

# WDR5, BRCA1, and BARD1 Co-regulate the DNA Damage Response and Modulate the Mesenchymal-to-Epithelial Transition during Early Reprogramming

Georgina Peñalosa-Ruiz,<sup>1</sup> Vicky Bousgouni,<sup>2</sup> Jan P. Gerlach,<sup>1</sup> Susan Waarlo,<sup>1</sup> Joris V. van de Ven,<sup>1</sup> Tim E. Veenstra,<sup>1</sup> José C.R. Silva,<sup>3</sup> Simon J. van Heeringen,<sup>1</sup> Chris Bakal,<sup>2</sup> Klaas W. Mulder,<sup>1,4,\*</sup> and Gert Jan C. Veenstra<sup>1,4,\*</sup>

<sup>1</sup>Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen 6500 HB, the Netherlands

<sup>2</sup>Dynamical Cell Systems Team, Division of Cancer Biology, Chester Beatty Laboratories Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

<sup>3</sup>Wellcome Trust Medical Research Council Cambridge Stem Cell Institute and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

<sup>4</sup>Co-senior author

\*Correspondence: [k.mulder@science.ru.nl](mailto:k.mulder@science.ru.nl) (K.W.M.), [g.veenstra@science.ru.nl](mailto:g.veenstra@science.ru.nl) (G.J.C.V.)  
<https://doi.org/10.1016/j.stemcr.2019.02.006>

## SUMMARY

Differentiated cells are epigenetically stable, but can be reprogrammed to pluripotency by expression of the OSKM transcription factors. Despite significant effort, relatively little is known about the cellular requirements for reprogramming and how they affect the properties of induced pluripotent stem cells. We have performed high-content screening with small interfering RNAs targeting 300 chromatin-associated factors and extracted colony-level quantitative features. This revealed five morphological phenotypes in early reprogramming, including one displaying large round colonies exhibiting an early block of reprogramming. Using RNA sequencing, we identified transcriptional changes associated with these phenotypes. Furthermore, double knockdown epistasis experiments revealed that BRCA1, BARD1, and WDR5 functionally interact and are required for the DNA damage response. In addition, the mesenchymal-to-epithelial transition is affected in *Brc1*, *Bard1*, and *Wdr5* knockdowns. Our data provide a resource of chromatin-associated factors in early reprogramming and underline colony morphology as an important high-dimensional readout for reprogramming quality.

## INTRODUCTION

Somatic cells can be reprogrammed to pluripotency by artificial expression of four transcription factors: OCT4, SOX2, KLF4, and c-MYC (OSKM) (Takahashi and Yamanaka, 2006). With varying efficiency, induced pluripotent stem cells (iPSCs) can be derived from a wide variety of cell types and they can differentiate into all cell lineages. Thus, they represent a promising resource for tissue regeneration and disease modeling.

Reprogramming occurs in two transcriptional and epigenetic waves (Polo et al., 2012). During the first wave, cell proliferation increases, and cells must overcome senescence and apoptosis (Xu et al., 2016). Dramatic metabolic changes and fast proliferation trigger the DNA damage response (DDR) through activation of the p53 pathway, inducing apoptosis on cells carrying substantial DNA damage (Marion et al., 2009). In agreement, DDR and DNA replication complexes are highly induced (Hansson et al., 2012), and DDR proteins, such as BRCA1, are required for efficient reprogramming (Gonzalez et al., 2013; Hansson et al., 2012). Senescence is also a barrier for reprogramming (Utikal et al., 2009). Some aging hallmarks, such as eroded telomeres (Lapasset et al., 2011; Marion and Blasco, 2010) and senescence-associated epigenetic marks (Ocampo

et al., 2016) need to be reset by OSKM for efficient reprogramming. At the transcriptional level, the somatic program is silenced concurrent with the acquisition of epithelial characteristics (Li et al., 2010; Samavarchi-Tehrani et al., 2010). Transcriptional changes during these stages are quite heterogeneous, whereas the pluripotency program is activated in a sequential and hierarchical manner (Buganim et al., 2012). Some of the early upregulated pluripotency genes include *SSEA1*, and the epithelial genes *Cdh1* and *Epcam*, and *Sall4*. At this point, reprogramming intermediates have only partially acquired the pluripotency program (Silva et al., 2008). Most of the cells will be trapped in such stages and only a small proportion will progress toward full pluripotency (Polo et al., 2012). Reactivation of *Sox2* (Buganim et al., 2012), *Nanog*, and *Esrrb* occurs during the second transcriptional and epigenetic wave and is also rate limiting to complete reprogramming (Apostolou and Stadtfeld, 2018).

During these transcriptional waves, chromatin dynamics involves the interplay of chromatin modifiers, transcriptional regulators and OSKM binding activities. Initially, OSK bind to open enhancers in mouse embryonic fibroblasts (MEFs) and consequently co-repressors, such as NCoR/SMRT are recruited to silence somatic genes (Zhuang et al., 2018). Also during early phases, H3K4me2 is rapidly





deposited at some pluripotency-associated loci (Xu et al., 2016). Accordingly, SET-MLL methyltransferase complexes, including their core component WDR5, have been shown to be crucial to facilitate reprogramming through H3K4me2/me3 deposition at pluripotency-associated regulatory regions (Ang et al., 2011; Wang et al., 2016). Other stem cell regulators reside within H3K9me3 heterochromatic domains (Apostolou and Stadtfeld, 2018). In concordance, activities of H3K9 methyl transferases EHMT1/2, SUV39H1/2, and SETDB1 constitute roadblocks of reprogramming (Soufi et al., 2012; Sridharan et al., 2013), whereas H3K9 demethylases such as KDM3A/B and KDM4C are facilitators (Chen et al., 2013). These and many other key chromatin regulators have been identified by RNAi (Cacchiarelli et al., 2015; Qin et al., 2014; Xu et al., 2016).

Despite the progress that has been made in characterizing the molecular changes during reprogramming, it is not fully understood how these dynamic changes are orchestrated. We have used high-content screening to assess the role of ~300 chromatin-associated proteins in colony phenotypes during early reprogramming. The combination of small interfering RNA (siRNA) screening with high-content microscopy allows simultaneous measurement of multiple morphological phenotypes and can reveal new associations among pathways (Fischer et al., 2015; Sero and Bakal, 2017). A similar approach has previously been used to define new gene networks involved in the final phase of iPSC formation (Golipour et al., 2012). We measured more than 20 colony features, including number of colonies, expression of early pluripotency markers, and other morphological and texture features, after individual knockdown of 300 chromatin modifiers. Selected hits from the primary screening were subjected to a transcriptome-based secondary screen. We identify several chromatin-associated genes that act together in the DDR and the mesenchymal-to-epithelial transition (MET) during early reprogramming.

## RESULTS

### High-Throughput Analysis of the Early Phase of Reprogramming

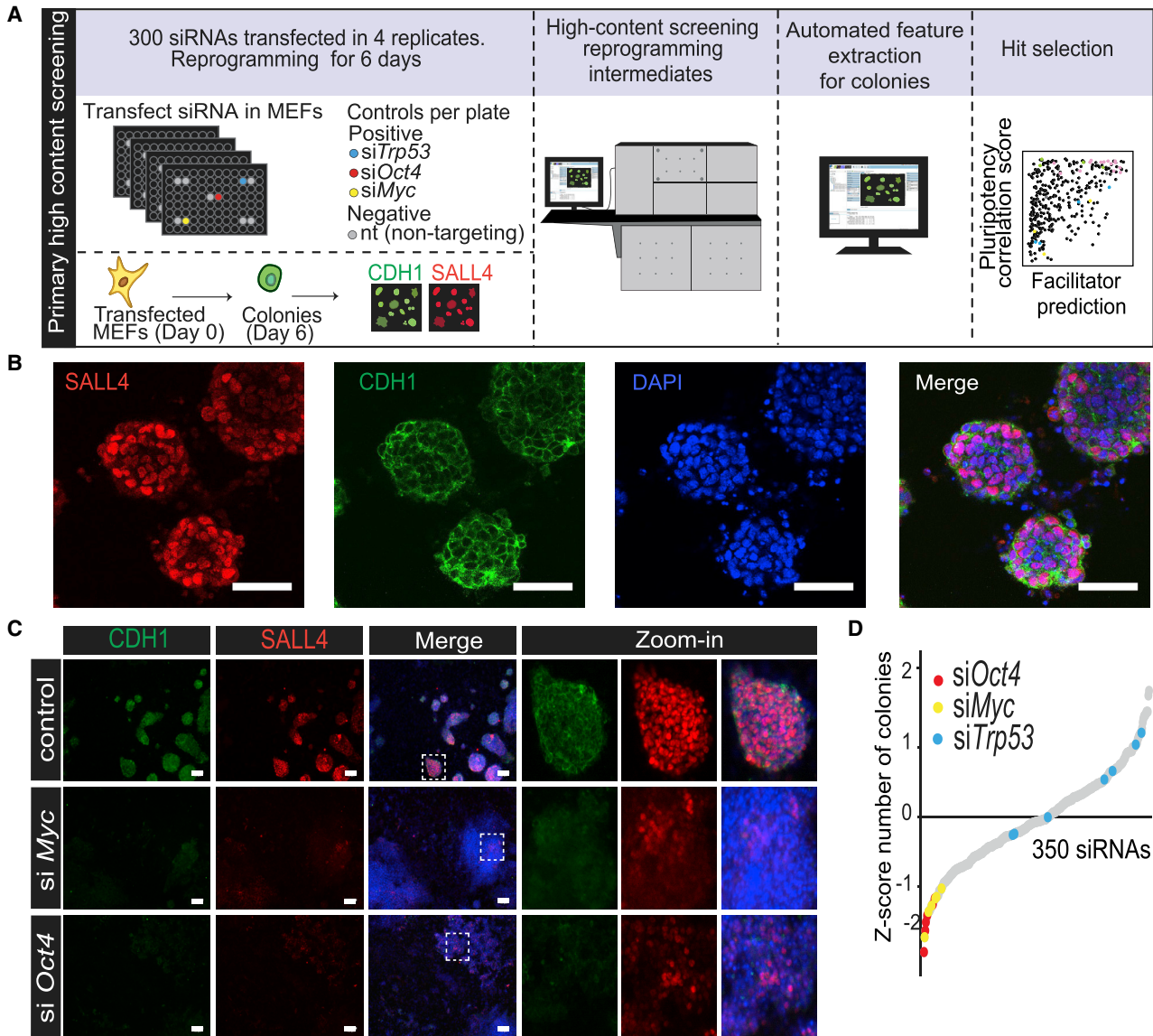
Reprogramming is associated with major changes in cell morphology, in part due to the MET (Li et al., 2010). Thus, we asked whether chromatin-mediated changes would affect reprogramming efficiency, colony morphology, and expression of early pluripotency markers. Moreover, we wondered how chromatin-associated factors might work together, as revealed by their similarities in a high-dimensional phenotypic space upon knockdown (Mulder et al., 2012; Wang et al., 2012). To define a set of

relevant chromatin-associated factors for an siRNA screen (Figure 1A), we used expression data (Chantzoura et al., 2015) to select genes with robust expression in MEFs or at least 4-fold upregulated expression in reprogramming cells. The custom siRNA library comprised siRNAs targeting 300 chromatin-associated factors. For each target, three independent siRNAs were pooled for transfections (Table S1).

We were specifically interested in the early phase of reprogramming, as chromatin is known to confer epigenetic stability to somatic cells. To test the function of chromatin-associated genes in early reprogramming, we used a relatively fast reprogramming system (Esteban et al., 2010; Vidal et al., 2014) in which colonies can be detected after 6 days of reprogramming (Figures 1B and S1A). Day 6 colonies show round, symmetric morphologies and robust expression of early markers CDH1, SSEA1, and SALL4, with gene expression of *Nanog* and *Esrrb* appearing later (Figures S1B–S1D). The specific staining of CDH1 and SALL4, respectively, at the cell surface and in the nucleus, strongly increased between days 3 and 6 (Figures 1B and S1), representing a suitable readout for the early phase of reprogramming.

The expression of genes was knocked down using siRNAs in MEFs infected with an inducible OSKM cassette lentivirus. Reprogramming was induced with doxycycline (dox) for 6 days (Figure S1A). The siRNA library consisted of six 96-well plates, with each plate containing seven non-targeting (nt) siRNA negative controls and three positive controls (siRNAs targeting *Trp53*, *Oct4*, and *c-Myc*). The screen was performed in quadruplicate. After 6 days of reprogramming, samples were fixed, stained for CDH1 and SALL4, and imaged using an automated high-content microscope. This allowed quantitation of morphology features such as colony size, symmetry and shape, and marker intensities, but also texture features. After data processing and colony feature extraction, the data were *Z* score normalized per plate (Bakal et al., 2007) and subjected to further analysis (Figure 1A).

To test the system, we disrupted reprogramming by knocking down OSKM factors *Oct4* (si*Oct4*) and *c-Myc* (si*Myc*). We also knocked down *Trp53* (si*Trp53*), which is expected to enhance reprogramming (Marion et al., 2009). si*Oct4* and si*Myc* colonies are flat and irregularly shaped, and show less intense SALL4 and CDH1 staining compared with the control (Figure 1C). As expected, the number of colonies observed in si*Oct4* and si*Myc* in our *Z* score-ranked data was low (Figure 1D). The si*Trp53* control showed a variable but positive effect on the number of colonies. These data confirm that colony morphology and early marker expression can be used as a readout for a disruption in the early reprogramming network. The screening focuses on early reprogramming events, and therefore it is unlikely to capture factors required later to attain full pluripotency.



**Figure 1. High-Throughput Analysis of the Early Phase of Reprogramming**

(A) Experimental design of high-content imaging siRNA screen.

(B) Representative image showing an immunofluorescence of reprogramming intermediates at day 6 stained for pluripotency markers CDH1 and SALL4 with DAPI counterstain. Scale bars, 50  $\mu$ m.

(C) Comparison of colony phenotypes of control, siMyc and siOct4 cells at reprogramming day 6, stained for SALL4 and CDH1. Scale bars, 100  $\mu$ m (images on the right, 4 $\times$  zoom-in of inset).

(D) siRNAs in the whole screen ranked from low to high Z scores, based on the number of colonies. Positive controls are highlighted in colors. Each siRNA represents the average Z score from four replicates (independent transfections in the same experiment). There are 350 siRNAs because controls are included in the rank.

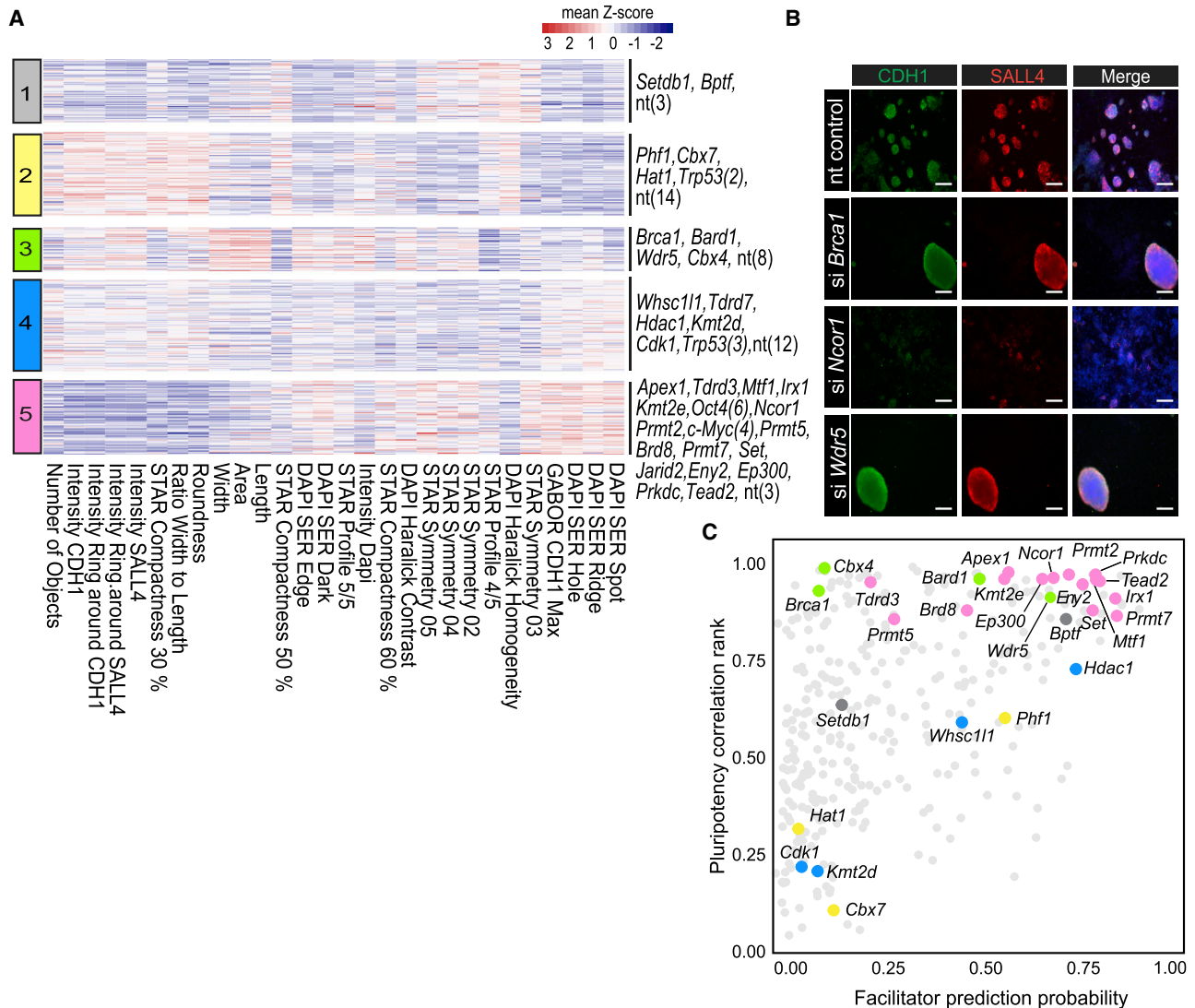
See also [Figure S1](#) and [Tables S1–S3](#).

### High-Content Microscopy Reveals Five Major Phenotypes of Colony Formation

The high-content analysis allowed us to measure a number of colony features ([Table S2](#)). This constitutes a multi-dimensional phenotypic space for analysis across many

conditions or perturbations ([Boutros et al., 2015](#)), and for the identification of functionally connected genes and processes ([Mulder et al., 2012](#); [Wang et al., 2012](#)).

We first defined the set of most discriminating features based on feature-to-feature pairwise correlations



**Figure 2. High-Content Microscopy Screen Reveals Five Major Phenotypes of Colony Formation**

(A) An average Z score for selected high-content features was calculated from quadruplicate samples (independent siRNA transfections from same experiment) and represented in a heatmap. Features are clustered by Euclidean distance and rows are clustered by K-means. Cluster number (left) and hits of the screen and the controls (right, number in brackets) are indicated.

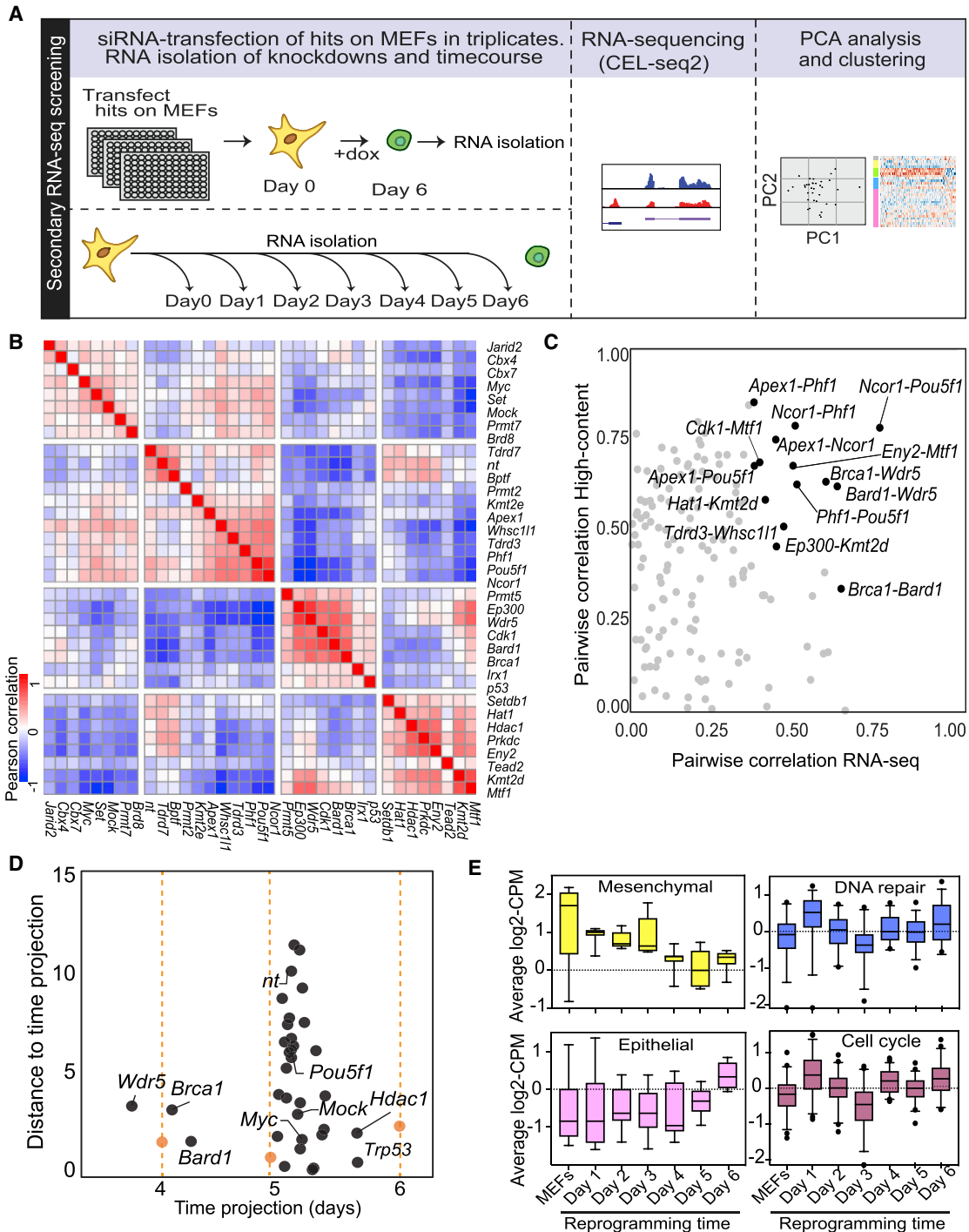
(B) Examples of knockdowns showing different phenotypes. Scale bars, 200  $\mu$ m.

(C) Pluripotency-associated hits were selected based on a combination of a probability prediction by machine learning based on known reprogramming facilitators (x), and a correlation analysis with the positive and negative controls (y). Selected top-hits are colored according to the cluster number (A; cf. Tables S3 and S4, Figure S2). Each data-point represents average of four replicates (A).

(Supplemental Information; Table S3). Using hierarchical (Figure S2) and K-means clustering (Figure 2A) we observed five main clusters that display different levels of pluripotency markers, number of colonies, symmetry features (ratio width to length, roundness), symmetry, threshold compactness, axial and radial (STAR) morphology features, and textural features (SER, Haralick, Gabor). Some textural features capture the distribution of intensity patterns across the image. These phenotypes

can be associated to structural changes in subcellular components (membranes, nucleoli, and chromatin). Other features or filters can detect how densely packed cells or cellular structures are (STAR features, see Supplemental Experimental Procedures). Although not all features have an intuitive biological interpretation, they are useful to discriminate cellular phenotypes and patterns not easily noted by the human eye in an unbiased manner (Boutros et al., 2015).





**Figure 3. Transcriptome-Based Secondary Screening and Time Course Expression Analysis**

(A) Overview of experimental design. Selected hits (30) and controls were transfected in triplicate and cultured until reprogramming day 6. The transcriptomes were analyzed together with a time series of control cells.

(B) siRNA-to-siRNA Pearson correlation heatmaps based on transcriptomes.

(C) Scatterplot representing pairwise siRNA correlations of gene expression values (x axis) and high-content imaging features (y axis). siRNA pairs with highest correlations in both approaches are highlighted.

(legend continued on next page)



Cluster 1 knockdowns have few colonies, in addition to lower intensities for SALL4 and CDH1 compared with clusters 2–4, suggesting a potential defect in reprogramming. Cluster 1 also shows lower scores for nuclear texture features (DAPI SER), suggesting structural changes in nuclei. More than half of the nt controls are found in cluster 2, which shows a high number of small, round, and compact colonies, and a robust expression of CDH1 and SALL4 (Figures 2A and 2B). Cluster 3 displays fewer, large colonies (high width, area, length scores), impaired compactness (STAR compactness), and detectable SALL4 and CDH1 expression (Figure 2A, cf. *Brca1* and *Wdr5*, Figure 2B). For cluster 4, no dramatic changes are observed compared with cluster 2, except for some DAPI SER texture features that score lower in cluster 2. Clusters 2 and 4 can be considered normal reprogramming, since they include most nt controls. However, they could still contain some enhancer factors, since *siTrp53* also resides there. Cluster 5 is characterized by substantially lower SALL4 and CDH1 staining intensities and a low number of irregularly shaped, less round (ratio with-to-length), and less compact colonies (cf. *Ncor1*, Figure 2B). Phenotypes of cluster 5 suggest a severe reprogramming impairment, as confirmed by the presence of *siMyc* and *siOct4* controls in that cluster.

Positive controls present in our screening (*Trp53*, *Myc*, and *Oct4*) are key for reprogramming. We reasoned that knockdown (high content) phenotypes similar to those of controls could potentially indicate a role in reprogramming progression. Therefore, we individually correlated each of the knockdowns to the positive controls (*Trp53*, *Myc*, and *Oct4*) based on all high-content features shown in Figure 2A using Pearson correlation (Table S4). Correlation scores of each knockdown to the positive controls were combined into a single ranking score for each knockdown (Figure 2C, y axis; cf. Experimental Procedures, Table S4). From this approach, we selected the 20 top-ranking knockdowns (Figure 2C).

In addition, the high-content phenotypes associated with known reprogramming facilitators (Table S4) present in our library (e.g., *Tet2*, *Jarid2*, and *Setdb1*), were used to train two independent machine-learning algorithms. Based on the high-content phenotypes of these facilitators, the algorithms classify the rest of the knockdowns, assigning them a score. Such a score indicates how well each knockdown is predicted to facilitate reprogramming (Figure 2C, x axis; cf. Supplemental Information; Table S4). We selected candidates from among the top-score predictions, but also some lower ranking siRNAs, because they

may represent roadblocks rather than facilitators of reprogramming (Figure 2C and Table S4). In total, 30 siRNAs were selected for an orthogonal transcriptome screen.

### A Transcriptome-Based Secondary Screening Uncovers Highly Correlated Phenotypes

We hypothesized that the phenotypes observed by microscopy might be reflected in their transcriptomes. Cells were transfected with siRNAs in triplicate, and day 6 RNA samples were subjected to CEL-Seq2-based RNA sequencing (Hashimshony et al., 2016). In addition to the 30 knockdowns, we also sequenced a day-by-day reprogramming time course of control cells (Figure 3A).

We performed principal-component analysis (PCA) to visualize transcriptome similarity based on the top 200 most variable transcripts across all samples (Figure S4A). Using these most variable transcripts, knockdown samples were clustered based on pairwise Pearson correlations (knockdown-to-knockdown; Figure 3B). This revealed groups of knockdowns with highly correlated transcriptomes, as noted in the intense red squares (Figure 3B). To analyze how the high-content phenotypes relate to the transcriptomes, we also calculated pairwise correlations of the same 30 siRNAs based on all high-content features (Figure 3C). This identified knockdown pairs that correlated in both their colony phenotype and their transcriptome (Figure 3C, highlighted in black). The strongest correlations were observed between the *Ncor1-Oct4* pair and a triplet consisting of *Wdr5*, *Brca1*, and *Bard1*. siRNAs with similar phenotypes (transcriptome and colony morphology) may reveal functional interactions. Indeed, NCoR1 has been shown to interact physically with OCT4 and MYC (Zhuang et al., 2018). However, the functional relationships between *Brca1*, *Bard1*, and *Wdr5* were unknown, therefore we decided to follow up on the effects of these siRNAs.

We performed siRNA deconvolution experiments measuring the number of SALL4-positive colonies of three independent siRNAs for *Wdr5*, *Brca1*, and *Bard1* to exclude off-target effects. This analysis resulted in phenotypes similar to the pooled siRNAs in at least two out of three siRNA sequences with the same target (Figure S3). In addition, high knockdown efficiencies of the *Brca1*, *Bard1*, and *Wdr5* mRNA targets were verified at day 3 of reprogramming (Figure S3). As reprogramming is a dynamic process, we wondered how cells progress toward the iPSC state in each of the knockdown conditions. Notably, in the PCA analysis, principal component 2 (PC2) correlates strongly

(D) Progression of reprogramming in knockdown cells (black) compared with cells of the time course (orange), based on PCA analysis of the transcriptomes and the projection of all data points on a curve fitted to the time course.

(E) Boxplots representing log transformed and normalized gene expression values from the CELSeq2 time course dataset show expression of different groups of genes (Experimental Procedures, Table S5, Figures S3 and S4). The box represents the interquartile range and the line is the median. (B)–(E) Averages of three independent RNA sequencing replicates.



with time ( $r^2 = 0.81$ ; Figures S4A and S4B). To model the reprogramming progression in each knockdown, we fitted a polynomial function to the time course data, and projected all other data on this fitted time line by shortest distance (Figures 3D and S4C; Experimental Procedures). The distance to the time line reflects transcriptome changes that are unrelated to progression of reprogramming. Such effects are to be expected as the siRNA library consisted of epigenetic and transcriptional regulators that may also affect processes not directly related to reprogramming. Most siRNA knockdowns, including the nt and mock transfected controls, have a transcriptome that is similar to control cells between day 5 and 6 of reprogramming, indicating a mild non-specific effect of transfection. Silencing *p53* and *Hdac1* modestly speeds up reprogramming relative to nt controls (Figure 3D). *Wdr5*, *Brca1*, and *Bard1*, show a strong delay in reprogramming with a short distance to the time projection of control cells, suggesting that most expression differences in these siRNA cells are due to the delay in reprogramming. si*Wdr5* cells were comparable with normal cells between day 3 and 4, while si*Bard1* and si*Brca1* cells were between day 4 and 5 (Figure 3D).

Next, we analyzed the gene expression changes associated with the colony morphologies of our high-content screen. For this, we identified the differentially expressed genes for each of the five morphological clusters (Figure 2A) using the secondary screen mRNA expression data. This highlighted a strong deregulated transcription program specific for morphological cluster 3 (containing *Wdr5*, *Brca1*, and *Bard1*) that was enriched for genes involved in cell-cell adhesion (Figure S4D; Table S5). Besides cell adhesion genes, this transcription program also contained *Zeb1* and *Twist2*, two regulators of mesenchymal cell fate. Indeed, genes in cluster 3 showed deregulation of mesenchymal and epithelial gene expression compared with the other clusters (Figure S4E; Table S5).

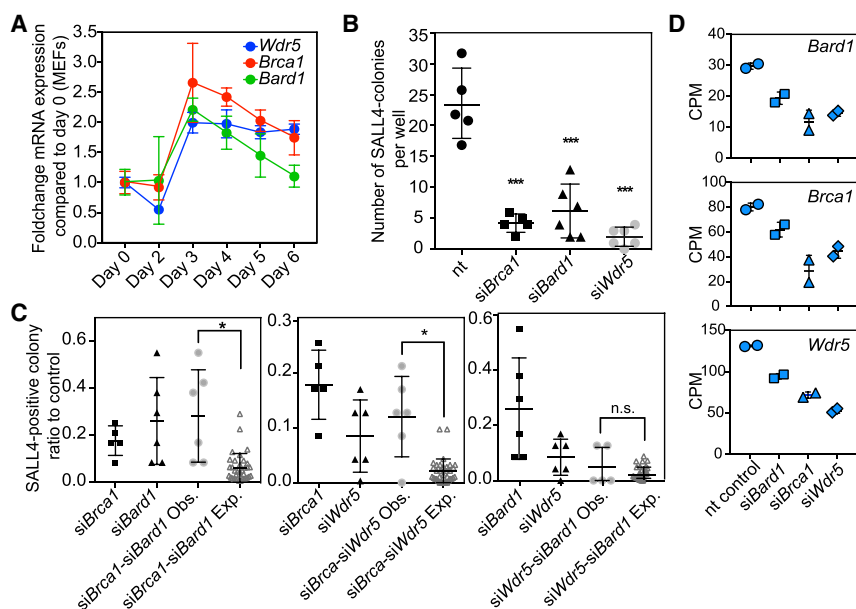
We then analyzed our time series data to relate the early block observed with si*Wdr5*, si*Brca1*, and si*Bard1* to the dynamics of genes involved in the MET, DNA repair, and cell-cycle regulation (Table S5). The block of reprogramming is observed around day 4 (Figure 3D), coinciding with the time of a major decrease of mesenchymal gene expression and preceding the activation of epithelial markers (Figure 3E). For DNA repair and cell-cycle genes there is an early wave of increased expression followed by downregulation, whereas a set of randomly selected genes are stably expressed over the time course of reprogramming (Figure S4F). This raised the possibility that WDR5, BRCA1, and BARD1 affect the repression of mesenchymal gene expression during early reprogramming. Moreover, based on the phenotypic and molecular co-correlation data we hypothesized that *Wdr5*, *Brca1*, and *Bard1* cooperate to control early stages of reprogramming.

### BRCA1, BARD1, and WDR5 Functionally Interact during Early Reprogramming

We asked whether *Wdr5*, *Bard1*, and *Brca1* genes have similar expression dynamics during early reprogramming. Interestingly, the three genes follow a similar qRT-PCR profile, peaking in expression at day 3, and then slowly going down (Figure 4A). To test the possibility of a functional interaction between these genes, the effect of double knockdowns was measured and compared with the effect of the single knockdowns regarding the number of colonies formed. In these experiments, the total amount of siRNA was kept the same (Experimental Procedures), preventing siRNA overloading and making sure that transfection conditions were comparable. All three single knockdowns displayed a significant reduction in number of SALL4-positive colonies compared with the nt control (Figure 4B). *Brca1-Bard1* double knockdown showed significantly more colonies than expected if the siRNAs were to have independent effects on the relative number of colonies (Figure 4C, left). This result was anticipated, as BRCA1-BARD1 are well-known physical interactors (Wu et al., 1996). Similarly, for both the *Wdr5-Brca1* and the *Wdr5-Bard1* double knockdowns, we also observed more colonies than expected, and this result was statistically significant for *Wdr5-Brca1* (Figure 4C). These results implicate the three genes in the same functional pathway. To test whether WDR5 is directly activating *Brca1* and *Bard1* gene expression, we determined the *Brca1* and *Bard1* expression levels after *Wdr5* knockdown (Figure 4D). Indeed, we found that this is the case at day 3, but also found that, in response to either *Bard1* or *Brca1* depletion, *Wdr5* expression was decreased. Taken together, *Brca1*, *Bard1*, and *Wdr5* are co-expressed, mutually depend on each other, and interact functionally in reprogramming.

### WDR5, BARD1, and BRCA1 Are Functionally Connected in the DDR Pathway

BRCA1 and BARD1 have a known function in double-strand break DNA repair. If BRCA1 and BARD1 functionally interact with WDR5, the prediction is that that all three knockdowns show an increase in DNA damage. The phosphorylated form of the histone variant H2A.X ( $\gamma$ H2A.X) represents a reliable biomarker for DNA damage as it is an immediate response to the presence of double-strand breaks (Sharma et al., 2012). Therefore, we employed fluorescence-activated cell sorting (FACS) analysis to measure  $\gamma$ H2A.X after knockdown (Figures 5A and 5B). Reprogramming cells (nt control) showed a significant decrease in DDR compared with non-reprogramming MEFs, in agreement with literature showing that reprogramming resolves DNA damage in somatic cells (Ocampo et al., 2016). Importantly, *Wdr5* knockdown showed a significantly increased level of  $\gamma$ H2A.X compared with the control (Figure 5A). Nearly 90% of the cells harbor  $\gamma$ H2A.X in *Wdr5*-depleted



**Figure 4. BRCA1, BARD1, and WDR5 Functionally Interact in Early Reprogramming**

(A) Gene expression of *Brca1*, *Bard1*, and *Wdr5* measured by qRT-PCR. Fold change was calculated relative to MEFs (day 0) gene expression. Each data point represents the mean value  $\pm$  SD of a duplicate from the same experiment.

(B) Dot plot showing the number of SALL4-positive colonies measured by in-cell western in control and *Brca1*, *Bard1*, and *Wdr5* knockdowns at day 6. Replicates are independent transfections from the same experiment. Statistical significance determined by one-way ANOVA (\*\*\*)  $p < 0.0005$ .

(C) SALL4-colony ratios of the single and double knockdowns compared with the non-targeting (nt) control, measured by in-cell western at day 6. Functional interaction is determined by comparing the mean difference in double knockdown colony ratios: observed (Obs.) versus expected (Exp.).

Replicates are independent transfections of an experiment performed at least twice. Statistical significance ( $*p < 0.05$ ) was calculated with two-tailed t-test.

(D) Dot plots to show *Wdr5*, *Brca1*, or *Bard1* gene expression as counts per million (CPM) reads in *siBard1*, *siBrca1*, *siWdr5*, and nt control. (B)–(D) Each dot or data point represents a replicate and the lines represent mean  $\pm$  SD.

cells (Figure 5B, bottom panel). As expected, *siBrca1* and *siBard1* also showed a high percentage of  $\gamma$ H2A.X-positive cells (Figures 5A and 5B). We also visualized  $\gamma$ H2A.X by immunofluorescence. At day 3 of reprogramming, knockdown cells and controls were stained for OCT4 and  $\gamma$ H2A.X (Figure 5C). In agreement with the results from the FACS analysis, nt control transfected (OCT4-positive) reprogramming cells showed strongly reduced focal  $\gamma$ H2A.X staining compared with the MEFs. In contrast, depletion of *Brca1*, *Bard1*, or *Wdr5* resulted in OCT4-positive,  $\gamma$ H2A.X-positive cells (Figure 5C). These results confirm the association of *Wdr5*, *Brca1*, and *Bard1* with the DDR during reprogramming.

### WDR5, BRCA1, and BARD1 Affect MET and DNA Repair Gene Expression

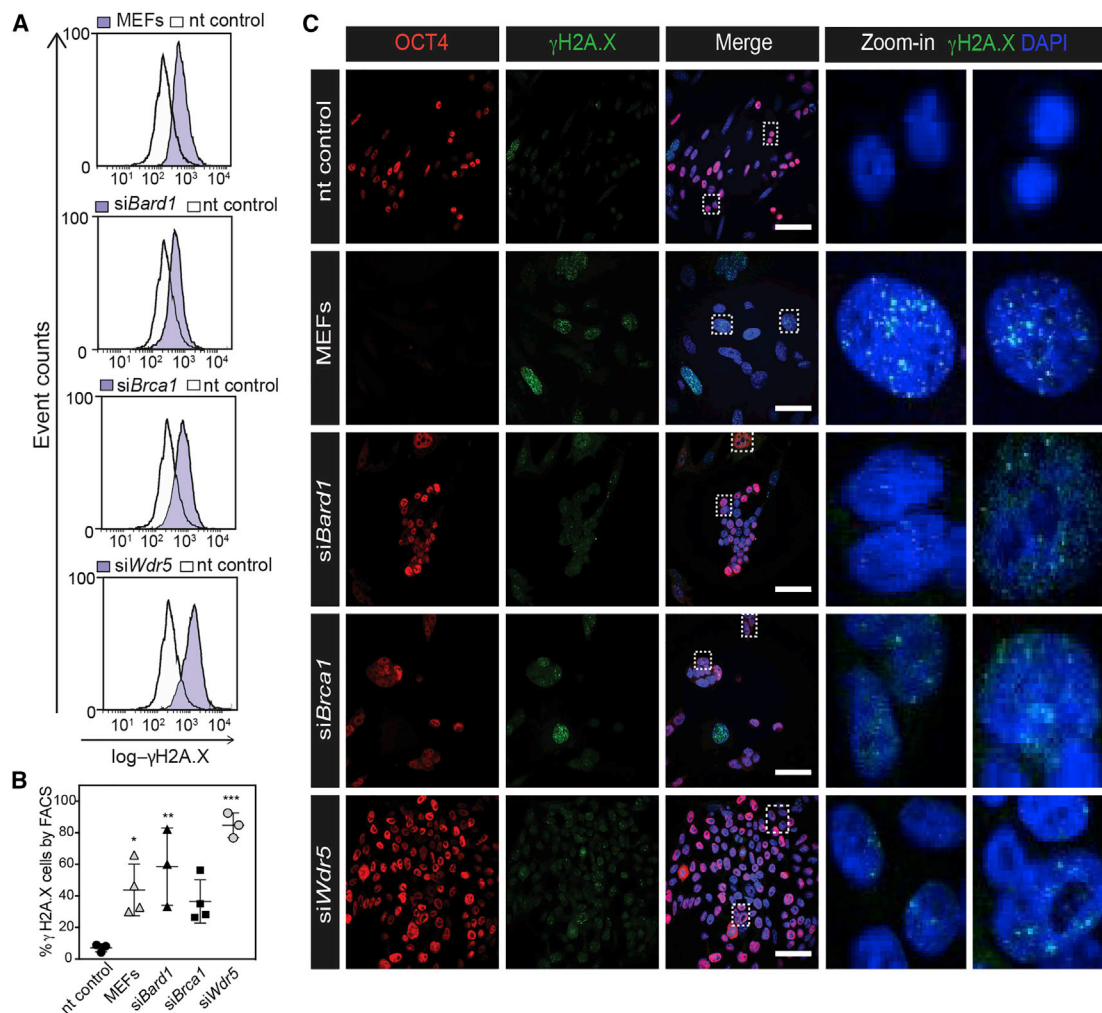
To gain more insight into the *Wdr5*, *Brca1*, and *Bard1* phenotypes and their link to MET (Figures 3E and S4C), we performed deep RNA sequencing at day 3 and day 6 of reprogramming. We called differentially expressed genes and found 753, 1,555, and 205 genes deregulated in, respectively, *Wdr5*, *Brca1*, and *Bard1* knockdown cells following 3 days of OSKM induction (Figure 6A). *Wdr5*-, *Brca1*-, and *Bard1*-depleted cells showed reduced expression of early pluripotency genes such as *Sall4*, and epithelial genes *Cdh1* and *Epcam* (Figure 6A).

Differentially expressed genes in each knockdown were further probed for overrepresented gene ontology (GO)

classes (Figure 6B; Table S6). *Brca1* knockdown causes a reduction in gene expression related to the cell cycle, response to DNA damage, and DNA repair (Figure 6B). We asked whether the effects on the DDR (Figure 5) are reflected in the transcriptome of *siWdr5* as well. To test this, DNA repair genes were probed in a gene set enrichment analysis (Mootha et al., 2003; Subramanian et al., 2005) comparing *siWdr5* and control transcriptomes. Indeed, the negative normalized enrichment score indicated decreased expression of DNA repair genes in the *siWdr5* compared with the control (Figure 6C, left). Furthermore, decreased expression of DNA repair genes in *siWdr5* was similar to that of *siBrca1* and *siBard1*, two well-known regulators of DDR (Figure 6C, right and Figure S5).

*Wdr5* and *Brca1* knockdowns shared a number of upregulated terms, including cell adhesion and developmental processes (e.g., skeleton or blood vessel development; Figure 6B). Regulation of cell proliferation was also affected in *Brca1*, *Bard1*, and *Wdr5* knockdowns. This GO term was enriched due to increased expression of, among others, *Tgfb-*, *Wnt-*, *Bmp-*, and *Fgf-*encoded growth factors (Table S6). In the RNA sequencing data from the knockdowns, we indeed observed higher expression values for these signaling factor genes compared with controls (Table S6; Figure 6D, left). These growth factors decrease cell proliferation and are involved in epithelial-to-mesenchymal transitions (EMT) (Barrallo-Gimeno and Nieto, 2005), potentially counteracting the MET required for reprogramming.





### Figure 5. WDR5, BARD1, and BRCA1 Are Functionally Connected in the DNA Damage Response Pathway

(A) Representative FACS histograms showing the cell distribution with log-intensity of  $\gamma$ H2A.X in reprogramming populations measured in different conditions (white, nt; purple, siRNA). Each sample was measured in at least three independent experiments.

(B) Dot plot representing the quantification of  $\gamma$ H2A.X-positive cells in each condition. Data points correspond to biological replicates from independent experiments. Statistical significance was determined by one-way ANOVA (\* $p < 0.05$ , \*\* $p < 0.005$ , \*\*\* $p < 0.0005$ ).

(C) Confocal images of reprogramming cells at day 3, stained for  $\gamma$ H2A.X (green) OCT4 (red), and DNA (DAPI, blue). Scale bars, 100  $\mu$ m (left-middle). Zoom-in (right): magnification from inset, showing characteristic  $\gamma$ H2A.X foci in all samples except nt control.

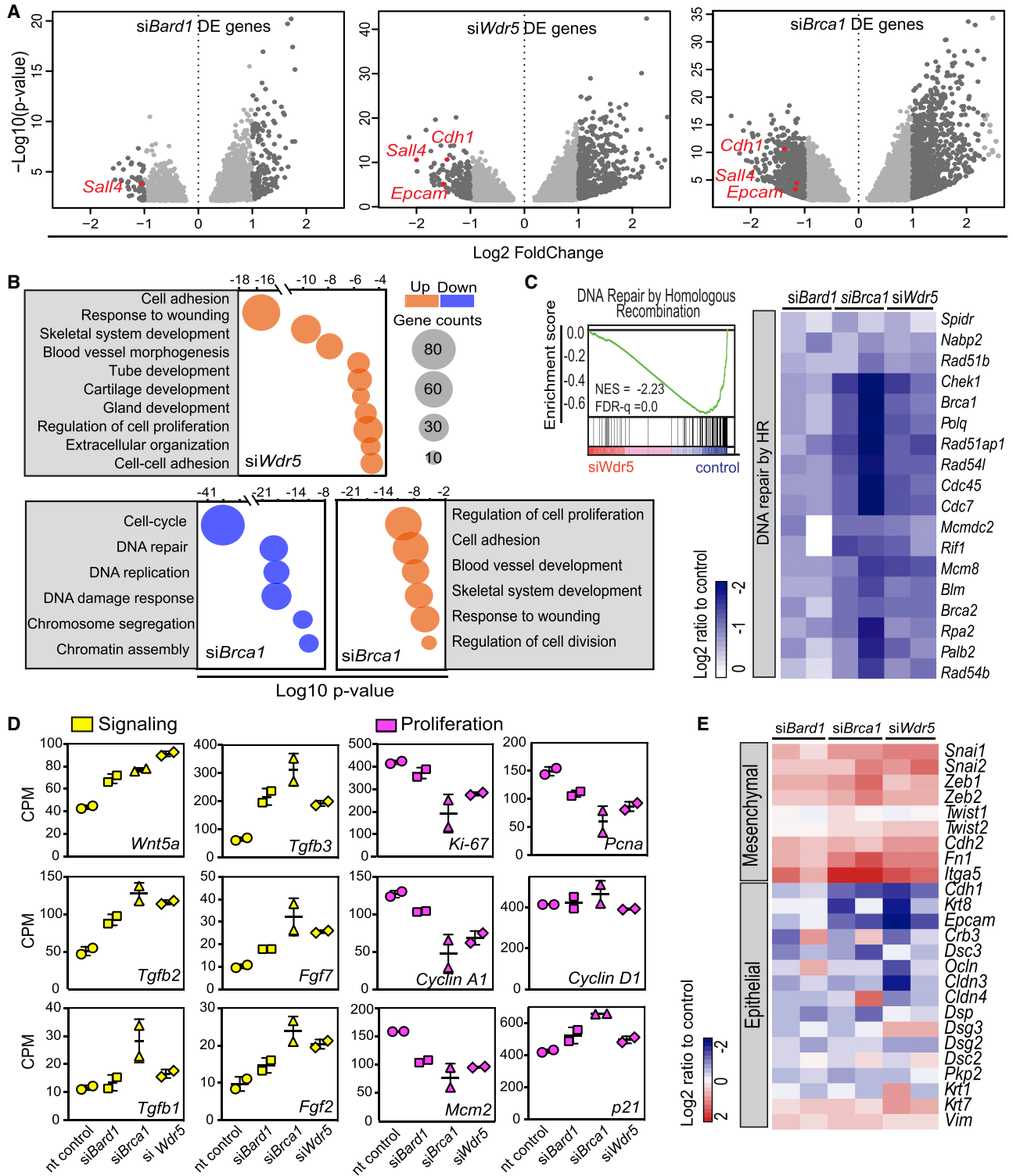
Several cell proliferation markers, including *Pcna*, *Ki-67*, and *Mcm2* showed decreased expression in all three knockdowns, while *p21* (*Cdkn1a*) was upregulated (Figure 6D, right). These data suggest an impaired MET upon *Brca1*, *Bard1*, and *Wdr5* depletion.

We assessed the gene expression levels of mesenchymal and epithelial markers in the three knockdowns and observed a clear increase in mesenchymal gene expression in the *Wdr5*, *Brca1*, and *Bard1* knockdown cells relative to control cells (Figures 6E and S5). Some epithelial genes were decreased (*Cdh1*, *Epcam*, and *Krt8*), whereas others did not change substantially or were increased (Figures 6E

and S5). Together, our data indicate that WDR5, BRCA1, and BARD1 cooperate in DDRs and that their absence affects MET progression during an early phase of reprogramming.

### DISCUSSION

Previously it has been shown that WDR5, a core component of SET-MLL methyltransferase complexes, interacts with OCT4 to activate the pluripotency network through H3K4 me2/me3 deposition (Ang et al., 2011). Our study shows that WDR5 also functionally interacts



**Figure 6. *Wdr5*, *Brca1*, and *Bard1* Depletion Affects Expression Profiles of MET and DNA Repair Genes**

(A) Volcano plots derived from two independent RNA sequencing replicates for *siBard1* (left), *siBrca1* (middle), and *siWdr5* differential gene expression at reprogramming day 3. Highlight: differentially expressed genes ( $\text{log}_2\text{-fold change} \geq 1$ , adjusted  $p < 0.05$ ).

(B) Bubble plot with enriched gene ontology terms (upregulated genes, orange; downregulated genes, blue) at day 3. Bubble sizes represent the number of genes.

(legend continued on next page)



with BRCA1-BARD1 to control DDR, and that MET is severely perturbed in the absence of these factors. The nature of the interaction remains to be elucidated. One possibility is that *Brca1* and *Bard1* are direct or indirect targets of the SET/MLL complexes, of which WDR5 is a subunit. In line with this possibility, chromatin immunoprecipitation analysis showed that WDR5 binds regulatory regions of *Brca1*, *Bard1*, and other genes involved in DNA repair (Ang et al., 2011). Moreover, we show that *Brca1* and *Bard1* transcripts are downregulated after silencing *Wdr5*. In addition, BRCA1 and BARD1 are involved in mitotic spindle organization and checkpoint gene regulation (Jin et al., 2009; Joukov et al., 2006; Wang et al., 2004), and MLL/WDR5 has been implicated in cell-cycle regulation, mitotic progression, and proper chromosome segregation as well (Ali et al., 2014, 2017; Liu et al., 2010).

DNA damage in reprogramming is, at least partly, associated with senescence (Ocampo et al., 2016), which can be rapidly induced by oxidative stress in primary cells (Xu et al., 2016). In agreement, low oxidizing conditions reduce the reprogramming barrier imposed by senescence (Utikal et al., 2009). BRCA1-BARD1 and WDR5 might alleviate a senescence-related block of reprogramming. Likewise, the requirement of a DDR could be related to the faster proliferation rates acquired early on in reprogramming (Polo et al., 2012; Ruiz et al., 2011). Embryonic stem cells, which proliferate in a similar fashion, indeed require additional genome surveillance mechanisms to cope with fast DNA replication (Ahuja et al., 2016).

Our study adds to the notion that colony morphology is linked to pluripotency (Abagnale et al., 2017; Kato et al., 2016; Narva et al., 2017) and is regulated by adhesion molecules, extracellular matrix, and cytoskeleton forces. Upon differentiation, these processes orchestrate morphological changes such as loss of colony compaction, increase of cell area, and colony flattening, together with changes in the pluripotency network (Narva et al., 2017). Therefore, colony morphology is an important readout for reprogramming quality. Moreover, screening of such multi-dimensional phenotypes is a powerful approach to identify functional interactions between genes. *Brca1*-, *Bard1*-, and *Wdr5*-depleted cells gave rise to fewer yet bigger, flat, symmetric colonies. One possibility is that these morphological changes are associated with a failure to downregulate

mesenchymal cell adhesion molecules. In addition, these cells fail to activate epithelial genes. Our study suggests a link between DDR and MET through *Brca1-Bard1* and *Wdr5* early in reprogramming. However, we cannot rule out that the MET phenotype is an indirect consequence of DDR impairment. Interestingly, the converse process of EMT may relate to DNA damage in kidney disease (Slaats et al., 2014) and cancer cells in culture (Chiba et al., 2012). Future work will further explore these relationships as well as gene-gene interactions that modify the phenotypic plasticity of reprogramming to induced pluripotency.

## EXPERIMENTAL PROCEDURES

### Data and Software Availability

All sequencing data are available at the GEO repository Superseries number GSE118680.

The code to reproduce reprogramming facilitator predictions by machine learning is available at <https://github.com/simonvh/facilitators-penalosa-ruiz/>. The code to reproduce the timeline projection is available at <https://github.com/TimeVeenstra/Time-Curve-Projection/> (<https://doi.org/10.5281/zenodo.1405746>).

### MEF Reprogramming and Culture Media

Passage 1–2 MEFs were seeded at a density of 10,000 cells per cm<sup>2</sup>. Next day, MEFs were transduced at an MOI of 1 with Tet-STEMCCA lentivirus (Sommer et al., 2009), rTA (Addgene, no. 20342), and 8 μg mL<sup>-1</sup> polybrene. Next day (day 0), cells were transferred to either 1% gelatin-coated plates or mitotically inactive feeder cells, in reprogramming medium (Vidal et al., 2014).

### siRNA Transfections and siRNA Screenings

A custom Silencer siRNA library targeting around 300 mouse genes encoding chromatin factors was designed (Thermo Scientific/Ambion, Table S1) and distributed in six plates. Each gene in the library was targeted with three different siRNAs, which were pooled for transfection. For the high-content screening, the six pooled plates were transfected in quadruplicate. Every plate contained the following controls: *siOct4* (*siPou5f1*), *siMyc*, *siTrp53*, and seven nt controls. Reverse transfections in a 96-well plate format were performed as follows: 20 μL of transfection mix was prepared in each well before adding the cell suspension. This transfection mix consisted of 40 nM of pooled siRNAs (considering 120 μL final volume), and 0.26 μL RNAiMAX lipofectamine (Thermo Scientific) diluted in OptiMem (Thermo Scientific). After incubation for 10 min, 100 μL of cell suspension (3,000–6,000

(C) Gene set enrichment analysis for DNA repair by homologous recombination (HR) comparing *siWdr5* versus control transcriptomes (left). NES, normalized enrichment score; FDR, false discovery rate. Heatmap for *siBrca1*, *siBard1*, and *siWdr5* samples showing genes for DNA repair by HR (log<sub>2</sub>-ratio relative to control).

(D) Gene expression (RNA sequencing CPM) of signaling genes (magenta) and cell proliferation markers (yellow) in control, *siBrca1*, *siBard1*, and *siWdr5* cells (each data point represents individual RNA sequencing replicates from independent experiments; bars, average ± SD).

(E) Heatmap representing the log<sub>2</sub>-ratio of mesenchymal and epithelial gene expression of the three knockdowns relative to control. See Figure S5 and Table S6.



cells) were added to each well. For transfections in a six-well plate format, the protocol was scaled up accordingly. Before adding 1.8 mL cell suspension with 100,000 cells, 220  $\mu\text{L}$  transfection mix was incubated in the wells for 10 min. The transfection mix consisted of 4  $\mu\text{L}$  RNAiMAX and a final concentration of 40 nM siRNA, all diluted in OptiMem.

### Immunostaining

Cells were cultured in 96-well Cell Carrier plates for microscopy (PerkinElmer). After 6 days of reprogramming, cells were washed with PBS and fixed with 4% paraformaldehyde (PFA) for 15 min. After blocking and permeabilization, samples were incubated overnight with mouse anti-CDH1 (Cell Signaling, 14472) and then with goat anti-mouse Alexa 488 for 2 h. Staining with rabbit anti-SALL4 (Abcam, ab29112) was done overnight, followed by 3 h incubation with goat anti-rabbit Alexa 568 and 40  $\mu\text{g mL}^{-1}$  DAPI. After antibody incubations, the cells were washed twice with PBS.

### High-Content Image Acquisition and Feature Selection

Plates were imaged with an Opera High-content Screening System (PerkinElmer) with a 4 $\times$  air lens. Images were imported into the Columbus software platform (PerkinElmer). To segment colonies imaged on multiple z planes, we used the maximum projection of z planes. SALL4 staining was used to find and segment the colonies. Automated image analysis was used for image region segmentation and for extraction of shape and morphology features. Image regions touching the edge were removed. For more details, see [Supplemental Information](#). After extracting all features for every plate from the automated pipeline, a Z score normalization was applied per plate (Bakal et al., 2007) based on the mean values per feature. To select relevant features, a feature-to-feature Pearson correlation was calculated. Features with a high pairwise correlation (>0.8) were considered redundant.

### RNA Sequencing and Analysis

CEL-seq2 sample preparation (Hashimshony et al., 2016) was performed with a few adaptations (see [Supplemental Information](#)). Transcripts were mapped to *Mus musculus* genome version mm10 with Bowtie2 (Langmead and Salzberg, 2012), UMI corrected using standard settings of the CELseq2 pipeline (<https://github.com/yanaailab/CEL-Seq-pipeline>), and matched to the gencode.vM13.annotation transcriptome. To relate knockdown data points to the progression of reprogramming, the transcriptomes were subjected to PCA. PC1 and PC2 were swapped (x axis: PC2) and all data (knockdown and time series) were rotated 15 $^\circ$  (cf. [Figure S4B](#)). A second-order polynomial curve was fitted to the time series (days 2–7), and all data points were projected on this curve (script: <https://doi.org/10.5281/zenodo.1405747>). For each data point, the projected x coordinate was used as a proxy for time, whereas the distance to the fitted time line (calculated using Pythagoras' theorem) was used as a proxy for gene expression differences unrelated to the process of reprogramming.

### FACS Analysis of DNA Damage

Reprogramming MEFs were transfected with siRNAs in six-well plates. After 3 days, cells were fixed on ice with 1% PFA for

15 min and incubated with 70% ice-cold ethanol at  $-20^\circ\text{C}$  for 2 h. Samples were then incubated with 100  $\mu\text{L}$  mouse anti-phospho-H2AX (Millipore, diluted 1:100 in 0.25% BSA 0.3% Triton/PBS) overnight at 4 $^\circ\text{C}$ . Subsequently, cells were washed and stained with 100  $\mu\text{L}$  Alexa 488 goat anti-rabbit 488 (diluted 1:500) for 2 h at room temperature. Finally, samples were incubated with propidium iodide overnight in a fridge and sorted using an FC 500 (Beckman Coulter) machine. Data analysis was done with Flowing software v.2.5. As positive control, reprogramming MEFs were treated with 400  $\mu\text{g mL}^{-1}$  mitomycin C for 3 days.

### Double Knockdowns and Functional Interactions

Reprogramming was started in 48-well plate formats, with transfection reagents and number of cells scaled accordingly. For the double knockdown, a mixture of two targeting siRNAs was used in a final concentration of 40 nM, meaning 20 nM of each siRNA. The corresponding single knockdowns were performed with 20 nM siRNA target + 20 nM nt control siRNA, to make it equivalent to the individual siRNA dose in the double knockdowns. In this way, individual as well as final siRNA concentrations are comparable in double and single knockdowns (Mulder et al., 2012). The observed SALL4-colony ratio was calculated dividing the double knockdown number of colonies by the average number of colonies of the control (six biological replicates). The expected SALL4-colony ratio was calculated by multiplying the ratios of the single knockdowns (Mani et al., 2008). A p value of <0.05 (two-tailed t-test) was considered significant. See [Supplemental Information](#) for details.

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.stemcr.2019.02.006>.

### AUTHOR CONTRIBUTIONS

Conceptualization, G.J.C.V., K.W.M., and G.P.-R.; Methodology, G.P.-R., V.B., G.J.C.V., K.W.M., C.B., and J.C.R.S.; Experiments, G.P.-R., V.B., J.P.G., S.W., and J.V.v.d.V.; Analysis, G.P.-R., V.B., J.P.G., G.J.C.V., S.J.v.H., and T.E.V.; Writing, G.P.-R., G.J.C.V., and K.W.M., with input from all authors.

### ACKNOWLEDGMENTS

The authors thank Dei M. Elurbe for useful suggestions on data analysis and processing of the sequencing files, Georgios Georgiou for bioinformatics support, Jessie A.G. van Buggenum for help with the colony counting script. Siebe van Genesen, Ann Rose Bright, Katie Tremble, and E. Janssen-Megens for valuable technical help and all Veenstra lab members for their input and help. V.B. and C.B. are funded by the Stand Up to Cancer campaign for Cancer Research UK, and Cancer Research UK Program Foundation Award to C.B. (C37275/1A20146). K.M. was supported by an NWO-VIDI grant (864.12.010).

Received: September 18, 2018

Revised: February 14, 2019

Accepted: February 14, 2019

Published: March 14, 2019





## REFERENCES

- Abagnale, G., Sechi, A., Steger, M., Zhou, Q., Kuo, C.C., Aydin, G., Schalla, C., Muller-Newen, G., Zenke, M., Costa, I.G., et al. (2017). Surface topography guides morphology and spatial patterning of induced pluripotent stem cell colonies. *Stem Cell Reports* *9*, 654–666.
- Ahuja, A.K., Jodkowska, K., Teloni, F., Bizard, A.H., Zellweger, R., Herrador, R., Ortega, S., Hickson, I.D., Altmeyer, M., Mendez, J., et al. (2016). A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. *Nat. Commun.* *7*, 10660.
- Ali, A., Veeranki, S.N., Chinchole, A., and Tyagi, S. (2017). MLL/WDR5 complex regulates Kif2A localization to ensure chromosome congression and proper spindle assembly during mitosis. *Dev. Cell* *41*, 605–622.e7.
- Ali, A., Veeranki, S.N., and Tyagi, S. (2014). A SET-domain-independent role of WRAD complex in cell-cycle regulatory function of mixed lineage leukemia. *Nucleic Acids Res.* *42*, 7611–7624.
- Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* *145*, 183–197.
- Apostolou, E., and Stadtfeld, M. (2018). Cellular trajectories and molecular mechanisms of iPSC reprogramming. *Curr. Opin. Genet. Dev.* *52*, 77–85.
- Bakal, C., Aach, J., Church, G., and Perrimon, N. (2007). Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* *316*, 1753–1756.
- Barrallo-Gimeno, A., and Nieto, M.A. (2005). The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* *132*, 3151–3161.
- Boutros, M., Heigwer, F., and Laufer, C. (2015). Microscopy-based high-content screening. *Cell* *163*, 1314–1325.
- Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* *150*, 1209–1222.
- Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrao, S., Zhang, X., et al. (2015). Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. *Cell* *162*, 412–424.
- Chantzoura, E., Skylaki, S., Menendez, S., Kim, S.I., Johnsson, A., Linnarsson, S., Woltjen, K., Chambers, I., and Kaji, K. (2015). Reprogramming roadblocks are system dependent. *Stem Cell Reports* *5*, 350–364.
- Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* *45*, 34–42.
- Chiba, N., Comaills, V., Shiotani, B., Takahashi, F., Shimada, T., Tajima, K., Winokur, D., Hayashida, T., Willers, H., Brachtel, E., et al. (2012). Homeobox B9 induces epithelial-to-mesenchymal transition-associated radioresistance by accelerating DNA damage responses. *Proc. Natl. Acad. Sci. U S A* *109*, 2760–2765.
- Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S., et al. (2010). Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* *6*, 71–79.
- Fischer, B., Sandmann, T., Horn, T., Billmann, M., Chaudhary, V., Huber, W., and Boutros, M. (2015). A map of directional genetic interactions in a metazoan cell. *Elife* *4*. <https://doi.org/10.7554/eLife.05464>.
- Golipour, A., David, L., Liu, Y., Jayakumar, G., Hirsch, C.L., Trcka, D., and Wrana, J.L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* *11*, 769–782.
- Gonzalez, F., Georgieva, D., Vanoli, F., Shi, Z.D., Stadtfeld, M., Ludwig, T., Jasin, M., and Huangfu, D. (2013). Homologous recombination DNA repair genes play a critical role in reprogramming to a pluripotent state. *Cell Rep.* *3*, 651–660.
- Hansson, J., Rafiee, M.R., Reiland, S., Polo, J.M., Gehring, J., Okawa, S., Huber, W., Hochedlinger, K., and Krijgsvelde, J. (2012). Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep.* *2*, 1579–1592.
- Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* *17*, 77.
- Jin, S., Gao, H., Mazzacurati, L., Wang, Y., Fan, W., Chen, Q., Yu, W., Wang, M., Zhu, X., Zhang, C., et al. (2009). BRCA1 interaction of centrosomal protein Nlp is required for successful mitotic progression. *J. Biol. Chem.* *284*, 22970–22977.
- Joukov, V., Groen, A.C., Prokhorova, T., Gerson, R., White, E., Rodriguez, A., Walter, J.C., and Livingston, D.M. (2006). The BRCA1/BARD1 heterodimer modulates ran-dependent mitotic spindle assembly. *Cell* *127*, 539–552.
- Kato, R., Matsumoto, M., Sasaki, H., Joto, R., Okada, M., Ikeda, Y., Kanie, K., Suga, M., Kinehara, M., Yanagihara, K., et al. (2016). Parametric analysis of colony morphology of non-labelled live human pluripotent stem cells for cell quality control. *Sci. Rep.* *6*, 34009.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.
- Lapasset, L., Milhabet, O., Prieur, A., Besnard, E., Babled, A., Ait-Hamou, N., Leschik, J., Pellestor, F., Ramirez, J.M., De Vos, J., et al. (2011). Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state. *Genes Dev.* *25*, 2248–2253.
- Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* *7*, 51–63.
- Liu, H., Takeda, S., Kumar, R., Westergard, T.D., Brown, E.J., Pandita, T.K., Cheng, E.H., and Hsieh, J.J. (2010). Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature* *467*, 343–346.



- Mani, R., St Onge, R.P., Hartman, J.L.t., Gjaever, G., and Roth, F.P. (2008). Defining genetic interaction. *Proc. Natl. Acad. Sci. U S A* *105*, 3461–3466.
- Marion, R.M., and Blasco, M.A. (2010). Telomere rejuvenation during nuclear reprogramming. *Curr. Opin. Genet. Dev.* *20*, 190–196.
- Marion, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPSC cell genomic integrity. *Nature* *460*, 1149–1153.
- Mootha, V.K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F., et al. (2003). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc. Natl. Acad. Sci. U S A* *100*, 605–610.
- Mulder, K.W., Wang, X., Escriu, C., Ito, Y., Schwarz, R.F., Gillis, J., Sirokmany, G., Donati, G., Uribe-Lewis, S., Pavlidis, P., et al. (2012). Diverse epigenetic strategies interact to control epidermal differentiation. *Nat. Cell Biol.* *14*, 753–763.
- Narva, E., Stubb, A., Guzman, C., Blomqvist, M., Balboa, D., Lerche, M., Saari, M., Otonkoski, T., and Ivaska, J. (2017). A strong contractile actin fence and large adhesions direct human pluripotent colony morphology and adhesion. *Stem Cell Reports* *9*, 67–76.
- Ocampo, A., Reddy, P., Martinez-Redondo, P., Platero-Luengo, A., Hatanaka, F., Hishida, T., Li, M., Lam, D., Kurita, M., Beyret, E., et al. (2016). In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell* *167*, 1719–1733.e2.
- Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPSC cells. *Cell* *151*, 1617–1632.
- Qin, H., Diaz, A., Blouin, L., Lebbink, R.J., Patena, W., Tanbun, P., LeProust, E.M., McManus, M.T., Song, J.S., and Ramalho-Santos, M. (2014). Systematic identification of barriers to human iPSC generation. *Cell* *158*, 449–461.
- Ruiz, S., Panopoulos, A.D., Herreras, A., Bissig, K.D., Lutz, M., Berggren, W.T., Verma, I.M., and Izpisua Belmonte, J.C. (2011). A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr. Biol.* *21*, 45–52.
- Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* *7*, 64–77.
- Sero, J.E., and Bakal, C. (2017). Multiparametric analysis of cell shape demonstrates that beta-PIX directly couples YAP activation to extracellular matrix adhesion. *Cell Syst.* *4*, 84–96.e6.
- Sharma, A., Singh, K., and Almasan, A. (2012). Histone H2AX phosphorylation: a marker for DNA damage. *Methods Mol. Biol.* *920*, 613–626.
- Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* *6*, e253.
- Slaats, G.G., Ghosh, A.K., Falke, L.L., Le Corre, S., Shaltiel, I.A., van de Hoek, G., Klasson, T.D., Stokman, M.F., Logister, I., Verhaar, M.C., et al. (2014). Nephronophthisis-associated CEP164 regulates cell cycle progression, apoptosis and epithelial-to-mesenchymal transition. *PLoS Genet.* *10*, e1004594.
- Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* *27*, 543–549.
- Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* *151*, 994–1004.
- Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. *Nat. Cell Biol.* *15*, 872–882.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* *102*, 15545–15550.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* *126*, 663–676.
- Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulal, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPSCs. *Nature* *460*, 1145–1148.
- Vidal, S.E., Amlani, B., Chen, T., Tsigos, A., and Stadtfeld, M. (2014). Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. *Stem Cell Reports* *3*, 574–584.
- Wang, C., Lee, J.E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016). Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc. Natl. Acad. Sci. U S A* *113*, 11871–11876.
- Wang, R.H., Yu, H., and Deng, C.X. (2004). A requirement for breast-cancer-associated gene 1 (BRCA1) in the spindle checkpoint. *Proc. Natl. Acad. Sci. U S A* *101*, 17108–17113.
- Wang, X., Castro, M.A., Mulder, K.W., and Markowitz, F. (2012). Posterior association networks and functional modules inferred from rich phenotypes of gene perturbations. *PLoS Comput. Biol.* *8*, e1002566.
- Wu, L.C., Wang, Z.W., Tsan, J.T., Spillman, M.A., Phung, A., Xu, X.L., Yang, M.C., Hwang, L.Y., Bowcock, A.M., and Baer, R. (1996). Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nat. Genet.* *14*, 430–440.
- Xu, Y., Zhang, M., Li, W., Zhu, X., Bao, X., Qin, B., Hutchins, A.P., and Esteban, M.A. (2016). Transcriptional control of somatic cell reprogramming. *Trends Cell Biol.* *26*, 272–288.
- Zhuang, Q., Li, W., Benda, C., Huang, Z., Ahmed, T., Liu, P., Guo, X., Ibanez, D.P., Luo, Z., Zhang, M., et al. (2018). NCoR/SMRT corepressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. *Nat. Cell Biol.* *20*, 400–412.