatin domains were slightly enriched ($P<0.05$). Genes occupied by both ING5 and bivalent chromatin, or by bivalent chromatin and DNA methylation, showed marginal or no evidence for co-regulation during differentiation ($P<0.01$ and non-significant, respectively, Fig. 5b). In contrast, strong enrichment was found for genes harbouring both ING5 and CpG methylation ($P<10^{-11}$, Fig. 5b), suggesting that they target a partially common gene-set to keep keratinocytes undifferentiated. ChIP-qPCR experiments showed that ING5, SmarcA5 and BPTF, co-occupy chromatin at tested loci lacking CpG methylation, suggesting

functional co-operativity on at least a subset of ING5 target genes (Supplementary Fig. 6a and data not shown).

Overrepresentation analysis indicated that genes harbouring ING5 and meCpG were enriched in the 60-80th percentile of expressed genes. In contrast, genes targeted by ING5, but not containing DNA methylation, were enriched in the 80-100th percentile (Fig. 5c). This suggests that the combination of ING5+meCpG results in a slightly lower gene expression level. In addition, only ING5+meCpG marked genes that are downregulated upon induction of differentiation are overrepresented in relatively highly expressed genes, whereas both up and downregulated ING5-only containing genes were overrepresented in the same group (Supplementary Fig. 6b). This could suggest a role for DNA methylation in downregulating these genes upon differentiation.

P63 is involved in maintenance of keratinocyte self-renewal[8,9]. A recent study revealed p63, predominantly ΔNp63, binding events in proliferating primary human keratinocytes on a genome-wide scale[36]. P63 localised to genes containing ING5, but not ING5+meCpG (Supplementary Fig. 7a). This suggests that p63 is involved in the gene set targeted by ING5, SmarcA5 and BPTF. Although we did not find direct physical interactions between p63, SmarcA5 and BPTF in soluble extracts of primary human keratinocytes (Fig. 2c), these interactions might occur on chromatin.

These experiments indicate that the epigenetic strategies within the functional network regulate three distinct sets of genes: one targeted by ING5, SmarcA5 and BPTF, one by the combination of ING5 and DNA methylation, and a third by bivalent chromatin domains (Fig. 6c).

We hypothesised that the genetic interactions within the self-renewal subnetwork might arise from higher-order redundancy, where the gene sets targeted by individual components are distinct, yet form functionally similar modules. In agreement with this, we found a high degree of overlap between the GO classes targeted by the two main gene sets, ING5+SmarcA5+BPTF, and ING5+meCpG (Fig. 6a, $P<10^{-17}$). Epidermal stem cells exit the cell cycle and detach from the basal layer when they differentiate[5]. The overlapping GO terms contained genes that control proliferation and cell adhesion (Fig. 6b). In addition, genes targeted by bivalent chromatin were enriched in components of the TGFβ-pathway (Fig. 6c, $P<0.05$), which is known to attenuate proliferation during differentiation[37].

ITGA6 and ITGB1 are targets of the subnetwork that interact genetically

If the observed genetic interactions indeed arise from higher-order redundancy, at least some of the genes downstream of the distinct arms of the self-renewal subnetwork should also display genetic interactions. Our genomic analysis revealed that several genes involved in integrin mediated adhesion (eg. ITGA6 and ITGB1), are targeted by distinct arms of the subnetwork (Fig. 6a-c). Integrin α6 (ITGA6) and β1 (ITGB1) are involved in attachment of epidermal stem cells to

their niche via biochemically and spatially distinct mechanisms. Integrin α6 heterodimerises with integrin β4 in hemidesmosomes, whereas integrin β1 heterodimers containing α2, α3, and α5 are in focal adhesions[6]. ITGA6 and ITGB1 are expressed in epidermal stem cells and are down-regulated upon differentiation (Fig. 3b). However, little is known about which factors control their expression.

Using ChIP, methylated DNA IP (meDIP) and bisulfite sequencing, we found that the integrin α6 (ITGA6) promoter is marked by ING5 and 5meC DNA methylation (Fig. 7a-d). ChIP followed by bisulfite conversion and pyrosequencing revealed that ING5 and DNA methylation co-exist on the same DNA molecule, excluding the possibility that the observed co-occupancy is due to signals derived from distinct cell populations (Fig. 7e). ING5 binding is lost upon differentiation while DNA methylation persists, indicating that these factors may cooperate functionally without being interdependent (Fig. 7f,g). Moreover, the same CpG residues are methylated before and after induction of differentiation, excluding a local shift in methylation from the non-CpG island to the CpG island position (data not shown).

Recent reports show that 5-hydroxymethylation regulates gene expression in embryonic stem cells[38,39]. We did not find evidence that the ITGA6 locus contains this mark (Fig. 7c and Supplementary Fig. 8a,b). In fact, global levels of 5-

hydroxymethylation are very low in both proliferating and differentiated primary keratinocytes (Supplementary Fig. 8b).

In contrast to ITGA6, the ITGB1 promoter features ING5, BPTF and SmarcA5 (Fig. 7a-c). Therefore ITGA6 and ITGB1 are exemplars of the two major gene-sets targeted by our self-renewal subnetwork (Fig. 6c). In addition, we found that p63 targets the ITGB1, but not ITGA6, locus, further indicating that the two integrin genes are regulated by distinct mechanisms in keratinocytes (Supplementary Fig. 7b).

To interrogate a possible genetic interaction between ITGA6 and ITGB1 in a functional assay, we measured the effect of silencing of each gene on the capacity of cells to remain adherent to the culture substrate. Combined ITGA6 and ITGB1 knock-down decreased the proportion of adherent cells to a greater extent than predicted from silencing them individually (Fig. 7h, $P<10^{-8}$ and Supplementary Fig. 8c). This shows that ITGA6 and ITGB1 indeed interact genetically and that the functional connections within our self-renewal subnetwork can, at least in part, be attributed to genetic interactions between its downstream targets.

Subnetwork components control differentiation in reconstituted human epidermis

Our data suggest that cells in which individual components of the self-renewal subnetwork have been silenced will be selectively expelled from the epidermal stem cell compartment and undergo terminal differentiation in the suprabasal layers. We used an organotypic skin reconstitution assay[16] to test this. GFP-IRES-shRNA lentiviral vectors silencing ING5, SmarcA5, BPTF, EZH2, UHRF1 or a control shRNA vector were introduced into primary human keratinocytes. 24 hours later, the cells were seeded onto de-epidermised dermis and cultured at the air-liquid interface (Fig. 8a). After 3 weeks the cells had reconstituted a morphologically normal interfollicular epidermis with distinct basal, spinous, granular and cornified layers (Fig. 8b-c).

After transduction, approximately 25-30% of keratinocytes expressed the GFP-IRES-shRNA construct (data not shown). If cells lacking any of the tested components have an increased tendency to differentiate compared to uninfected cells in the same population, there will be a diminution of the frequency and position of GFP positive clones in reconstituted epidermis. Indeed, all five chromatin modifiers reduced clone formation (Fig. 8d-f). In addition, three-dimensional wholemount imaging showed that by 3 weeks many of the clones that did form were no longer anchored in the basal layer, indicating that they were in the process of undergoing terminal differentiation and being shed from the surface of the reconstituted epidermis (Fig. 8c-g). This is consistent with the observation that components of the self-renewal network regulate integrin genes (Fig. 6b and 7). We conclude that the novel factors identified in our screen are

indeed important for controlling epidermal differentiation in a physiological context and regulate genes involved in maintenance of interactions between stem cells and their niche.

DISCUSSION

Our siRNA-based screens identified both known and novel epigenetic regulators of epidermal stem cell fate. The computational approach we developed revealed a network of genetic interactions involving ING5, SmarcA5, BPTF, EZH2 and UHRF1. We showed that cells depleted of these genes are unable to contribute to maintenance of reconstituted human epidermis because they are expelled from the basal layer and undergo terminal differentiation.

The epigenetic factors we identified target at least two distinct sets of genes involved in epidermal self-renewal and differentiation. One set is also targeted by p63, suggesting that ING5 and p63 may work together in epidermal keratinocytes. However, ING5 depletion does not completely phenocopy knock-down of p63[40]. This is not unexpected as p63 probably uses a range of chromatin factors to exert its effects and ING5 may likewise interact with other sequence-specific transcription factors.

Although the two gene sets are different, many of the proteins they encode are involved in identical cellular functions, such as mitosis and integrin mediated adhesion. We confirmed that ITGA6 and ITGB1 are regulated by complementary

components of the ING5, SmarcA5, BPTF, EZH2 and UHRF1 subnetwork and interact genetically. Thus genetic interactions among the epigenetic modifiers are underpinned by genetic interactions between their downstream targets. It has long been recognised that the α6 and β1 integrins co-operate to anchor epidermal stem cells to the underlying basement membrane, but until now the means by which the ITGA6 and ITGB1 genes are co-ordinately downregulated during differentiation were obscure.

A key function of biological networks is to confer resistance to genetic and environmental perturbations[1]. This may be achieved by simple redundancy between biochemically equivalent network components. For instance, redundancy between EZH1 and EZH2[12] and HDAC1 and HDAC2[41] is observed in mouse epidermis. We now show that higher-order redundancy can be achieved when a network controls discrete, yet functionally overlapping, downstream functions. Co-ordinate epigenetic regulation of distinct gene sets is an important and previously unrecognised fail-safe mechanism that protects epidermal stem cells from premature differentiation.

Materials & Methods.

Cell culture and differentiation.

Primary normal human keratinocytes (oral lka, foreskin kc and km strains), obtained with appropriate ethical consent, were cultured on feeders as described[42]. Prior to induction of differentiation, cells were grown, feeder-free, in Keratinocyte Serum Free Medium (KSFM containing 30 μg/ml Bovine Pituitary Extract and 0.2 ng/ml EGF; Gibco) for 2-3 days. At ~70% confluency cells were incubated in KSFM containing 10 μM AG1478 (Calbiochem), 200 ng/ml recombinant human BMP2/7 (R&D systems), both, or 10% foetal bovine serum (PAA).

siRNA nucleofection.

siRNA nucleofections were performed with the Amaxa 96-well shuttle system (Lonza). Keratinocytes were grown in KSFM to ~70% confluency, harvested and resuspended in cell line buffer SF. $2x10^5$ cells were used for each 20 μl transfection (program FF-113) with 1-2 μM siRNA duplexes. This is equivalent to 5-10 nM siRNA in conventional liposome-based transfections. Transfected cells were incubated at ambient temperature for 5-10 minutes and subsequently resuspended in pre-warmed KSFM. Silencer Select siRNAs were used (Ambion/Applied Biosystems).

siRNA library.

We designed a custom library (Silencer Select product, purchased from Ambion) targeting a comprehensive set of human genes encoding known or putative chromatin-factors. We included factors containing any of the following domains or functions: PHD, BROMO, CHROMO, PWWP, tandem BRCT, TUDOR, BAH, MBT, SET (including DOT1L), JMJC, JMJN, PRMT, HAT, HDAC, SIRT, DNMT, MBD, and SNF2 ATP-dependent remodelers. After manual curation for redundant entries a final list of 332 chromatin-associated factors was obtained (see Supplementary Table 1).

siRNA screening and data processing.

We used passage 2 Ika keratinocytes for the siRNA screens. Our custom library of 332 siRNA pools (3 duplexes/pool) was plated in four 96-well plates. Following transfection, keratinocytes were manually dispensed into twenty 96-well plates (8,000 cells/well) containing pre-warmed KSFM. This allowed analysis of quadruplicate plates for each of the five treatment groups (vehicle, AG1478, BMP2/7, AG1478+BMP2/7 and 10% serum). Medium was refreshed the next day. 72 hours after transfection cells were differentiated for 48 hours. Cells were fixed in 4% paraformaldehyde (10 min, RT), washed and permeabilised in PBS+0.2% Triton X-100 (10 min, RT). Following blocking (PBS+10%serum, 30 min, RT), cells were stained using Transglutaminase I specific antibodies (1:2000 mouse monoclonal BC.1 in blocking buffer) for at least 1 hour (RT). After three washes, cells were stained with IR800 anti-mouse secondary antibodies (1:2000, LiCor) and a DNA stain, DRAQ5 (1:2000, Biostatus Ltd), in blocking buffer for 1

hour at RT. Cells were washed 3 times and a final volume of 100 µl TBS was added. Plates were scanned and analysed using the LiCor Odessey system and software with consistent settings throughout the whole screen.

As a transfection efficiency dependent background control we included two independent siRNAs targeting TG1. After scanning and quantification, this background was subtracted and TG1 levels for each individual well were normalised to DRAQ5 signal to give a measure of differentiation/cell for each population of siRNA transfected cells. High data quality was ensured by confirming high pearson-correlation coefficients (Pearson-correlation>0.95) of each replicate versus the mean of the quadruplicates.

A Z-score was subsequently calculated for each population [43].

$$Z = \frac{X - A}{\delta}$$

Where X is the background corrected normalised intensity of a specific well, A and δ are the mean and standard deviation of background corrected normalised intensities of all test siRNAs on the plate, respectively. This calculation effectively transforms the data (without affecting its distribution) to an average of zero and a standard deviation of one. This standardised format allowed us to compile the results of all conditions and experiments into one dataset of ~6,600 quantitative measurements. The three best replicates were used for hierarchical clustering and network analysis.

Hierarchical clustering, statistical analysis and network visualisation.

Unsupervised hierarchical clustering and matrix visualisation was performed using R (http://www.r-project.org/), GenePattern[44] or the online Matrix2PNG tool[45]. Over/under-representation (hypergeometric test) and gene set enrichment analysis (GSEA) was performed using web-based tool GeneTrail[46]. Significance of the genetic interactions in Fig. 3e was calculated using an unpaired two-tailed t-test as described[3,30] after checking normality of the data (n=6). The non-parametric Man-Whitney test was used to calculate significant differences in the skin reconstitution assays as the data were not normally distributed. Networks were visualised using Cytoscape[47].

Human nuclear protein interaction network.

We used the highly curated Proteins Interacting in the Nucleus database (PINdb)[25] to extract human protein complexes. This contained information on protein interaction partners of ~100 of the factors represented in our siRNA library. We considered proteins residing in the same multi-subunit complex to interact with each other, rather than focussing on the sparse and incomplete information on direct physical interactions.

Bayesian mixture model to infer functional interactions.

We developed a posterior association network (*PAN*) to predict functional interactions between genes. A *PAN* encodes *a posteriori* beliefs of functional association types on edges and perturbation phenotypes on nodes. *PAN*

quantifies the statistical significance of functional interactions by Bayesian mixture modelling of gene association densities.

This package is available in the R language at:

http://www.bioconductor.org/packages/devel/bioc/html/PAN.html.

Finite mixture models have been used to identify co-expressed genes from gene expression data[48]. An efficient methodology was proposed by Ji et al., which models densities of correlation coefficients of gene expression levels by a mixture of a finite number of beta distributions[49]. We propose a beta-mixture distribution to model associations of perturbation screens, built on the assumption that the distribution of gene association scores computed from rich phenotyping screens is a mixture of three components representing positive (+), negative (-) and lack (x) of association (x). We employ a stratification strategy to take into consideration potential prior knowledge for the functional network such as protein-protein interactions.

To infer the beta-mixture model from data, we performed MAP (maximum *a posterior*) based on the EM algorithm[50]. The algorithm alternates between computing the expectation of the log-posterior probability based on the current estimates for the latent variables and maximizing the expected log-posterior. Having estimated the parameters in the beta-mixture model, we computed posterior probabilities for each gene pair belonging to the positive, negative or lack of association component. To perform a model selection for each edge, a posterior odd in favor of signal (association) to noise (lack of association) was

computed. A cutoff score of 10, interpreted as 'strong' evidence in *Bayesian inference*[51], was set to filter out non-significant edges.

Using Bayesian mixture modelling, we predicted a functional network of 158 chromatin factors with 837 interactions of strong statistical significance. To further investigate the modularity among the chromatin factors, we searched for coherent functional modules in the network by performing hierarchical clustering on second-order cosine similarities—cosine similarities of functional profiles of cosine similarities between genes. The second-order similarities are a highly desirable measure to group genes with similar interaction patterns by comparing functional profiles of genes instead of their own functions[29]. To assess the uncertainty of the clustering analysis, we computed a *p*-value for each cluster using multiscale bootstrap resampling (10,000 times, more details in the R package *pvclust*[52]).

Chromatin Immunoprecipitation (ChIP) and ChIP-sequencing.

Antibodies used were: ING5 (AbNova), SmarcA5 (Active Motif), BPTF (Abcam), Flag (M2, Sigma). ChIP was performed essentially as described[18]. Briefly, formaldehyde crosslinked material corresponding to ~10^7 keratinocytes (grown on feeders) was incubated with 10 μg antibody overnight. After centrifugation to pellet precipitated material, antibodies were captured with 100 μl protG-coated magnetic beads (Dynal) for 2-4 hours. All steps were performed at 4^0C. After five consecutive washes, the material was reverse-crosslinked overnight at 65^0C,

treated with proteinase K and RNAse, purified by organic extraction and collected by ethanol precipitation.

ChIP experiments were analysed using SYBR green based quantitative PCR. IP efficiency was calculated with regard to a dilution series of input material. Differences between chromatin preparations were corrected by normalisation to an intergenic region on the left arm of chromosome 2. This region was not enriched in any ChIP we performed, except when using a total H3 antibody. To correct for aspecific signal we subsequently corrected for ChIPs performed with a negative antibody. After verification of ChIP efficiency, material was prepared for sequencing as described[53]. The raw sequence data were put through our in-house pipeline. This included base-calling, alignment to the genome using BWA and filtering out potential PCR duplications. For the plots depicted in Fig. 3a, we used the cumulative sequence tag count in 1 kilobase bins across the genome. Using binned data we defined bivalent chromatin domains as a cumulative score of > 50 in -/+ 2kb of the transcription start site (TSS) for H3K4me3 and > 150 across -/+ 10 kb of TSS for H3K27me3. This measure is relatively lenient to allow analysis of a larger number of genomic regions. Data for single histone modifications, RNA pol II and CTCF in normal human keratinocytes was available from the ENCODE consortium[31] (http://genome.ucsc.edu/ENCODE/downloads.html). Locations of CpG islands in the human genome were downloaded from (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database). Methylated

promoter regions in undifferentiated primary human keratinocytes were extracted from Sen *et al* [16]. Methylated CpG islands were determined by intersecting these resources.

Western blots:

Keratinocytes were lysed in 1X sample buffer and incubating at 95^0C for 10 minutes. Proteins were separated on 4-12% polyacrylamide gradient gels, transferred to PVDF membrane and detected with specific antibodies (SMARCA5, Bethyl; EZH2, Cell signaling). Uncropped blots are available in Supplementary Fig. 8.

RNA extraction, RT-qPCR and expression profiling.

Total RNA was isolated using the RNeasy kit (Qiagen), including a DNaseI digestion step. cDNA was generated using the superscript III supermix for qPCR (Invitrogen). Quantitative PCR analysis of cDNA was performed using Taqman probes and Taqman fast chemistry (Applied Biosystems). ChIPs were analysed using powerSYBR green (Applied Biosystems) and self designed primers. Taqman probes and ChIP primer sequences are available upon request. Genome-wide expression profiling (on kc keratinocytes) was performed using the Illumina BeadArray platform and standard protocols. Data were processed using Genespring GX10 software.

Skin reconstitution assays:

Keratinocytes (km passage 2) cultured on feeders to ~50-70% confluence were infected with pGIPZ based Lentiviral vectors (Open Biosystems) expressing miR30 embedded shRNAs. Green fluorescent protein was expressed from the same promoter via an IRES. 24 hours after infection, cells were harvested and seeded on irradiated de-epidermised human dermis[16] in 6-well trans-well plates with feeders and cultured at the air-liquid interface for 3 weeks. Cultures were imaged using the ICys system (CompuCyte) for 2 dimensional imaging and a 2 photon Leica confocal microscope for 3 dimensional imaging. Quantification of the unprocessed data was performed using ICys and Volocity software. Three-dimensional rendering of organotypic cultures were performed using Volocity software.

References:

1 Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101-113, doi:10.1038/nrg1272 nrg1272 [pii] (2004).

2 Boutros, M. & Ahringer, J. The art and design of genetic screens: RNA interference. *Nat Rev Genet* **9**, 554-566, doi:nrg2364 [pii] 10.1038/nrg2364 (2008).

3 Mani, R., St Onge, R. P., Hartman, J. L. t., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc Natl Acad Sci U S A* **105**, 3461-3466, doi:0712255105 [pii] 10.1073/pnas.0712255105 (2008).

4 Perez-Perez, J. M., Candela, H. & Micol, J. L. Understanding synergy in genetic interactions. *Trends Genet* **25**, 368-376, doi:S0168-9525(09)00132-2 [pii] 10.1016/j.tig.2009.06.004 (2009).

5 Blanpain, C. & Fuchs, E. Epidermal stem cells of the skin. *Annu Rev Cell Dev Biol* **22**, 339-373, doi:10.1146/annurev.cellbio.22.010305.104357 (2006).

6 Watt, F. M. Role of integrins in regulating epidermal adhesion, growth and differentiation. *EMBO J* **21**, 3919-3926, doi:10.1093/emboj/cdf399 (2002).

7 Green, H. The birth of therapy with cultured cells. *Bioessays* **30**, 897-903, doi:10.1002/bies.20797 (2008).

8 Pellegrini, G. *et al.* p63 identifies keratinocyte stem cells. *Proc Natl Acad Sci U S A* **98**, 3156-3161, doi:10.1073/pnas.061032098 98/6/3156 [pii] (2001).

9 Koster, M. I. & Roop, D. R. The role of p63 in development and differentiation of the epidermis. *J Dermatol Sci* **34**, 3-9, doi:S0923181103002238 [pii] (2004).

10 Eckert, R. L., Crish, J. F., Banks, E. B. & Welter, J. F. The epidermis: genes on - genes off. *J Invest Dermatol* **109**, 501-509, doi:S0022202X97890458 [pii] (1997).

11 Driskell, I. *et al.* The histone methyltransferase Setd8 acts in concert with c-Myc and is required to maintain skin. *EMBO J* **31**, 616-629, doi:emboj2011421 [pii]
10.1038/emboj.2011.421 (2012).
12 Ezhkova, E. *et al.* EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev* **25**, 485-498, doi:gad.2019811 [pii]
10.1101/gad.2019811 (2011).
13 Ezhkova, E. *et al.* Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* **136**, 1122-1135, doi:S0092-8674(09)00006-3 [pii]
10.1016/j.cell.2008.12.043 (2009).
14 Luis, N. M. *et al.* Regulation of human epidermal stem cell proliferation and senescence requires polycomb- dependent and -independent functions of Cbx4. *Cell Stem Cell* **9**, 233-246, doi:S1934-5909(11)00343-2 [pii]
10.1016/j.stem.2011.07.013 (2011).
15 Mejetta, S. *et al.* Jarid2 regulates mouse epidermal stem cell activation and differentiation. *EMBO J* **30**, 3635-3646, doi:emboj2011265 [pii]
10.1038/emboj.2011.265 (2011).
16 Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. & Khavari, P. A. DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463**, 563-567, doi:nature08683 [pii]
10.1038/nature08683 (2010).
17 Sen, G. L., Webster, D. E., Barragan, D. I., Chang, H. Y. & Khavari, P. A. Control of differentiation in a self-renewing mammalian tissue by the histone demethylase JMJD3. *Genes Dev* **22**, 1865-1870, doi:22/14/1865 [pii]
10.1101/gad.1673508 (2008).
18 Connelly, J. T. *et al.* Actin and serum response factor transduce physical cues from the microenvironment to regulate epidermal stem cell fate decisions. *Nat Cell Biol* **12**, 711-718, doi:ncb2074 [pii]
10.1038/ncb2074 (2010).
19 Gosselet, F. P., Magnaldo, T., Culerrier, R. M., Sarasin, A. & Ehrhart, J. C. BMP2 and BMP6 control p57(Kip2) expression and cell growth arrest/terminal differentiation in normal primary human epidermal keratinocytes. *Cell Signal* **19**, 731-739, doi:S0898-6568(06)00265-8 [pii]
10.1016/j.cellsig.2006.09.006 (2007).
20 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, doi:ncb1750 [pii]
10.1038/ncb1750 (2008).
21 Kuramoto, N. *et al.* Development of ichthyosiform skin compensates for defective permeability barrier function in mice lacking transglutaminase 1. *J Clin Invest* **109**, 243-250, doi:10.1172/JCI13563 (2002).
22 Matsuki, M. *et al.* Defective stratum corneum and early neonatal death in mice lacking the gene for transglutaminase 1 (keratinocyte transglutaminase). *Proc Natl Acad Sci U S A* **95**, 1044-1049 (1998).
23 Vermeulen, M. *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-980, doi:S0092-8674(10)00951-7 [pii]
10.1016/j.cell.2010.08.020 (2010).
24 Doyon, Y. *et al.* ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell* **21**, 51-64, doi:S1097-2765(05)01849-6 [pii]
10.1016/j.molcel.2005.12.007 (2006).
25 Luc, P. V. & Tempst, P. PINdb: a database of nuclear protein complexes from human and yeast. *Bioinformatics* **20**, 1413-1415, doi:10.1093/bioinformatics/bth114
bth114 [pii] (2004).

26 Rahman, S. *et al.* The Brd4 Extraterminal Domain Confers Transcription Activation Independent of pTEFb by Recruiting Multiple Proteins, Including NSD3. *Mol Cell Biol* **31**, 2641-2652, doi:MCB.01341-10 [pii]
10.1128/MCB.01341-10 (2011).
27 Indra, A. K. *et al.* Temporally controlled targeted somatic mutagenesis in embryonic surface ectoderm and fetal epidermal keratinocytes unveils two distinct developmental functions of BRG1 in limb morphogenesis and skin barrier formation. *Development* **132**, 4533-4544, doi:132/20/4533 [pii]
10.1242/dev.02019 (2005).
28 Kashiwagi, M., Morgan, B. A. & Georgopoulos, K. The chromatin remodeler Mi-2beta is required for establishment of the basal epidermis and normal differentiation of its progeny. *Development* **134**, 1571-1582, doi:dev.001750 [pii]
10.1242/dev.001750 (2007).
29 Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425-431, doi:327/5964/425 [pii]
10.1126/science.1180823 (2010).
30 Horn, T. *et al.* Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat Methods* **8**, 341-346, doi:nmeth.1581 [pii]
10.1038/nmeth.1581 (2011).
31 Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90, doi:nature04815 [pii]
10.1038/nature04815 (2006).
32 LeRoy, G., Loyola, A., Lane, W. S. & Reinberg, D. Purification and characterization of a human factor that assembles and remodels chromatin. *J Biol Chem* **275**, 14787-14790, doi:10.1074/jbc.C000093200
C000093200 [pii] (2000).
33 Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:S0092-8674(06)00380-1 [pii]
10.1016/j.cell.2006.02.041 (2006).
34 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49, doi:nature09906 [pii]
10.1038/nature09906 (2011).
35 Champagne, K. S. *et al.* The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. *Proteins* **72**, 1371-1376, doi:10.1002/prot.22140 (2008).
36 Kouwenhoven, E. N. *et al.* Genome-wide profiling of p63 DNA-binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. *PLoS Genet* **6**, e1001065, doi:10.1371/journal.pgen.1001065 (2010).
37 Li, A. G., Koster, M. I. & Wang, X. J. Roles of TGFbeta signaling in epidermal/appendage development. *Cytokine Growth Factor Rev* **14**, 99-111, doi:S1359610103000054 [pii] (2003).
38 Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398-402, doi:nature10008 [pii]
10.1038/nature10008 (2011).
39 Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348, doi:nature10066 [pii]
10.1038/nature10066 (2011).
40 Truong, A. B., Kretz, M., Ridky, T. W., Kimmel, R. & Khavari, P. A. p63 regulates proliferation and differentiation of developmentally mature keratinocytes. *Genes Dev* **20**, 3185-3197, doi:20/22/3185 [pii]
10.1101/gad.1463206 (2006).
41 LeBoeuf, M. *et al.* Hdac1 and Hdac2 act redundantly to control p63 and p53 functions in epidermal progenitor cells. *Dev Cell* **19**, 807-818, doi:S1534-5807(10)00468-5 [pii]
10.1016/j.devcel.2010.10.015 (2010).
42 Rheinwald, J. G. & Green, H. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell* **6**, 331-343 (1975).

43 Birmingham, A. *et al.* Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* **6**, 569-575, doi:nmeth.1351 [pii]
10.1038/nmeth.1351 (2009).
44 Reich, M. *et al.* GenePattern 2.0. *Nat Genet* **38**, 500-501, doi:ng0506-500 [pii]
10.1038/ng0506-500 (2006).
45 Pavlidis, P. & Noble, W. S. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19**, 295-296 (2003).
46 Backes, C. *et al.* GeneTrail--advanced gene set enrichment analysis. *Nucleic Acids Res* **35**, W186-192, doi:gkm323 [pii]
10.1093/nar/gkm323 (2007).
47 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303
13/11/2498 [pii] (2003).
48 McLachlan, G. J. & Peel, D. *Finite mixture models*. Vol. 299 (Wiley-Interscience, 2000).
49 Ji, Y., Wu, C., Liu, P., Wang, J. & Coombes, K. R. Applications of beta-mixture models in bioinformatics. *Bioinformatics* **21**, 2118-2122, doi:bti318 [pii]
10.1093/bioinformatics/bti318 (2005).
50 Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 1-38 (1977).
51 Jeffreys, H. *Theory of probability*. 3rd edn, 432 (Oxford University Press, 1998).
52 Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-1542, doi:btl117 [pii]
10.1093/bioinformatics/btl117 (2006).
53 Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240-248, doi:S1046-2023(09)00047-4 [pii]
10.1016/j.ymeth.2009.03.001 (2009).

Acknowledgements

We thank A. Brenkman, J. Carroll, N. Rosenfeld, M. Narita, D. Odom and Watt lab members for critical discussion and comments. This work was supported by Cancer Research UK (CRUK), Hutchinson Wampoa, University of Cambridge (F.M.W.) and a Marie Curie Fellowship (PIEF-GA-2008-220642) to K.W.M. We are indebted to the core facilities of the CRUK Cambridge Research Institute for excellent technical assistance.

Author contributions

K.W.M conceived the study, designed, performed and analysed experiments and performed bioinformatic analysis. X.W. developed the Bayesian mixture model and network analysis. C.E. and G.S. contributed to the skin reconstitution assays. Y.I. performed bisulfite sequencing analysis. R.F.S. performed bioinformatic analysis. J.G. performed mRNA coexpression analysis. G.D. analysed data and contributed to experimental design. S.U-L. performed 5hmeDNA dot-blot analysis. P.P. performed mRNA coexpression analysis. A.M. oversaw the work of Y.I. and S. U-L. F.M. analysed data, performed bioinformatic analysis and contributed computational tools. F.M.W. designed experiments, analysed the data and oversaw the study. K.W.M and F.M.W. wrote the manuscript with input from all authors.

Author information

Genome-wide mRNA expression and ChIP-seq data are deposited in GEO, accession number GSE34558. The data are also available at http://www.markowetzlab.org/. The authors declare no competing interest. Correspondence and requests for materials should be addressed to K.W.M. (klaas.mulder@cancer.org.uk) or F.M.W. (fiona.watt@cancer.org.uk).

Figure legends:

Figure 1: **Chromatin-wide siRNA-based screen. a,** Schematic representation of the experimental strategy. An siRNA library targeting 332 known and putative chromatin-factors was transfected in to primary human keratinocytes. 72 hrs after transfections, cells were treated with the indicated agents to induce differentiation. Endogenous TG1 levels were quantified using an immunofluorescence based assay and normalised to cell numbers. **b,** RT-qPCR analysis of differentiation markers following 48 hours treatment with the indicated agents. mRNA levels were normalised to 18S. **c,** Expression profiling GO overrepresenation analysis of genes UP and DOWN regulated under the indicated conditions (also see Supplementary Fig. 1d).

Figure 2: siRNA screen reveals known and novel players controlling epidermal differentiation. a, Heatmap representation of the Z-scores of TG1 levels for all 332 knock-downs after two-dimensional clustering (cosine distance). Clusters of genes potentially involved in self-renewal (purple) and differentiation (orange) are highlighted. **b**, Protein complexes showing a statistically significant enrichment for self-renewal or differentiation effects (P<0.05, by GSEA). Edges denote physical interactions. Node color represents Z-score under vehicle treatment according to Fig. 1a. **c**, Nuclear extracts of keratinocytes immunoprecipitated using BPTF and SmarcA5 antisera and Western-blotted with SmarcA5 antibodies. Specificity of the SmarcA5-BPTF interaction is shown using

a negative control antibody (FLAG) and a non-interacting protein (p63). **d,** Keratinocytes were transfected with siRNAs targeting BPTF or SmarcA5. Non-targeting siRNAs were used as a control. 24 hours after transfection, cells were seeded at clonal density and cultured for two weeks. Number of colonies per plate was counted. Average ±SD, n=2.

Figure 3: **A Bayesian statistical model predicts functional/genetic interactions. a,** Schematic representation of the computational approach to define high confidence putative functional connections among all 332 chromatin-factors. Genes selected for detailed follow-up are highlighted. **b,** Expression of subnetwork components is downregulated upon differentiation. Keratinocytes were treated with vehicle or AG1478. After 48 hours, mRNA was isolated and samples subjected to RT-qPCR analysis. Asterisk indicates $P<0.05$ using a two tailed t-test. Average ±SD, n=6. **c,** Silencing individual subnetwork components induces expression of multiple differentiation markers. Keratinocytes were transfected with siRNAs targeting the indicated genes. After 72 hours, RT-qPCR analysis was performed. Data were normalised to GAPDH and represented as log2 fold over control siRNA. **d,** High incidence of genetic interactions among subnetwork components. A genetic interaction is assigned where the observed effect of the double knock-down is significantly different from the calculated expected value. Aggravating (yellow) and alleviating (blue) interactions show higher and lower TG1 levels than expected, respectively. Single knock-down data and non significant interactions are depicted in green and white,

respectively. Size of the data-points represents p-value (two tailed unpaired t-test, observed versus expected). Error bars indicate the SEM. See Supplementary Fig. 3 and 4 for controls.

Figure 4: **Genome-wide identification of subnetwork target genes. a,** Genomebrowser tracks of raw input and ING5 ChIP-seq tag counts. **b,** ING5 marks actively transcribed genes. Heatmaps represent ChIP signal in 1 bk bins around the transcription start site (TSS) of all protein coding RefSeq genes in the human genome, ranked according to ING5 signal. Right panel indicates DNA methylation status of each gene. **c,** ING5 binding signal correlates with expression level. Average ING5 ChIP-seq tag counts (-/+ 10 kb of the TSS) of the top, middle and bottom 25% of transcribed genes (as determined by log2 intensity on a microarray) were plotted. **d,** Correlations of genome-wide occupancies of the indicated factors and modifications determined using Pearson correlation. Cluster analysis revealed that global ING5 signals correlate well with active histone marks and RNA polymerase II signals. Color scale ranges from 0 to 0.7.

Figure 5: **Subnetwork components target distinct gene sets. a,** Genes differentially expressed in differentiated keratinocytes are enriched in ING5, DNA methylation and bivalent domains. Pie-charts indicate proportions of ING5, meCpG and bivalent domain-containing up or down regulated genes. P-values were calculated using a hypergeometric test. **b,** Components of the self-renewal

network target at least three distinct gene-sets. Overlaps between ING5, methylated CpG islands or bivalent chromatin targets are represented as Venn diagrams (circle size proportional to number of genes). P-values were calculated using a hypergeometric test. **c,** Over/Under representation analysis of genes containing ING5, meCpG or both, in 20th-percentile groups of genes stratified to their expression level. Significance was tested using a hypergeometric distribution test.

Figure 6: **Targeted gene sets encode proteins with overlapping functions. a,** GO term over/under representation analysis (hypergeometric test with multiple testing correction, $p<0.05$) was performed on the up or down regulated genes targeted by ING5-SMARCA5-BPTF, or by ING5-meCpG. A second independent hypergeometric test was used to determine significance of overlap of the GO terms enriched in either gene-set. **b,** Representative gene classes (GO terms) over-represented in both major gene sets involved in epidermal differentiation and self-renewal. P-value derived from a hypergeometric test after multiple testing correction ($p<0.05$). **c,** Schematic representation of levels of redundancy conveyed by the self-renewal subnetwork.

Figure 7: **Genes regulated by distinct arms of the self-renewal subnetwork display genetic interactions themselves. a,** Genome browser tracks of ING5 ChIP-seq signals on the integrin alpha 6 (ITGA6) and beta 1 (ITGB1) loci, respectively. CGI, CpG island. **b,** ChIP-qPCR of ING5, BPTF and SmarcA5

binding to ITGA6 and ITGB1 loci. Enrichment is expressed as fold over negative antibody (Flag) after correction to a negative genomic region. **c,** MeDIP-qPCR analysis for 5meC and 5hmeC on the ITGA6 and ITGB1 loci indicated strong 5meC of the ITGA6, but not ITGB1 locus. **d,** Bisulfite conversion followed by cloning and Sanger sequencing showed localised non CGI methylation of ITGA6 promoter. **e,** ChIP followed by bisulfite conversion and pyrosequencing showed that ING5 and DNA methylation occur on the same DNA molecule in the cell. **f,** ChIP of ING5 in proliferating and AG1478-differentiated keratinocytes showed loss of ING5 occupancy upon differentiation. **g,** MeDIP in proliferating and AG1478-differentiated keratinocytes shows that DNA methylation persists on the ITGA6 promoter following differentiation. **h,** ITGA6 and ITGB1 genetically interact. SiRNA transfected keratinocytes were cultured for six days, and then the number of attached cells was quantified. The expected level of cell adherence was calculated based on the single knock-downs. P-value was calculated with an unpaired two-tailed t-test. n=3 independent transfections.

Figure 8: **Epidermal reconstitution assays verify physiological relevance of self-renewal subnetwork components. a,** Outline of experimental set-up. Cultures were analysed by two and three-dimensional wholemount imaging. **b**, Hematoxilin and eosin stained section of reconstituted human epidermis after 3 weeks in culture. **c**, GFP-positive control clone (brown, with haematoxylin counterstain) extending from basal layer (dashed line) to differentiated suprabasal layers. **d**,**e**, Three dimensional rendering of whole-mount images of

control and ING5 shRNA expressing lenti-virus infected cultures. **f,** Quantification of GFP+ clone number in organ cultures (n=3-4 independent infections and cultures). **g**, Quantification of individual GFP+ clone position relative to basement membrane (BM) (n=3-4 independent infections and cultures). Quantifications were performed on unprocessed images. Asterisk indicates $P<0.05$ using a Mann-Whitney (non-parametric) test. Scale bars indicate 150 microns.

Supplementary figures:

Supplementary Fig. 1: **siRNA screen quality controls. a,** Immunofluorescence based read-out of primary screen. Specificity of the TG1 antibody is shown using two independent siRNAs against TG1. **b,** Cell number corrected TG1 protein expression levels for the indicated treatment groups (average ±SD, n=6). **c,** Kinetics of induction of differentiation markers. **d,** Venn diagram of differentially expressed genes in primary human keratinocytes differentiated with the indicated agents. **e,** Colony formation assay of cells treated with vehicle, AG1478, BMP2/7 or AG1478+BMP2/7. Treatment was started 24 hours after seeding and plates were harvested after 14 days. **f,g,** Raw (TG1/DRAQ5) data and corresponding Z-scores indicate high reproducibility within experimental replicates (upper panel; Pearson-correlation 0.94) and between biological replicates (lower panel; Pearson-correlation 0.91), 25% of the library. Error bars: SEM of experimental replicates. **h,** Comparing results from two different siRNA library plates indicates siRNA position does not influence outcome.

Supplementary Fig. 2: **siRNA pool deconvolution, false positive and negative rate estimation. a,** Primary human keratinocytes were transfected with 3 independent siRNAs against SmarcA5 or EZH2. Protein extracts were prepared 72 hours after transfection and subjected to western blot analysis. Non-targeting and GAPDH targeting siRNAs were included as controls. Actin was used as a loading control. **b-c,** siRNA screen hits resulting in spontaneous differentiation or that inhibited differentiation were confirmed with three independent siRNAs. TG1 protein levels under the indicated conditions and knock-down efficiencies were determined for each siRNA. Twelve out of thirteen genes reproduced the screen results with at least 2 independent siRNAs in a least one of the conditions. This indicates that the false positive rate may be as low as 10%. ND is not determined. **d,** Knock-down efficiencies of 18 siRNA pools across a wide range of screen Z-scores were determined. Only one gene displayed <50% knock-down in combination with a Z-score >-2 and <2 all 5 conditions. This suggests that the false negative rate may be <10%.

Supplementary Fig. 3: **Expression of self-renewal subnetwork components in double knock-down experiment. a,** RT-qPCR analysis of mRNA levels of network components in each of the double knock-down groups. Data are expressed relative to non-targeting siRNA transfections after normalisation to GAPDH as an endogenous control. Genes intended to be targeted by the siRNAs are indicated in red, the other network components in green. Error bars

indicate standard deviation. **b,** Combinatorial knock-down of components of the highly connected subnetwork involved in epidermal self-renewal. Single and double knock-down effects on TG1 mRNA levels were quantified by RT-qPCR (n=6 independent transfections). **c,** Highly similar results are obtained with two independent siRNA sets. The ratio of observed over expected TG1 mRNA levels was plotted for two independent experiments using two independent siRNAs for each gene of the subnetwork. The high correlation between the experiments indicates that the genetic interactions are supported by two siRNAs and are unlikely to be due to off-target effects.

Supplementary Fig. 4: **Double knock-down experiment of non highly connected nodes (control subnetwork). a,** RT-qPCR analysis of mRNA levels of genes in each of the double knock-down groups. Data are expressed relative to non-targeting siRNA transfections after normalisation to GAPDH as an endogenous control. Genes intended to be targeted by the siRNAs are indicated in red, the other network components in green. Error bars indicate standard deviation. **b,** RT-qPCR analysis of TG1 mRNA levels in all combination of siRNA double knock-downs. Data are displayed as log2 fold over control knock-downs after correction for GAPH as an endogenous control. Error bars depict SEM (n=4 independent transfections). **c,** Genetic interaction plot as in Fig. 3d. None of the gene pairs displayed strong genetic interactions. This is evident from the minor deviation from the expected values, Pearson correlation = 0.96.

Supplementary Fig. 5: **Reproducibility and specificity of ING5 ChIP(-seq). a,** Genome browser track showing ING5 ChIP-seq signals on a 11 Mb region of chromosome 2 from two independent biological replicates. For the first replicate cells were grown on feeders in the presence of serum in FAD. For the second replicate cells were grown under feeder and serum free conditions in KSFM. **b,** Heatmaps representing ChIP signal in 1 bk bins around the transcription start site (TSS) of all protein coding RefSeq genes in the human genome for the two biological replicates. **c,** ChIP-qPCR validation of a number of loci enriched for ING5 in our ChIP-seq datasets. Enrichment is expressed as fold over a negative antibody (Flag) after correction to a negative genomic region. **d,** ING5 ChIP-qPCR signal decreases proportionally to siRNA mediated ING5 knock-down. Control or ING5 targeting siRNA transfected cells were crosslinked and subjected to ChIP-qPCR using Flag or ING5 specific antibodies. Although the ING5 signal is not completely lost, it is proportional to the knock-down efficiency obtained. This argues that our ChIP signal is ING5 specific.

Supplementary Fig. 6: **ING5, SmarcA5 and BPTF target overlapping gene sets. a,** Twenty ING5 enriched loci and a negative control region were tested for BPTF and SMARCA5 enrichment by ChIP-PCR analysis. A high concordance (as measured by Pearson correlation) between ING5, BPTF and SmarcA5 enrichment indicates co-occupancy of the three factors on these loci. **b,** Overrepresentation analysis of genes containing ING5, meCpG or both, and differentially regulated during keratinocyte differentiation, in 20^{th}-percentile

groups of genes stratified to their expression level. Significance was tested using a hypergeometric distribution test.

Supplementary Fig. 7: **P63 targets ING5, but not ING5+meCpG gene set. a,** P63 target genes overlap with genes differentially expressed during keratinocyte differentiation and targeted by ING5, but not ING5+meCpG, in proliferating cells. This suggests that p63 targets the ING5+SmarcA5+BPTF gene set. **b,** ITGB1, but not ITGA6 promoter is occupied by p63 in proliferating keratinocytes, as determined by ChIP-qPCR.

Supplementary Fig. 8: **Primary human keratinocytes contain very low global levels of 5-hydroxymethylated DNA. a,** 5hmeCpG spike-in control. MeDIP with 5-hydroxymethylcytosine antibodies specifically recovered 5-hydroxymethylated but not methylated spike-in control DNA. **b,** Serial dilutions of purified total genomic DNA from keratinocytes treated with vehicle or AG1478 were subjected to spot blot analysis with antibodies specific for DNA methylation or 5-hydroxymethylated DNA. DNA from mouse ES cells, HCT116 cells, human placenta and synthetic modified DNA was included as positive and negative controls. **c,** Efficiency of double knock-down of ITGA6 and ITGB1 as analysed by RT-qPCR relative to control knock-down.

Supplementary Fig. 9: **Uncropped images from the western-blots in Fig 2c and Supplementary Fig. 2a.**

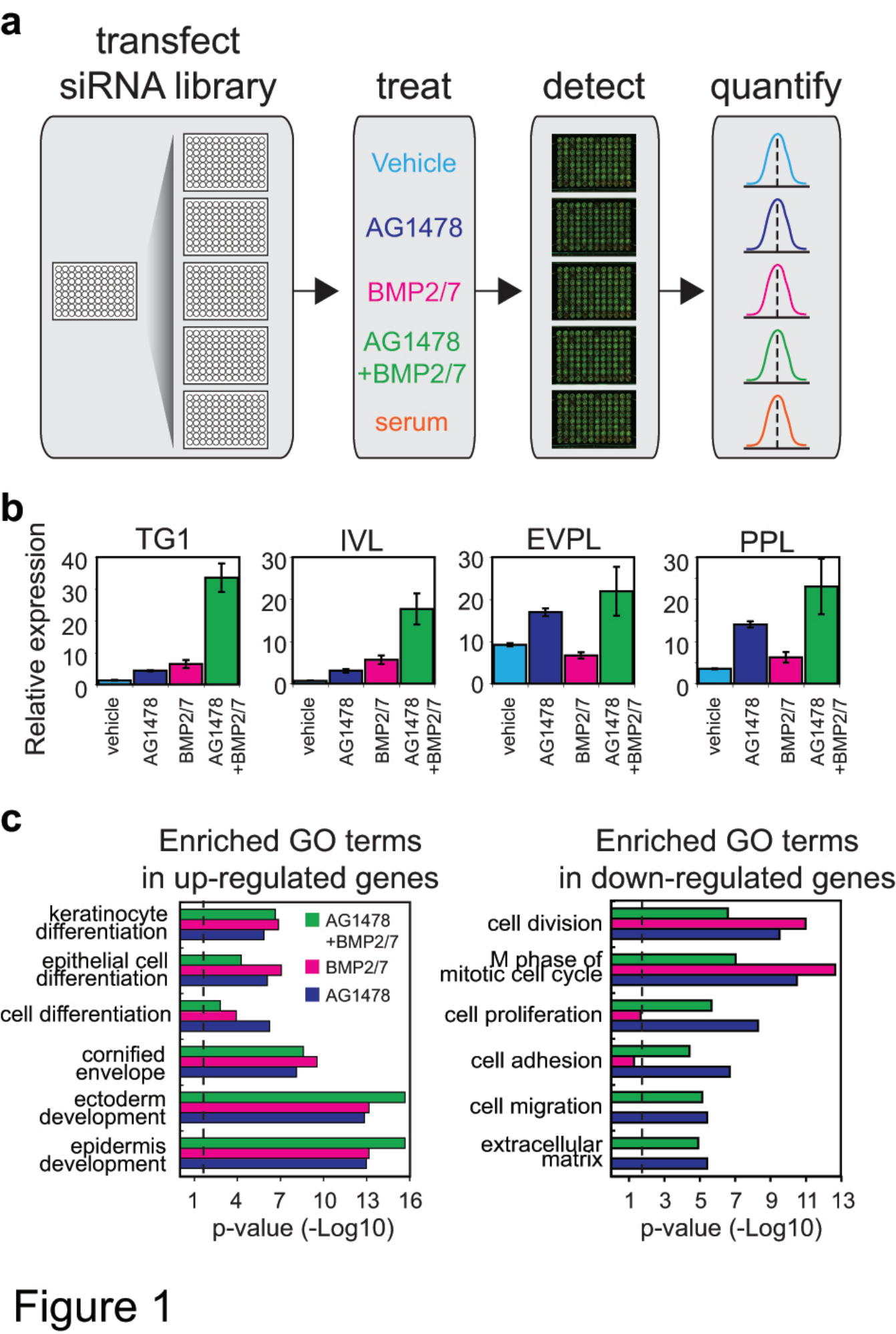
a
transfect siRNA library
treat
detect
quantify
Vehicle
AG1478
BMP2/7
AG1478 +BMP2/7
serum
b
Relative expression
TG1
IVL
EVPL
PPL
vehicle
AG1478
BMP2/7
AG1478 +BMP2/7
c
Enriched GO terms in up-regulated genes
keratinocyte differentiation
epithelial cell differentiation
cell differentiation
cornified envelope
ectoderm development
epidermis development
AG1478 +BMP2/7
BMP2/7
AG1478
p-value (-Log10)
Enriched GO terms in down-regulated genes
cell division
M phase of mitotic cell cycle
cell proliferation
cell adhesion
cell migration
extracellular matrix
p-value (-Log10)

Figure 1

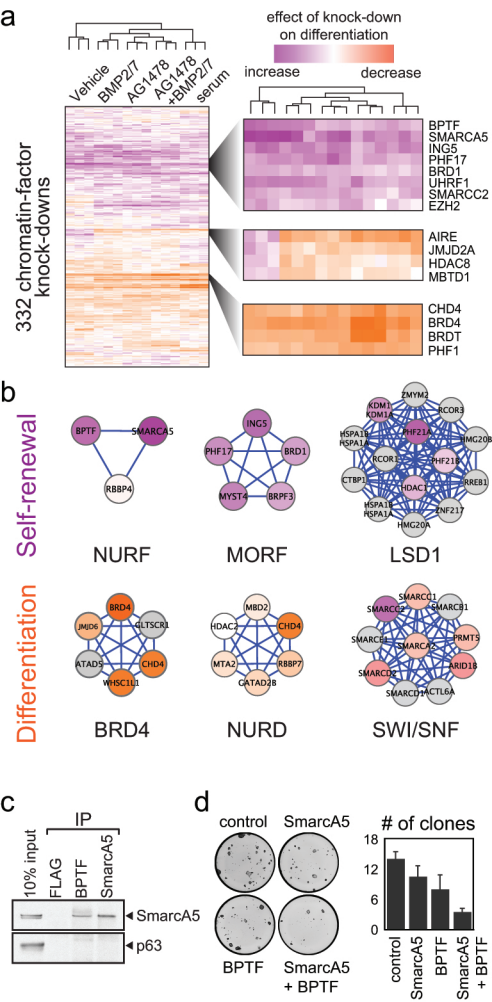
a
effect of knock-down on differentiation
increase
decrease
Vehicle
BMP2/7
AG1478
AG1478 +BMP2/7
serum
332 chromatin-factor knock-downs
BPTF
SMARCA5
ING5
PHF17
BRD1
UHRF1
SMARCC2
EZH2
AIRE
JMJD2A
HDAC8
MBTD1
CHD4
BRD4
BRDT
PHF1
b
Self-renewal
Differentiation
BPTF
SMARCA5
RBBP4
ING5
PHF17
BRD1
MYST4
BRPF3
ZMYM2
KDM1A
RCOR3
HSPA1A
PHF21A
IMG20B
RCOR1
PHF21B
CTBP1
HDAC1
RREB1
HSPA1B
ZNF217
HMG20A
NURF
MORF
LSD1
BRD4
JMJD6
CLTSCR1
ATAD5
CHD4
WHSC1L1
MBD2
HDAC2
CHD4
MTA2
RBBP7
CATAD2B
SMARCC1
SMARCC2
SMARCB1
SMARCE1
PRMT5
SMARCA2
ARID1B
SMARCD1
ACTL6A
BRD4
NURD
SWI/SNF
c
IP
10% input
FLAG
BPTF
SmarcA5
SmarcA5
p63
d
control
SmarcA5
BPTF
SmarcA5 + BPTF
# of clones
18
12
6
0
control
SmarcA5
BPTF
SmarcA5 + BPTF

Figure 2

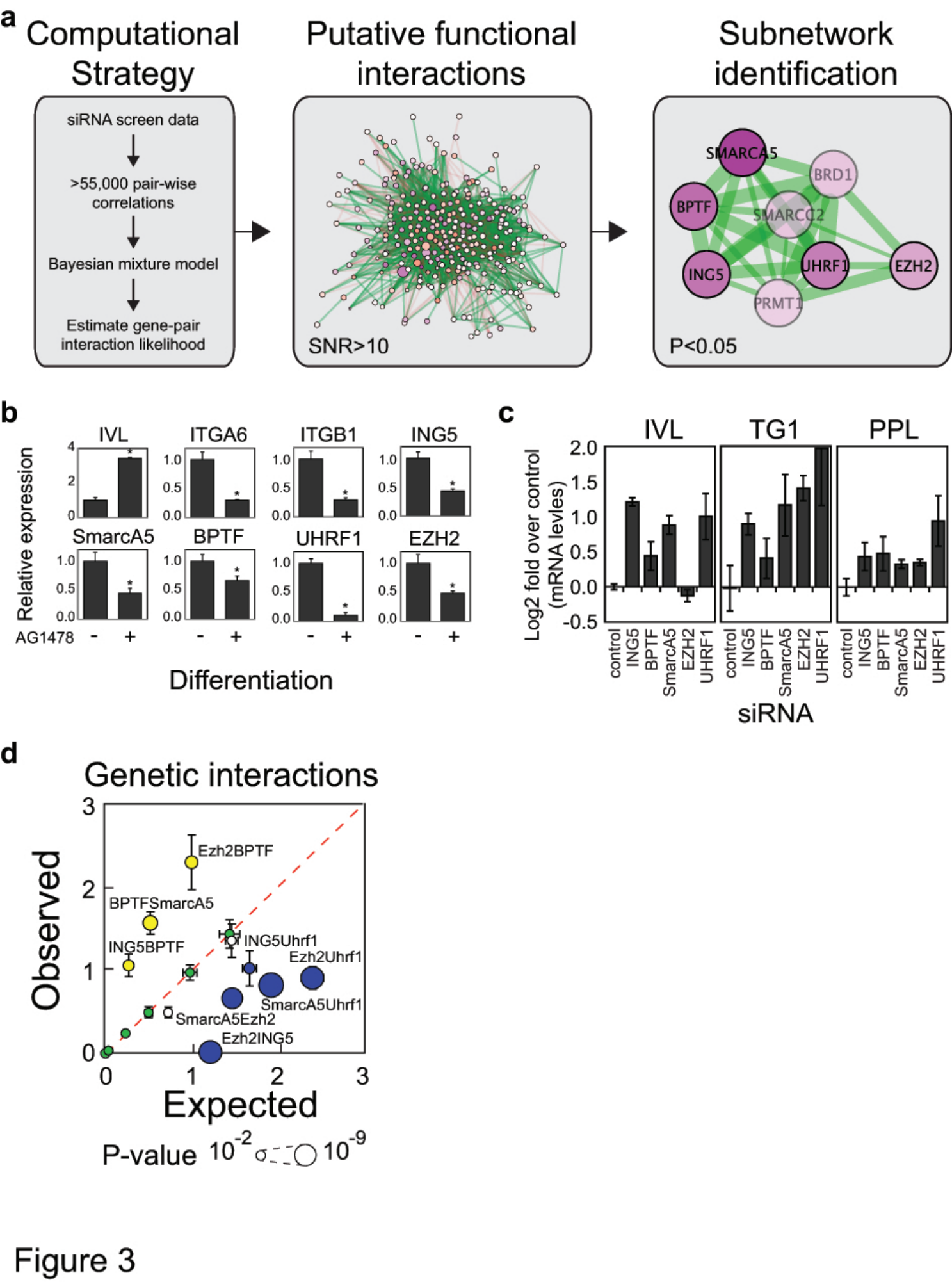
a
Computational Strategy
siRNA screen data
>55,000 pair-wise correlations
Bayesian mixture model
Estimate gene-pair interaction likelihood
Putative functional interactions
SNR>10
Subnetwork identification
SMARCA5
BRD1
BPTF
SMARCC2
ING5
UHRF1
EZH2
PRMT1
P<0.05
b
IVL
ITGA6
ITGB1
ING5
SmarcA5
BPTF
UHRF1
EZH2
Relative expression
AG1478 - + - + - + - +
Differentiation
c
IVL
TG1
PPL
Log2 fold over control (mRNA levles)
control
ING5
BPTF
SmarcA5
EZH2
UHRF1
siRNA
d
Genetic interactions
Observed
Expected
Ezh2BPTF
BPTFSmarcA5
ING5BPTF
ING5Uhrf1
Ezh2Uhrf1
SmarcA5Uhrf1
SmarcA5Ezh2
Ezh2ING5
P-value 10-2 10-9

Figure 3

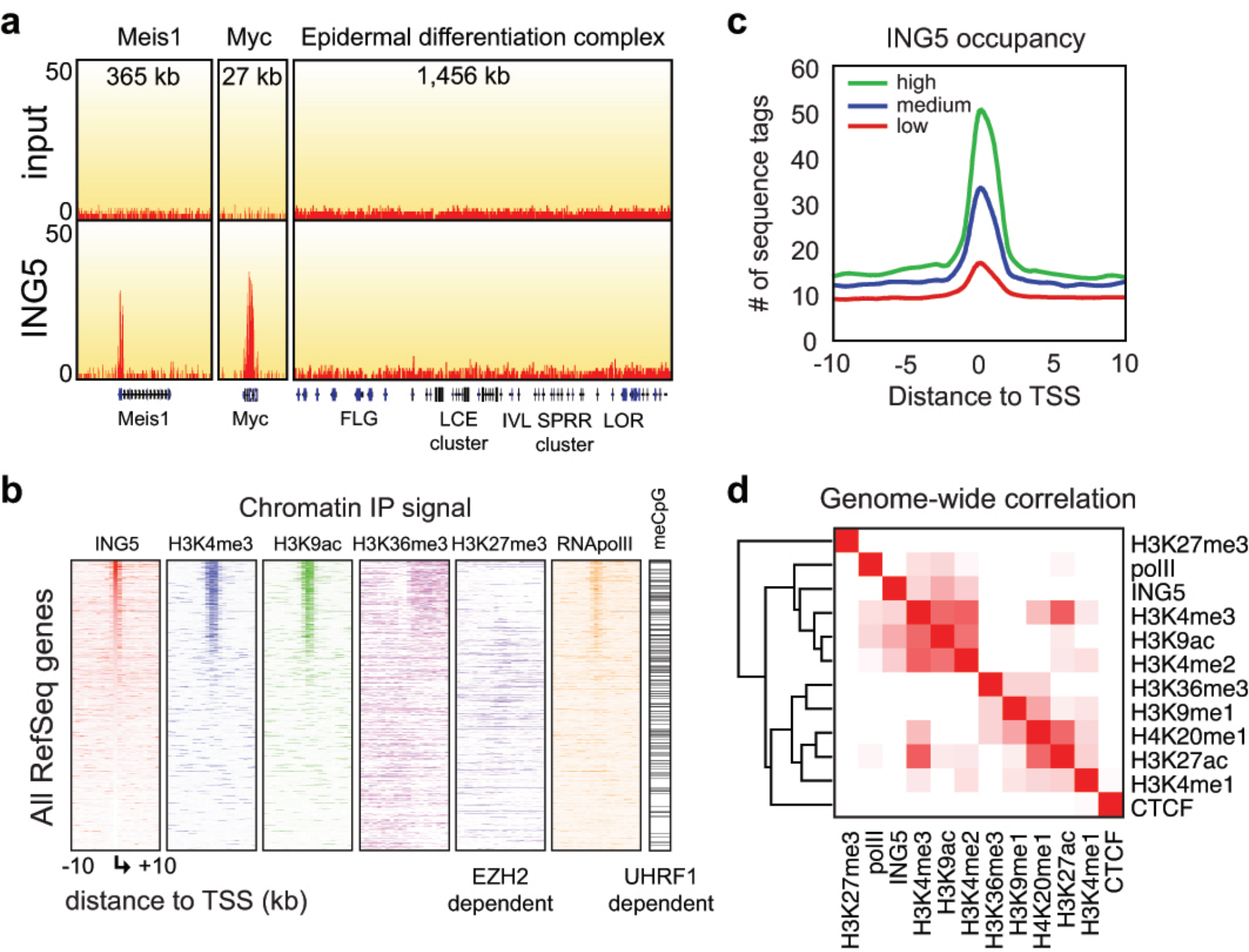
a
Meis1
Myc
Epidermal differentiation complex
365 kb
27 kb
1,456 kb
input
ING5
50
0
Meis1
Myc
FLG
LCE cluster
IVL
SPRR cluster
LOR
c
ING5 occupancy
high
medium
low
# of sequence tags
Distance to TSS
b
Chromatin IP signal
ING5
H3K4me3
H3K9ac
H3K36me3
H3K27me3
RNApolII
meCpG
All RefSeq genes
-10
+10
distance to TSS (kb)
EZH2 dependent
UHRF1 dependent
d
Genome-wide correlation
H3K27me3
polII
ING5
H3K4me3
H3K9ac
H3K4me2
H3K36me3
H3K9me1
H4K20me1
H3K27ac
H3K4me1
CTCF

Figure 4

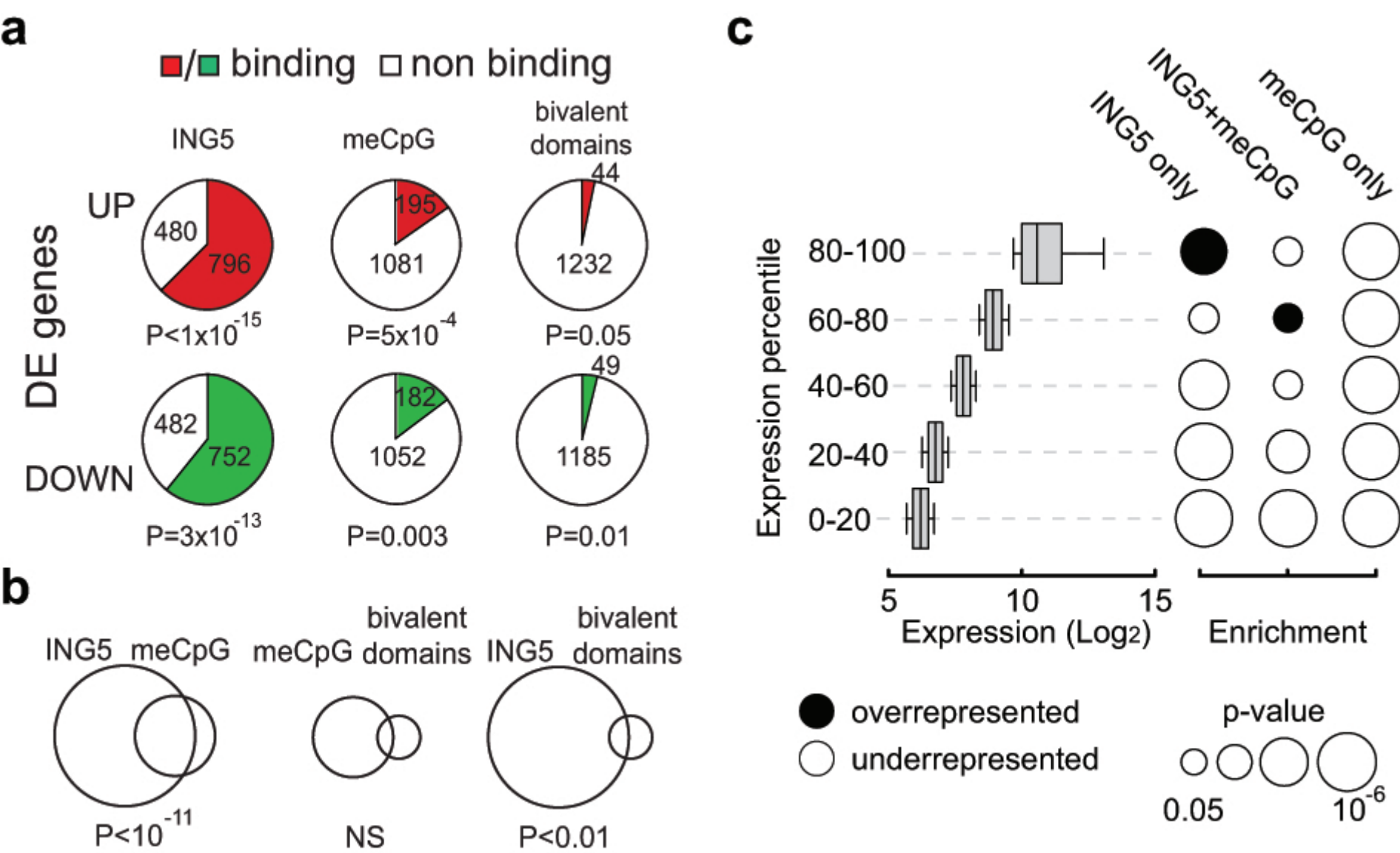
a
binding
non binding
ING5
meCpG
bivalent domains
DE genes
UP
DOWN
480
796
195
1081
44
1232
482
752
182
1052
49
1185
P<1x10-15
P=5x10-4
P=0.05
P=3x10-13
P=0.003
P=0.01
b
ING5
meCpG
bivalent domains
P<10-11
NS
P<0.01
c
ING5 only
ING5+meCpG
meCpG only
Expression percentile
80-100
60-80
40-60
20-40
0-20
5
10
15
Expression (Log2)
Enrichment
overrepresented
underrepresented
p-value
0.05
10-6

Figure 5

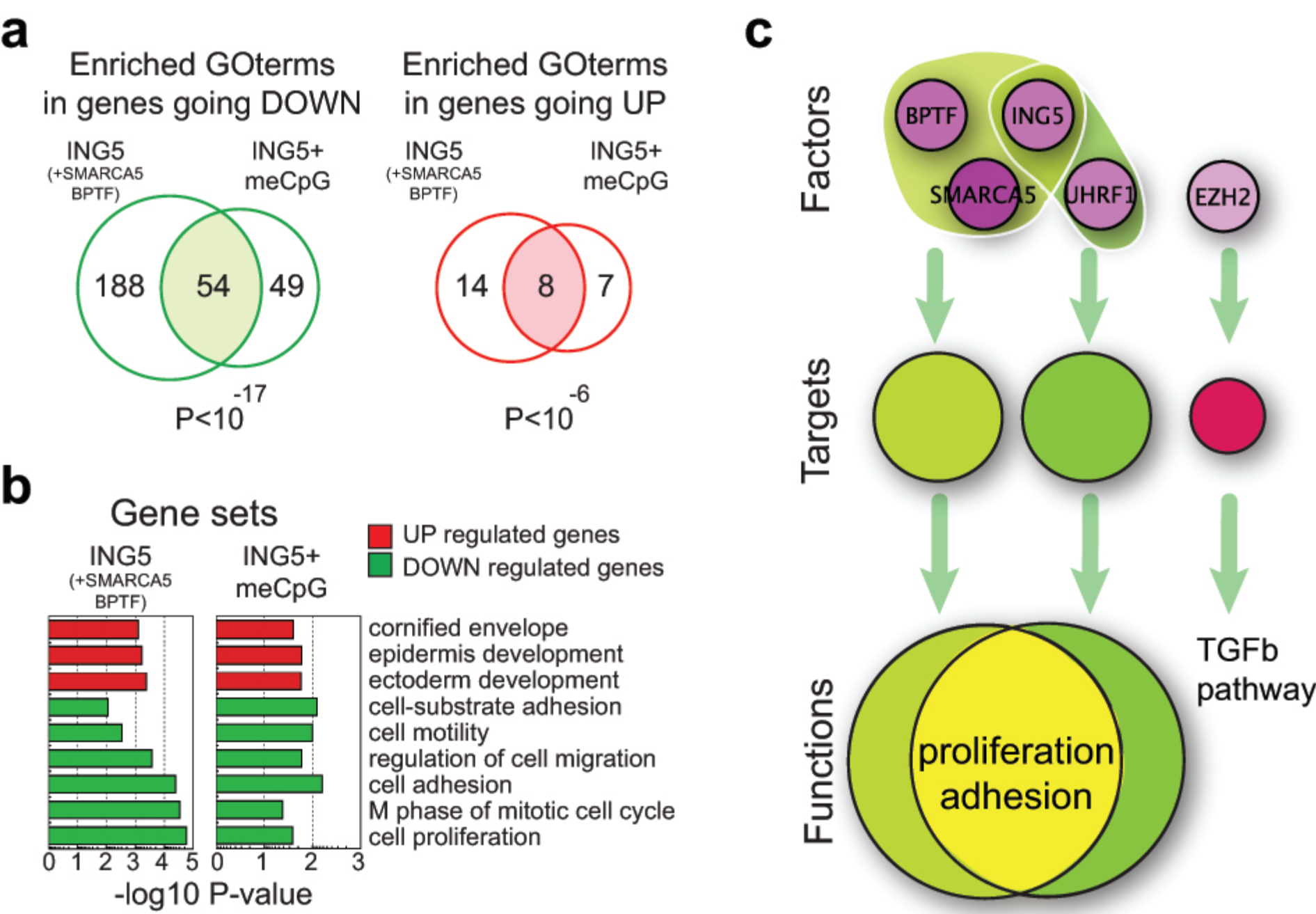
a
Enriched GOterms in genes going DOWN
ING5 (+SMARCA5 BPTF)
ING5+ meCpG
188
54
49
P<10-17
Enriched GOterms in genes going UP
ING5 (+SMARCA5 BPTF)
ING5+ meCpG
14
8
7
P<10-6
b
Gene sets
ING5 (+SMARCA5 BPTF)
ING5+ meCpG
UP regulated genes
DOWN regulated genes
cornified envelope
epidermis development
ectoderm development
cell-substrate adhesion
cell motility
regulation of cell migration
cell adhesion
M phase of mitotic cell cycle
cell proliferation
0 1 2 3 4 5
0 1 2 3
-log10 P-value
c
Factors
BPTF
ING5
SMARCA5
UHRF1
EZH2
Targets
Functions
proliferation
adhesion
TGFb pathway

Figure 6

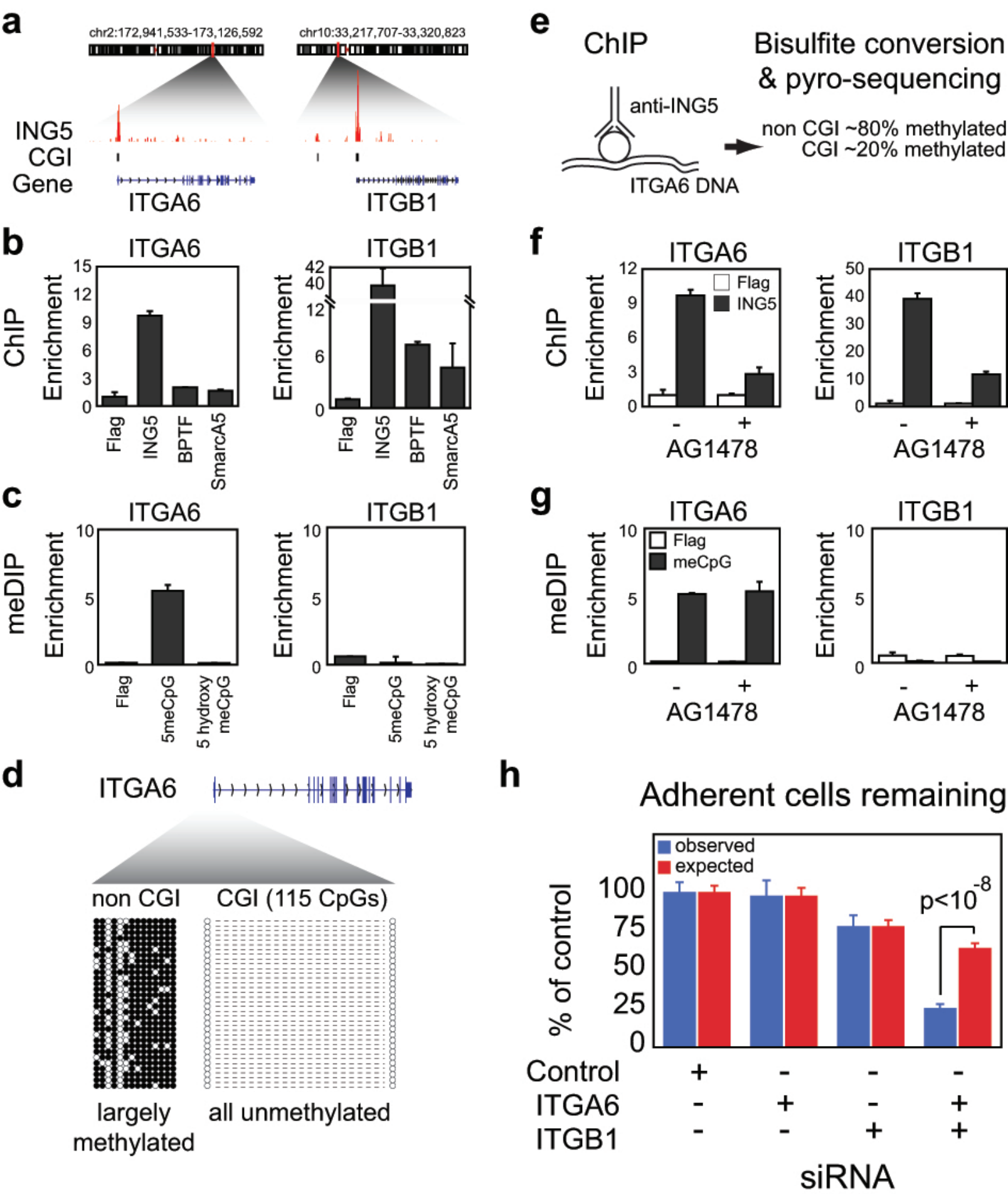
a
chr2:172,941,533-173,126,592
chr10:33,217,707-33,320,823
ING5
CGI
Gene
ITGA6
ITGB1
b
ITGA6
ITGB1
ChIP
Enrichment
Flag
ING5
BPTF
SmarcA5
c
meDIP
5meCpG
5 hydroxy meCpG
d
non CGI
CGI (115 CpGs)
largely methylated
all unmethylated
e
ChIP
anti-ING5
ITGA6 DNA
Bisulfite conversion & pyro-sequencing
non CGI ~80% methylated
CGI ~20% methylated
f
AG1478
g
meCpG
h
Adherent cells remaining
observed
expected
% of control
p<10-8
Control
ITGA6
ITGB1
siRNA

Figure 7

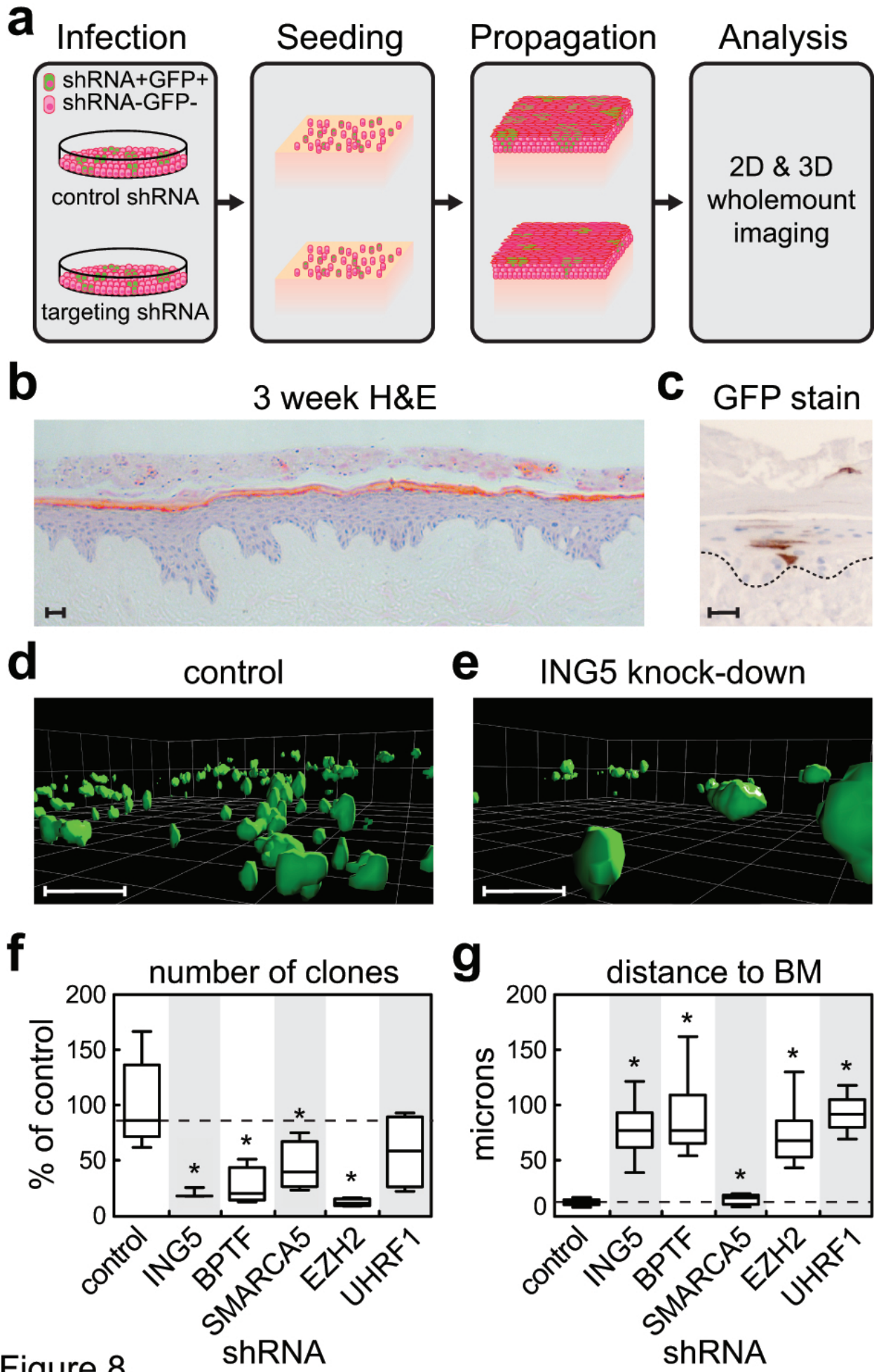
a
Infection
Seeding
Propagation
Analysis
shRNA+GFP+
shRNA-GFP-
control shRNA
targeting shRNA
2D & 3D wholemount imaging
b
3 week H&E
c
GFP stain
d
control
e
ING5 knock-down
f
number of clones
% of control
200
150
100
50
0
control
ING5
BPTF
SMARCA5
EZH2
UHRF1
shRNA
g
distance to BM
microns
200
150
100
50
0
control
ING5
BPTF
SMARCA5
EZH2
UHRF1
shRNA

Figure 8