SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
EPIGENETICS
TUMOUR SUPPRESSORS
STEM CELLS
TRANSCRIPTOME

Received
7 June 2012

Accepted
6 August 2012

Published
21 August 2012

Correspondence and requests for materials should be addressed to X.W. (bchwxy@nus.edu.sg)

* These authors contributed equally.

# Reversal of Aberrant Cancer Methylome and Transcriptome upon Direct Reprogramming of Lung Cancer Cells

Dashayini Mahalingam[1]*, Chiou Mee Kong[1]*, Jason Lai[1]*, Ling Lee Tay[1], Henry Yang[2] & Xueying Wang[1]

[1]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, [2]Cancer Science Institute of Singapore, National University of Singapore, Singapore.

**Recent reports on direct reprogramming of cancer cells (iPCs) which results in reduced tumorigenic potential has attributed the importance of epigenetics in tumorigenesis, but lacked genome-wide analysis. Here we describe successful generation of iPCs from non-small cell lung cancer (NSCLC) cell lines. Following reprogramming, they resembled embryonic stem and induced pluripotent stem cells in pluripotency markers expression, gene expression patterns and *in vitro* differentiation ability. Genome-wide methylation analysis revealed that aberrantly methylated promoters which were mostly developmental-associated genes and tumor suppressors; as well as commonly upregulated genes in NSCLC i.e. *KRT19* and *S100P* were reversed in iPCs upon reprogramming. Also, the reversal of oncogenes and tumor suppressors status were partially explainable by DNA methylation. These findings suggest that DNA methylation patterns explain the downstream transcriptional effects, which potentially caused the reduced tumorigenicity in iPCs, thus providing evidence that reprogramming reverses the aberrantly dysregulated genes in NSCLC both epigenetically and transcriptionally.**

Lung neoplasm has been documented as the leading cause of death by cancer amongst men and second amongst women consistently[1]. Neoplasia is widely thought to be driven by genomic instability which is due to both reversible and irreversible alterations[2]. The epigenetic alterations primarily constitute the former while genetic mutations constitute the latter. Epigenetics which largely regulate gene expression can result in either the aberrant silencing of tumor suppressors or upregulation of oncogenes thereby contributing to tumorigenesis[3]. On the other hand, somatic gene mutations may result in additional functional consequences, apart from gene expression regulation which eventually promotes genomic instability[4]. It is currently thought that a handful of spurious mutations is both necessary and sufficient to cause malignant transformation[5]. Another form of genetic mutation involves irreversible chromosomal insults as demonstrated by the classic experiment performed by Theodor Boveri on sea urchin eggs[6], consistent with the observation that Down syndrome patients suffer from higher incidence of cancer[7]. Chromosomal damage or aneuploidy indeed affects thousands of genes and disturbs the stoichiometry of the cell, resulting in dysregulation of cell functions which eventually leads to cancer[8].

Indeed, if cancer is a genetic disease by virtue of somatic mutations, it would be of interest to elucidate whether recapitulation of the cancer phenotype is achievable with direct reprogramming[9,10] as has been shown with familial diseases[11,12]. However, several groups have recently reported that reprogrammed cancer cells are less tumorigenic compared to the parent cancer cell[13–15], suggesting that epigenetics are pertinent in cancer progression. However, a genome-wide analysis is still lacking. We therefore hypothesized that direct reprogramming may have reversed the aberrant epigenetic alterations in cancer cells which are important in cancer progression. To this end, we embarked on a study involving genome-wide analyses of DNA methylation and gene expression patterns of induced pluripotent cancer cells (iPC). Here, we describe the successful reprogramming of non-small cell lung cancer (NSCLC) i.e. H358 and H460. We found that colonies derived from H358 (iPCH358) and H460 (iPCH460) were morphologically indistinguishable from the reprogrammed IMR90 (iPSIMR90) and embryonic stem cells (ES), i.e. H1 and HES-3. iPC were pluripotent as demonstrated by presence of ES cell markers and ability to differentiate into three germ layers *in vitro*. Furthermore, the reprogrammed cells shared similar gene expression patterns with H1. In addition, key NSCLC biomarkers such as aberrantly methylated promoters (AMPs) and important prognostic factors were also reversed in iPC and this was followed through in the *in vitro* differentiated iPC (post-iPC) cells. In parallel to the downregulation of pro-angiogenic oncogenes (OGs), we also found upregulation of tumor suppressors (TSs) in the reprogrammed NSCLC lines. Our study provides

evidence that direct reprogramming amends the aberrant epigenetic signatures of cancer cells which may contribute to abrogation of their malignant properties.

## Results

**Generation of iPC from NSCLC lines.** Reprogramming of cancer cells appears to be challenging due to various genetic alterations including aberrant gene silencing in cancers which may interfere with the process. Here, we show that two NSCLC cell lines i.e. H358, an adenocarcinoma, and H460, a large cell carcinoma, were successfully reprogrammed following the forced expression of Yamanaka's four transcription factors i.e. *OCT3/4*, *KLF4*, *c-MYC* and *SOX2*[9]. As controls for the reprogramming experiment as well as reference to the NSCLC cell lines, we reprogrammed IMR90, a human fetal lung fibroblast, using the same factors and obtained iPS colonies. Our results showed that IMR90 was more readily reprogrammed compared to the two NSCLC cell lines as the former formed ES-like colonies at day-8 post-infection, before the cells were seeded onto feeder cells (Fig. 1**a**). The latter however, showed clusters of cells that were irregular but upon seeding onto feeder cells at day 15 post-infection, we found that the morphology of these colonies were identical to the colonies obtained from IMR90 as well as H1 and HES-3 (Fig. 1**a,b**).

**Morphologically and transcriptionally indistinguishable to H1.** Upon obtaining the ES-like colonies, we evaluated their pluripotency status following alkaline phosphatase (AP) staining. Our colonies derived from H358 and H460 cancer cells showed positive for AP activity and stained purple similar to iPSIMR90 and H1 (Fig. 1**b**). Immunofluorescence data further confirmed that these colonies expressed ES cell markers, as they all stained positive for TRA-1-60 and Nanog (Fig. 1**b**). Real-time PCR (qPCR) results showed that the iPCs expressed pluripotency markers such as *SOX2*, *NANOG*, *FGF4*, and *OCT3/4* at levels comparable to iPS cells (Fig. 1**c**). Moreover, all colonies depicted higher telomerase activity (TA) as compared to their respective parental cells (Fig. 1**d**). In addition, our global gene expression analysis also revealed that iPC, post-iPC (piPC), iPS and H1 were clustered together, indicating similarity in their gene expression profile (Fig. 2**a,c**). On the other hand, data from methylation array indicated that iPC, piPC and iPS have similar methylation profiles, but differed from H1 (Fig. 2**b,c**). We attribute this difference to the early passage of iPS and iPC (passage 4–8) used in this array[16,17]. Nonetheless, since gene expression is downstream of DNA methylation regulation, we postulated that the deviation of iPS and iPC from H1 in methylation profile is non-consequential in pluripotency maintenance, as the gene expression patterns showed that the reprogrammed cells did not diverge largely from H1. However, it is possible that the generated iPS and iPC have restricted commitment lineages[18]. In addition, gene ontology (GO) analysis of hypomethylated promoters in iPC compared to their respective cancer parents showed enrichment of developmental associated genes (Fig. 2**d**).

**Spontaneous *in vitro* differentiation of iPC.** In order to assess the differentiation ability of the reprogrammed cells, embryoid bodies (EBs) were formed *in vitro* and transferred to gelatin-coated plates to generate post-iPS (piPS) or piPC (Fig. 3**a**). Our data provides evidence that, like iPS, iPC cells derived from lung cancer have the ability to differentiate into the three germ layers *in vitro*. When compared to their respective parental cells, the EBs, piPS and piPC cells revealed up-regulation in ectoderm markers i.e *CDX2* and *PAX6*, genes that identify mesoderm layer i.e *Brachyury* and *MSX1*, as well as endoderm genes such as *GATA4* and *FOXA2* (Fig. 3**b**).

**Direct reprogramming hypomethylates AMPs in NSCLC.** We next, went on to investigate if reversible alterations in cancer cells were reverted upon reprogramming. To address this, we first generated a list of known AMPs in NSCLC through literature search (Table S1). We found 237 unique AMPs, out of which, 217 were interrogated by Illumina Infinium Human Methylation 27 k BeadChip array. Indeed, these genes were over-represented among all methylated promoters in H358 and H460 lung cancer cells but under-represented among all methylated promoters in IMR90 normal lung fibroblasts (Fig. 4**a**). We first categorized the promoters that are aberrantly hypermethylated in NSCLC by comparing them to IMR90. We found 105 and 94 AMPs in H358 and H460, respectively, and 84 AMPs were shared between H358 and H460 (Fig. 4**a**). Interestingly, we observed that 71 (67.6%) and 50 (53.2%) of the AMPs in H358 and H460 respectively became hypomethylated upon reprogramming (Fig. 4**b**). Among them, 44 promoters overlapped, many of which represent developmental associated genes such as *HOX* and *PAX* gene clusters, as well as tumor suppressors such as *APC*, *TIMP3* and *WRN* (Fig. 4**b**, Table S2). Methylation-specific PCR (MSP) also verified this observation. *APC* along with *HOXA5*, *HOXA7*, *HOXC9* and *HOXD13* were all hypomethylated in iPC colonies (Fig. 4**c**). Bisulfite genomic sequencing was used to evaluate and quantify the methylation level at every cytosine-guanosine dinucleotides (CpGs) contained within the amplified sequence. Results revealed that the CpGs of *HOXA5* were not methylated appreciably in iPC cells as compared to their parent cancer cells (Fig. S1**a**). We therefore showed evidence that aberrant DNA methylation in cancer was reversed by direct reprogramming.

We then asked whether the hypomethylated AMPs translate into gene upregulation. Here, we showed that 25 (35.2%) and 13 (26%) of the hypomethylated AMPs resulted in adjacent gene upregulation ($\geq$ 2-fold change, false discovery rate (FDR) $\leq$ 0.05) in iPCH358 and iPCH460, respectively (Fig. 4**d,e**). The remaining genes did not change in transcript levels. The latter observation is explainable by the fact that promoter hypomethylation does not necessarily result in gene upregulation concurrently, but has the potential to cause upregulation subsequently[19,20]. Our qPCR results concurred with gene expression array data. Transcription of Homeobox genes i.e. *HOXA5, HOXA7* and *HOXD13* were significantly upregulated in all iPC colonies when compared to respective cancer cells (Fig. 4**f**). *RPRM*, a gene known to be heavily methylated in lung cancer and its low expression correlated with poor prognosis[21] was also restored upon reprogramming, consistent with the hypomethylation observed (Fig. 4**c,f**).

**Downregulation of NSCLC biomarkers upon reprogramming.** We then assessed if the genes commonly upregulated in NSCLC (UR), would be downregulated upon reprogramming. Similar to AMPs, we generated a list of URs based on literature and published gene expression data of NSCLC clinical samples deposited in GEO database (GSE19188)[22] (Table S3). In our list of 420 unique genes, 391 genes were interrogated in the Illumina HumanHT-12 array. From this master list, we identified 110 and 59 genes to be upregulated ($\geq$ 2-fold change, FDR $\leq$ 0.05) in H358 and H460, respectively, when compared to IMR90 and we define these genes as URs in our cancer samples. Among these, 52 (47.3%) and 25 (42.4%) were downregulated in iPCH358 and iPCH460, respectively, and were over-represented for genes downregulated upon reprogramming (Fig. 5**a**, Table S4).

The URs that were downregulated in H358 and H460 upon reprogramming include important prognostic factors such as *KRT19* (also known as *CK19* or *CYFRA 21-1*), *S100P*, *KRT7*, *PPAP2C* and *AGR2*[23–27]. Our validation with qPCR concurred with the gene expression array data (Fig. 5**b**). Interestingly, URs that were downregulated upon reprogramming can be explained by DNA hypermethylation in iPCs (Fig. 5**c**). This was further confirmed for *S100P* and *KRT19* genes in which MSP and qPCR results showed complete methylation and gene silencing respectively, in all iPC colonies as compared to the parental cancer cells (Fig. 5**c,d**). The

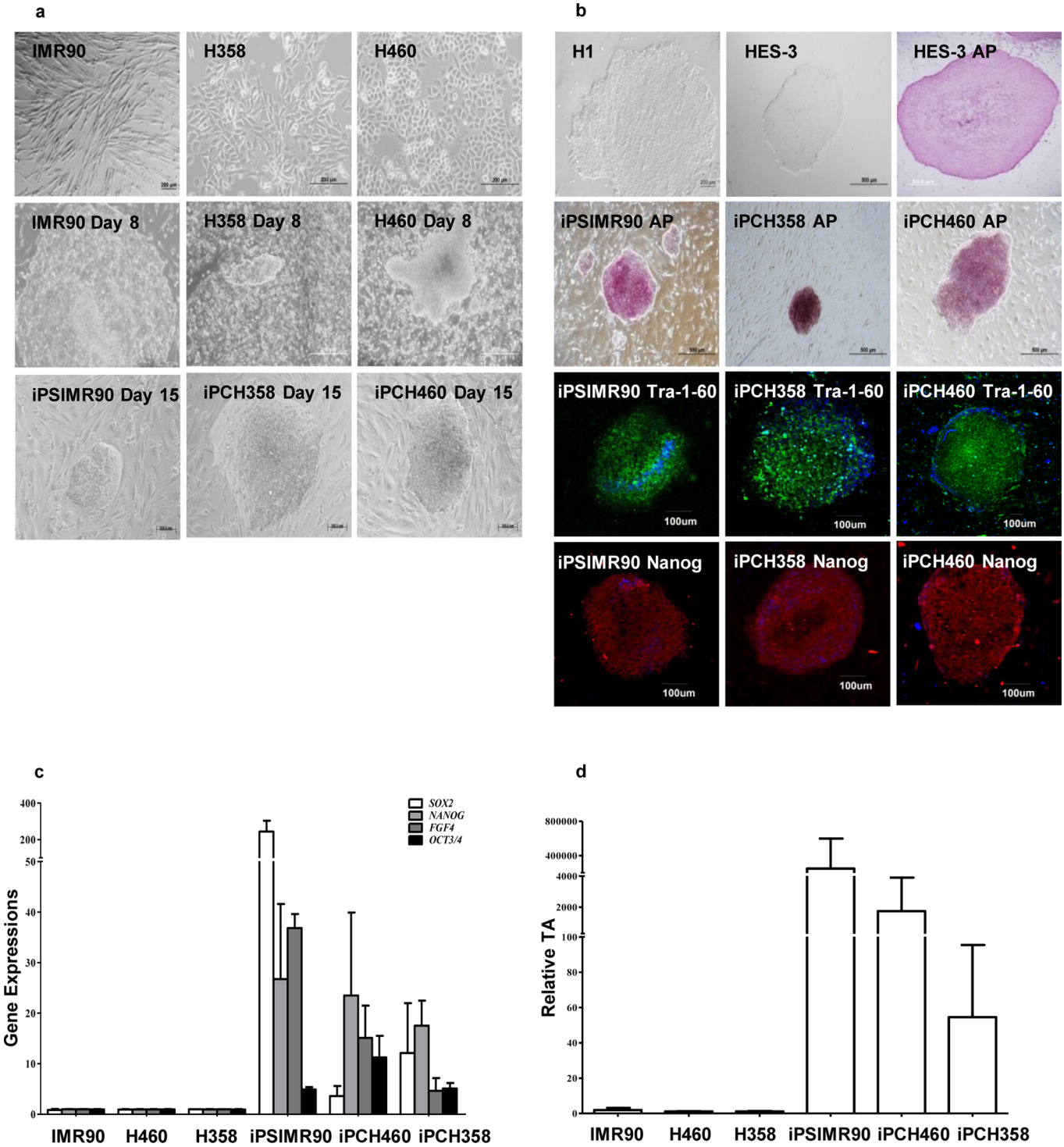



**Figure 1 | Generation and characterization of iPC from NSCLC lines.** (a) Morphology of parental IMR90 fibroblasts, H358 and H460 cancer cells. Formation of ES-like colonies in IMR90, H358 and H460 cells on day 8 post-infection of four factors. Upon seeding onto feeder layer on day 15 post-infection, iPS and iPCs formed flat and round edged colonies like (b) ES cells i.e. H1 and HES-3. Similar to H1, iPSIMR90, iPCH358 and iPCH460 colonies stained positive for AP. Representative images of individual colonies derived from IMR90, H358 and H460 showed positive staining for pluripotency markers i.e. TRA-1-60 (green) and Nanog (red) respectively. Nuclei were stained with Hoechst 33342 (blue). Scale bars: 20 - 500μm. (c) iPS and iPC expressed high levels of ES cell markers namely *SOX2*, *NANOG*, *FGF4* and *OCT3/4* as compared to their parental cells. The mRNA expression was normalized to *GAPDH* mRNA expression. (d) iPS and iPC depicted higher TA when compared to their respective parental cells. Data are presented as mean ± SD.

methylation status in iPC cells was further quantified using bisulfite sequencing. H358iPC and H460iPC cells were highly methylated in *KRT19* gene with methylation scores 86% and 96%, respectively (Fig. S1**b**). This suggests that epigenetics, largely DNA methylation, played a role in the dysregulation of URs to cause malignant transformation. To rule out any possibility that the remaining URs maintained a high transcript level in iPCs, we therefore checked whether these transcript levels were comparable to H1. To our surprise, we

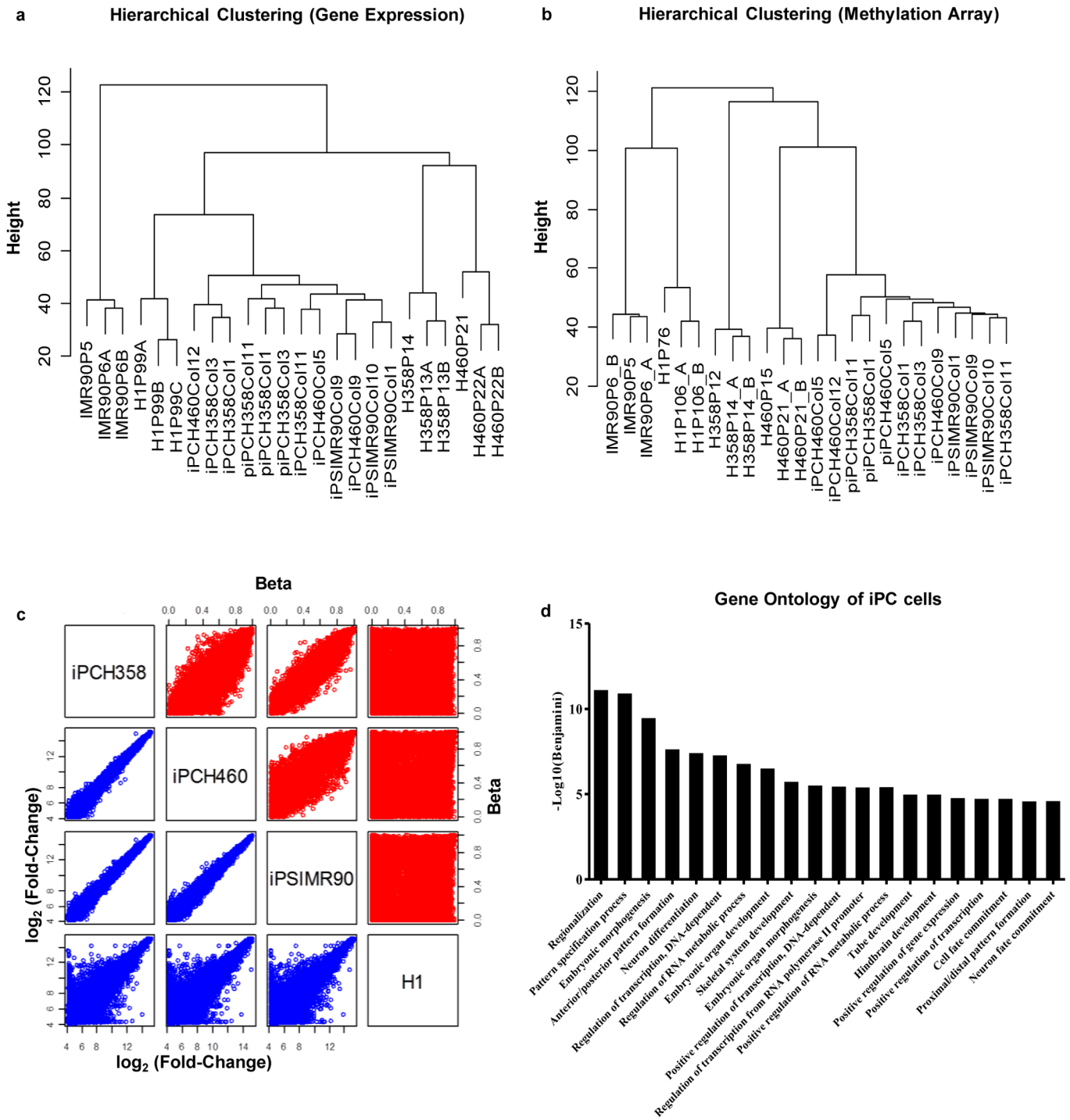



**Figure 2 | Genome-wide assessment of gene expression and CpG methylation profiles of iPC derived from lung cancer.** (a) Hierarchical clustering showing gene expression patterns of the parents, iPCs, post-iPCH358 (piPCH358), iPS and H1. (b) Clustering of samples based on methylation levels of the parents, iPCs, iPS, piPCH358, post-iPCH460 (piPCH460) and H1. (c) Paired scatter plot illustrating the similarities of gene expression levels between iPS/Cs and H1 (Blue). Methylation levels, on the other hand, were only similar between iPS and iPC but differed from H1 (Red). (d) GO analysis of hypomethylated promoters in iPC compared to cancer parents showing enrichment in developmental associated genes.

did not find any evidence to reject the null hypothesis. Therefore, it is likely that these URs were required at high expression levels to maintain the iPCs in an ES-like state. Nonetheless, we have shown that reprogramming indeed reverses the aberrantly upregulated genes in NSCLC both epigenetically as well as transcriptionally.

**Oncogenes and tumor suppressors.** In order to compare the gene regulation of tumor suppressors (TS) and oncogenes (OG) in our experiment, firstly, we obtained a list of TS and OG from the Memorial Sloan-Kettering Cancer Center Database (cbio.mskcc.org/CancerGenes; accessed on 5 Dec 2011)[28] and determined the aberrant regulation of these genes in H358 and H460 as compared to IMR90 cells.

Among the 495 OG obtained from the database, we identified 42 and 29 genes that were aberrantly upregulated in H358 and H460, respectively. Among these, 25 (59.5%) and 14 (48.3%) were

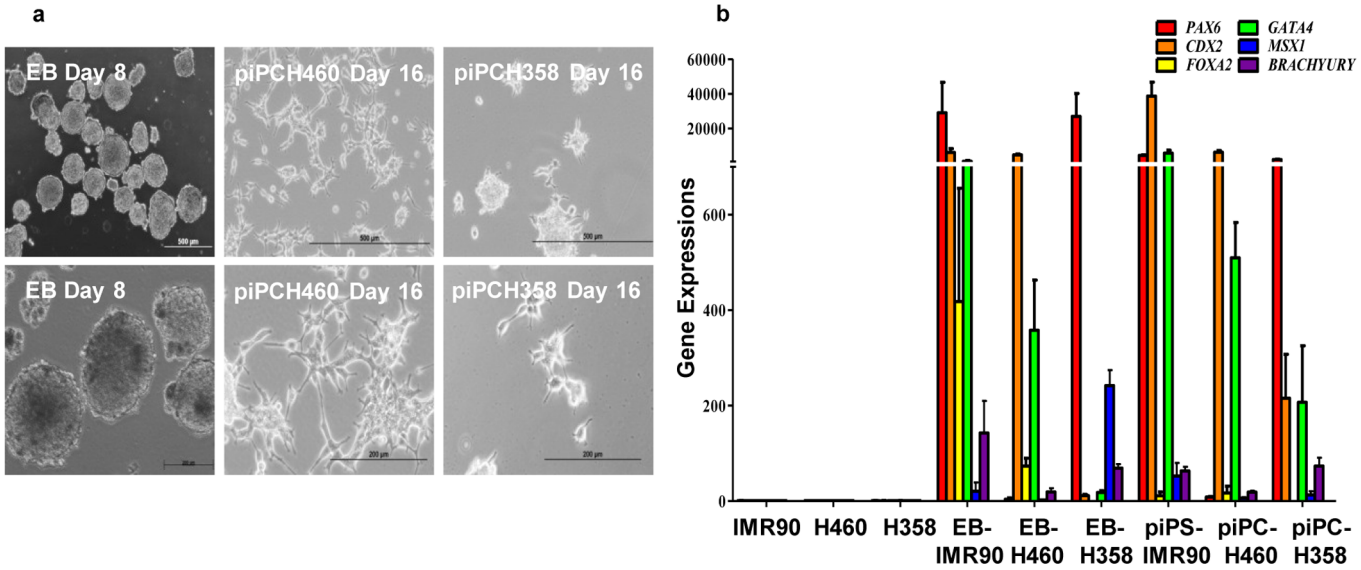



**Figure 3** | ***In vitro* differentiation of iPC cells through EB formation.** (a) Generation of EBs and piPCs cells. EBs were maintained as floating culture for 8 days before returning to adherent culture for another 8 days. Scale bars, 200 μm or 500 μm. (b) qPCR results demonstrated that EBs and piPC from H358 and H460 expressed markers of the three embryonic germ layer i.e. *CDX2* and *PAX6*, (ectoderm), *Brachyury* and *MSX1* (mesoderm), and *GATA4* and *FOXA2* (endoderm). The mRNA expression was normalized to *GAPDH* mRNA expression. Data are presented as mean ± SD.

downregulated upon reprogramming (Fig. 6**a**). We found that the gene expression patterns in H358 and iPCH358 were satisfactorily explained by DNA methylation, whereby DNA hypomethylation in H358 attributed to OG aberrant upregulation and DNA hypermethylation in iPCH358 silenced these aberrantly upregulated genes (Fig. 6**b**, Table S5). However, this trend is not statistically significant for H460 and iPCH460 (Fig. 6**b**, Table S5) and we speculate that other gene regulation mechanisms were utilized in these cells. Nonetheless, the top three oncogenes which were upregulated in both the cancer cell lines i.e. *EFNA1, CXCL1* and *CXCL2* were pro-angiogenic factors[29–31] which became downregulated upon reprogramming in all iPC and piPC cells (Table S5). *ID1*, an oncogene that promotes lung cancer cell proliferation[32] was also observed to be downregulated in the reprogrammed NSCLC line.

As for TS, we obtained 873 genes from the database. From this list, approximately 87 and 74 TS were aberrantly downregulated in H358 and H460, respectively, when compared to IMR90. We found that 21 and 6 of these TS were significantly upregulated upon reprogramming (Fig. 6**c**). Of these, TSs *CADM1*[33] and *PLAGL1*[34] were transcriptionally elevated in both reprogrammed H358 and H460 cell lines as compared to their respective parental cells. Nonetheless, we noted that the total percentage of TS upregulation were low and found that the large, remaining bulk of these genes also have comparably low expression levels in H1 (Table S6), suggesting that TS genes possibly need to be maintained at low levels for cell proliferation and survival[35]. Following this, we investigated whether DNA methylation could explain the regulation of these TS. We observed that the promoters of TS were significantly hypermethylated in H358 and hypomethylated in iPCH358, but the same observation was not detected in H460 and iPCH460 (Fig. 6**d**). We concluded that the dysregulation of oncogenes (OGs) and tumor suppressors (TSs) in NSCLC were reversed upon reprogramming and were partially explainable by intricate DNA methylation regulation.

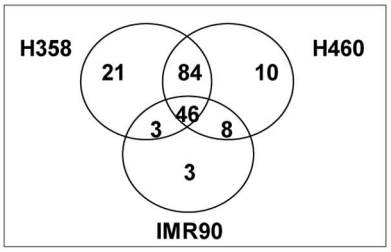
## Discussion

The reprogramming of cancer cells have been reported in mice melanoma[15,36], human melanoma, prostate cancer[37], chronic myeloid leukemia[38], a panel of gastrointestinal cancer cells[14], lung[39] and breast cancer cells[13]. Here we describe the successful reprogramming of NSCLC cell lines, namely H358 and H460. The evidence for pluripotency in iPCH358 and iPCH460 were manifested through AP staining, pluripotency markers expression and *in vitro* differentiation assay by EBs formation which exhibited the presence of markers that identify the three germ layers. Moreover, we also present the first extensive characterization of the methylome and transcriptome of iPS and iPC through DNA microarray technologies. Although our methylation data may be affected by epigenetic memory of early passages of iPS and iPC, this did not affect its downstream regulation of the transcriptomes. Indeed, the transcriptome of iPS, iPC and piPC were indistinguishable from each other and were closely related to ES cells. This observation is indeed novel and reveals that cancer cells can be reprogrammed to attain ES-like characteristics. Motivated by this, we assessed the reversible changes that account for tumorigenesis such as aberrant hypermethylation of promoters as well as abnormal upregulation of genes in NSCLC. Furthermore, we also investigated the fate of oncogenes and tumor suppressors in our cancer cell lines upon reprogramming.

Previously, Ron-Bigger *et al.* (2010) reported that reprogramming could reverse hypermethylated promoters, particularly tumor suppressor gene *p16* in hTERT immortalized human lung fibroblast (WI-38), however the established subclone of hTERT cells may not be entirely similar to cancer cells[40]. In our study, we have reprogrammed established cancer cell lines to satisfactorily address the question whether direct reprogramming may reverse AMPs in cancer. Considering that DNA hypermethylation is associated with silencing of gene transcription[19], these aberrantly hypermethylated promoters, which are largely enriched for differentiation-associated genes in lung cancer[41], possibly confer growth advantages to cancer cells. Our study showed that direct reprogramming were able to perturb the epigenetics of lung cancer cells by causing the reversal of these AMPs and in some instances, resulted in active gene transcription. However, it is interesting to note that despite using early passages of iPCs to assess this question, which has been shown to cause interference due to epigenetic memory from the parental cells, we were still successful in proving this point[16,17]. Indeed, it is very plausible that our data is underestimated. It was reported previously that genes expressed in fully differentiated lung cells are repressed in lung cancer and *vice versa*[42,43] and was explainable by DNA methylation[41]. It was

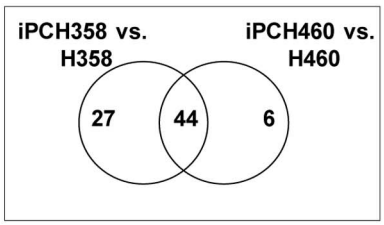
a

| Samples | AMP | Non-AMP | Total | Odds Ratio | P values |
|---|---|---|---|---|---|
| H358 | 154 | 6833 | 6987 | 2.66 | 6.40E-12 |
| H460 | 148 | 6850 | 6998 | 2.33 | 1.80E-09 |
| IMR90 | 60 | 6354 | 6414 | 0.48 | 2.80E-07 |

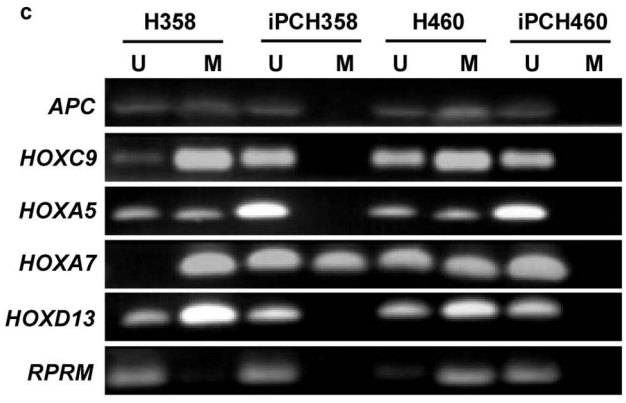



b

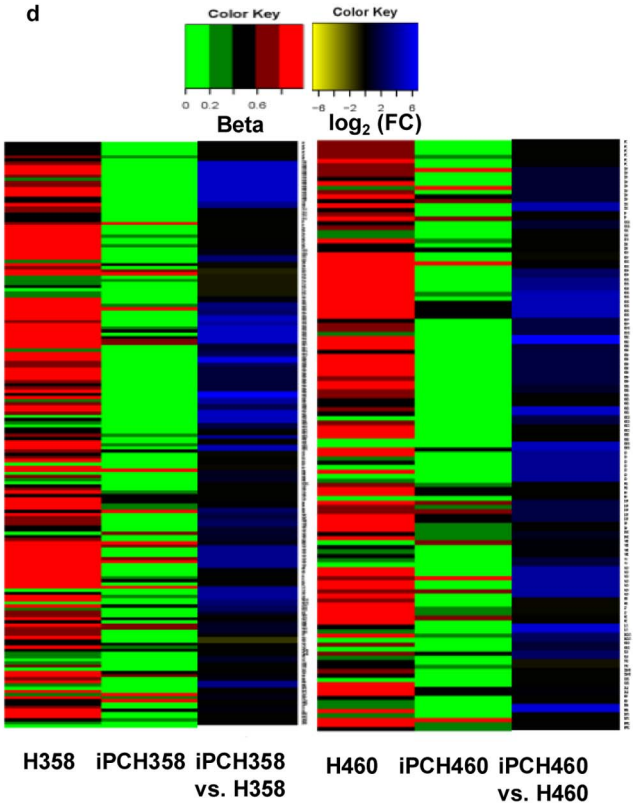
| Samples | AMP | Non-AMP | Total | Odds Ratio | P values |
|---|---|---|---|---|---|
| iPCH358 vs. H358 | 71 | 3072 | 3143 | 2.59 | 2.00E-06 |
| iPCH460 vs. H460 | 50 | 1745 | 1795 | 3.36 | 9.10E-09 |




c




d




e

| Samples | No. of Genes (%) | |
|---|---|---|
| | iPCH358 vs. H358 | iPCH460 vs. H460 |
| Upregulated | 25 (35) | 13 (26) |
| Downregulated | 0 (0) | 0 (0) |
| Status quo | 46 (65) | 37 (74) |
| **Total** | **71 (100)** | **50 (100)** |

f

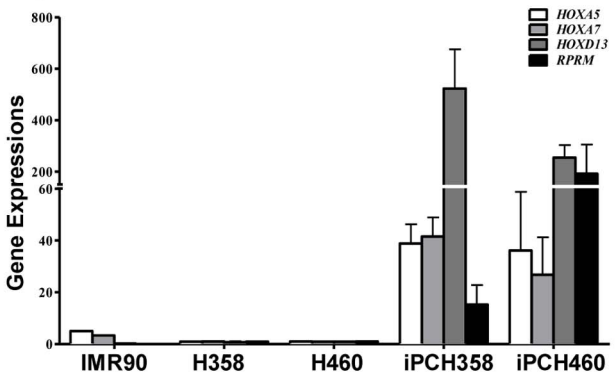
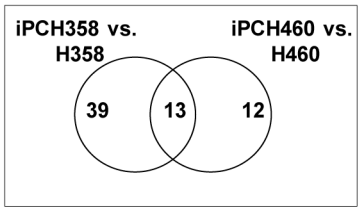


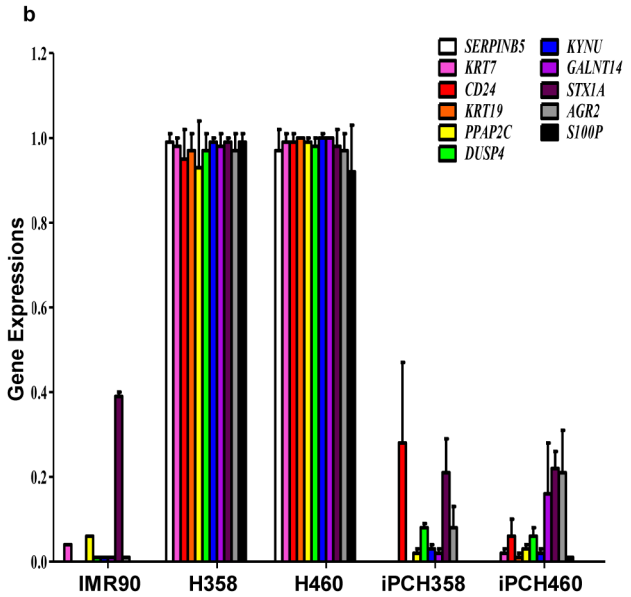

**Figure 4 | Fate of AMPs upon direct reprogramming of lung cancer cells.** (a) Table and Venn diagram showing the number of identified AMPs methylated in H358, H460 and IMR90. P-value and odds ratio (OR) was calculated using Fisher's exact test against the expected background; OR>1: over-representation, OR<1: under-representation. (b) Table and Venn diagram illustrating the number of identified AMPs that are hypomethylated in iPCH358 and iPCH460. Fisher's exact test was used to calculate P-value and OR. (c) MSP analyses further verified the methylation status of AMP genes extracted from the array. (d) Heat map representing all hypomethylated AMPs identified in (b). Methylation pattern is depicted in green (unmethylated) and red (methylated). Adjacently, the gene expression patterns showed here in yellow (downregulated) and blue (upregulated) illustrate gene upregulation in iPCs upon hypomethylation. **(e)** Quantification of the number of genes upregulated, downregulated or status quo among all hypomethylated AMPs in iPCH358 and iPCH460. (f) qPCR on *HOX* gene clusters and other AMPs corroborated with the gene expression data. The mRNA expression was normalized to *GAPDH* mRNA expression. Data are presented as mean ± SD.

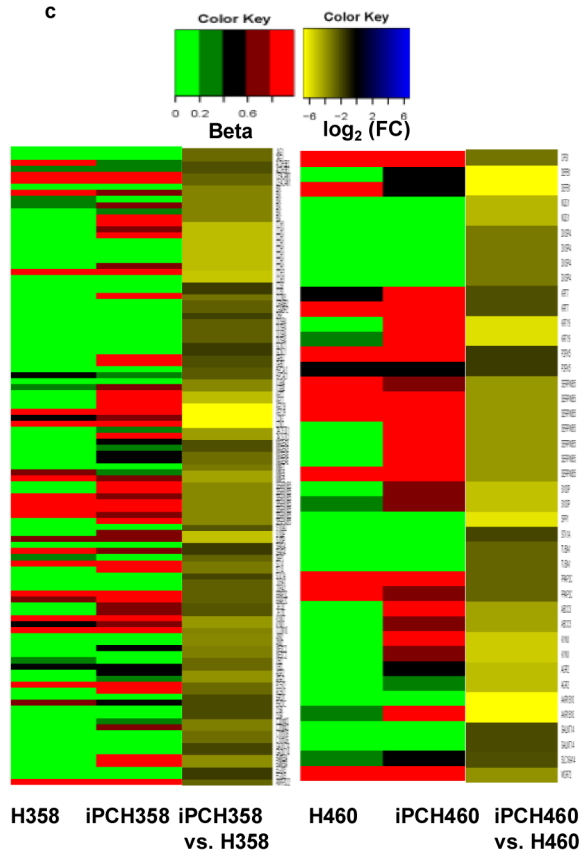
suggested that this observation possibly implies that lung cancer development is related to a dedifferentiation event and exactly how it confers growth advantage to result in malignancy remains unclear. Nonetheless, our study shows that following reprogramming, the iPCs no longer harbor the same aberrant DNA methylation mark and may no longer exhibit malignancy.

a

| Samples | UR | Non-UR | Total | Odds Ratio | P values |
|---|---|---|---|---|---|
| iPCH358 vs. H358 | 52 | 883 | 935 | 29.94 | <2.20E-16 |
| iPCH460 vs. H460 | 25 | 712 | 737 | 30.68 | <2.20E-16 |



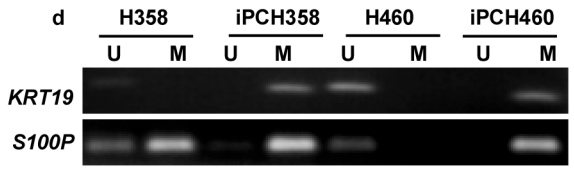




**Figure 5 | Aberrantly upregulated URs were reversed upon direct reprogramming of NSCLC cell lines.** (a) Identified URs that were downregulated in iPCs were over-represented in all downregulated genes in iPC compared to parents. Fisher's exact test was conducted to determine the OR (>1:over-representation) and P-value. Number of genes shared between iPCH358 and iPCH460 were illustrated in the Venn diagram. (b) qPCR showing downregulation of UR genes in iPCH358 and iPCH460 that were initially upregulated in the parental cancer cells. The mRNA expression was normalized to *GAPDH* mRNA expression. Data are presented as mean ± SD. (c) Heat map illustrating the downregulation of URs in iPCs as a consequence to DNA hypermethylation. (d) MSP analyses on *KRT19 and S100P* showing consistent data with the methylation array.

Another reversible alteration we observed in NSCLC was the commonly upregulated genes. Since the advent of DNA microarray technologies to interrogate mRNA transcript levels, assays have been performed to identify biomarkers to predict lung cancer survival in patients[22,44,45]. By compiling a list of genes that are commonly upregulated in NSCLC compared to normal adjacent tissue, we report that these markers that are found to be aberrantly upregulated in H358 or H460 were subsequently downregulated upon reprogramming. And this list includes important prognostic factors of lung cancer such as *KRT19* and *S100P*[23,24]. Moreover, our array data revealed that the regulation of these genes can be satisfactorily explained by DNA methylation. Therefore, it is interesting to note that these markers are made absent upon induction of pluripotency. Again, supposing that these prognostic factors are pertinent in cancer progression[46,47], we find that direct reprogramming may result in loss of malignancy. We, therefore present evidence here that epigenetic changes were significant in explaining the regulation of these genes and hence the importance of epigenetics in NSCLC progression. Indeed, the observed results are fascinating in that the prognostic factors as well as DNA methylation markers that are crucial for NSCLC progression seem to be reversed upon direct reprogramming. However, this observation does not discount the possibility of reacquiring these phenotypes upon directed differentiation to lung lineage cells. Unfortunately to date, there are no established protocols for directed differentiation to lung lineage cells. Nonetheless, we assessed whether these cancer markers manifested in the *in vitro* differentiated piPC cells. To our surprise, we did not find any aberrant dysregulation of these genes as well as DNA methylation markers and we conclude that direct reprogramming of cancer cells resulted in the reversion to normal DNA methylation and gene expression regulation (Table S4).

We further analyzed the effect of direct reprogramming on a panel of TSs and OGs. Surprisingly, we found that these genes i.e. pro-angiogenic factors such as *EFNA1, CXCL1* and *CXCL2* as well as *ID1* which work in concert to promote tumorigenesis were reversed to the normal expression levels in iPC and remained so in piPC (Table S5,S6). On the other hand, TSs such as *CADM1* and *PLAGL1* were upregulated in the NSCLC lines upon reprogramming. Interestingly, we found that the regulation of these genes in H358 were prominently explainable by DNA methylation but not in H460. This reveals

a

| Samples | OG | Non-OG | Total | Odds Ratio | P values |
|---|---|---|---|---|---|
| iPCH358 vs. H358 | 25 | 910 | 935 | 47.79 | <2.20E-16 |
| iPCH460 vs. H460 | 14 | 723 | 737 | 38.34 | <2.20E-16 |

c

| Samples | TS | Non-TS | Total | Odds Ratio | P values |
|---|---|---|---|---|---|
| iPCH358 vs. H358 | 21 | 690 | 711 | 13.7 | 7.00E-16 |
| iPCH460 vs. H460 | 6 | 424 | 430 | 6.43 | 0.0005 |

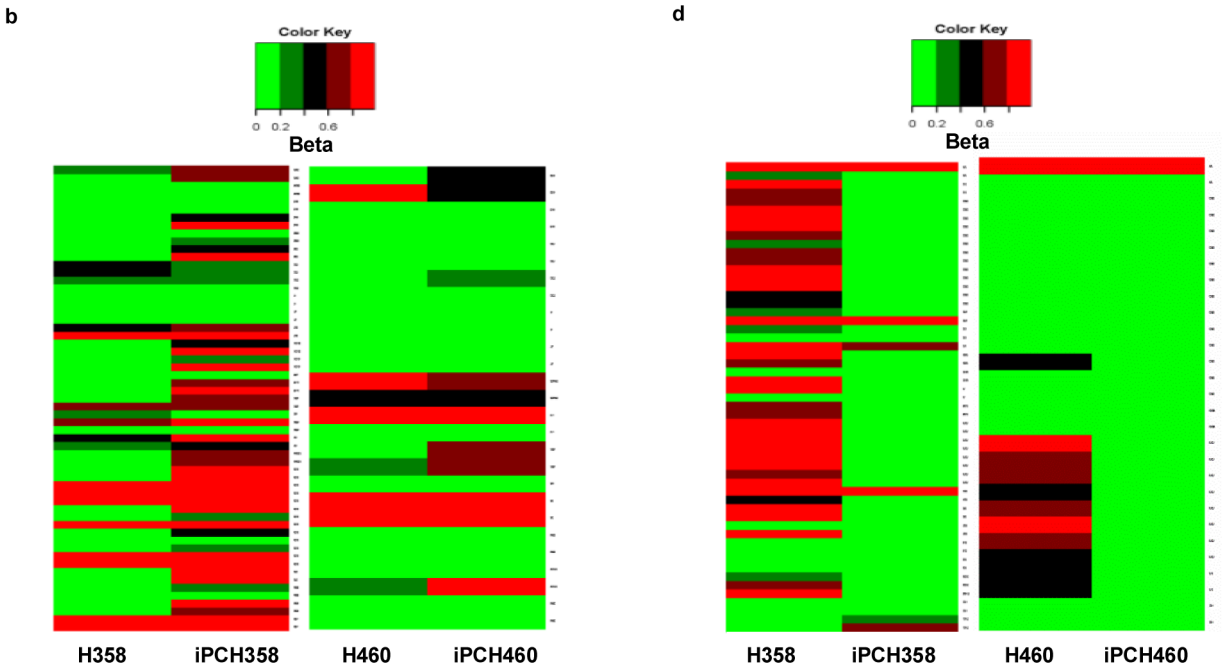



**Figure 6 | The fate of oncogenes (OG) and tumor suppressors (TS) following direct reprogramming of lung cancer cells.** (a) Identified OG in parental cancer cells were downregulated upon direct reprogramming and were over-represented among all downregulated genes (OR>1:over-representation). (b) Heat map illustrating DNA methylation controls the gene expression changes of OG in H358 upon direct reprogramming, but not in H460. (c) Similarly, identified TS were upregulated in cancer iPCs upon direct reprogramming. (d) Heat map illustrating hypomethylation of promoters in iPCH358 on adjacent upregulated TS, but not so in iPCH460.

that the mechanism behind aberrant dysregulation of tumor suppressors and oncogenes are more robust and may include other mechanisms such as gene deletions and gene amplifications, to name a few.

Our extensive study has revealed a better understanding of cancer. In the wisdom of Thomas S. Kuhn, in his landmark thesis on paradigms[48], direct reprogramming is indeed a new tool to study and understand cancer cells, which may result in paradigm shifts. By globally resetting the epigenetic state of lung cancer cells through direct reprogramming, our study provides evidence that these cells may become reticent by reversing aberrant epigenetic changes in NSCLC which in turn affects the gene regulation. However, it is of interest to study whether the directed differentiation of these iPCs to different commitment lineage will result in malignant manifestation phenotypically as well as epigenetically, although, our *in vitro* differentiation assay suggests that this may not be true.

Reversal of aberrant cancer methylome, which explains regulation of prognostic factors as well, by direct reprogramming provides evidence that DNA methylation is important for tumorigenesis. By extension, targeting epigenetics factors to inhibit tumor growth as shown clinically[49], is the way forward. Epigenetic based therapy by itself or in combination with current available drugs for NSCLC may be an improved and better therapeutic regimen for lung cancer patients in the near future. However, development of drugs with higher degree of precision and targeting will be desired. It would be of interest to elucidate the indirect roles of Yamanaka's four factors in the delicate regulation of epigenetics in a cell, i.e. hypomethylation or hypermethylation at specific loci. Better understanding of this mechanism would certainly contribute to a more sophisticated and effective treatment of cancer than currently tested non-specific DNA methylation inhibitors or DNA demethylating agents.

## Methods

**Cell lines and culture conditions.** Cell lines used in our study include human embryonic lung fibroblasts IMR90 (ATCC no. CCL-186), NSCLC lines i.e. adenocarcinoma NCI-H358 (ATCC no. CRL-5807) and large cell carcinoma NCI-H460 (ATCC no. HTB-177), as well as ES cells i.e. H1 and HES-3. All cell lines were maintained as recommended by the American Type Culture Collection (ATCC). iPS, iPC and ES cells on the other hand, were cultured and expanded on irradiated mouse embryonic fibroblasts (iMEFs) in medium consisting of DMEM/F12 (Invitrogen), 20% Knockout Serum Replacement (Invitrogen), 1 mM L-glutamine (Invitrogen), 100 μM nonessential amino acids, 100 μM β-mercaptoethanol (Sigma-Aldrich), and supplemented with 4 ng/uL basic fibroblast growth factor (bFGF) (Invitrogen). Pluripotent cells on Matrigel (BD Biosciences) were maintained using mTeS®1 (STEMCELL Technologies) medium.

**Transfection and infection.** Human iPC and iPS cells were established using Yamanaka's protocol with slight modification[9]. Infectious lentiviral particles were produced by transfecting 293T cells with pLenti6/UbC/mSlc7a1 (Addgene plasmid 17224) using Lipofectamine. Retroviral vectors (pMXs) used in this study include of pMXs-hOCT3/4 (Addgene plasmid 17217), pMXs-hSOX2 (Addgene plasmid 17218), pMXs-hKLF4 (Addgene plasmid 17219) and pMXs-hC-MYC (Addgene plasmid 17220). The viral transfectants were collected, mixed and filtered before infecting human fibroblasts and cancer cells. Individual ES-like colonies were picked and passaged on six-well plates on iMEFs and maintained in hES medium.

**RNA isolation, reverse transcription and qPCR.** Total RNA was extracted by using RNeasy Mini Kit (Qiagen) and reverse-transcribed using Reverse Transcriptase Enzyme (Promega) as well as oligo dT primers (Promega), according to manufacturer's instruction. The synthesized cDNA was diluted and SYBR Green

qPCR Master Mix (Applied Biosystems) was used for the detection. Relative expression was calculated using the $\Delta\Delta C_t$ method. All samples were performed in duplicates and error bars represent standard deviation of the relative values. *GAPDH* housekeeping gene was set as the reference. Primers used in qPCR are shown in Table S7.

**Alkaline phosphatase (AP) staining.** AP staining was performed using the Alkaline Phosphatase Detection Kit (Millipore), in accordance with the manufacturer's instruction.

**Telomerase activity (TA) assay.** TA was quantified using TeloExpress Quantitative Telomerase Detection Kit (ExpressBio) which utilizes the telomerase repeat amplification protocol (TRAP)-based qPCR method, according to the manufacturer's instruction. TA in each sample was calculated based on comparing the $C_t$ values of the standard curve generated from 10-fold dilutions of telomerase control oligo with known copy numbers of the telomeric repeats.

***In vitro* differentiation.** piPS and piPC were derived from *in vitro* spontaneous differentiation assay through EBs formation as described by Miyoshi and colleagues[14]. Briefly, iPS and iPC cells were differentiated into EBs following the floating culture technique in ultra-low attachment plates (Corning) in the absence of bFGF for 8 days. EBs were collected at day 8 and analyzed for markers of the three embryonic germ layers. After 8 days of floating culture, EBs were then transferred to a gelatin-coated plate and cultured in the same medium for another 8 days to allow the attachment of the EBs.

**Gene expression profiling.** The quality of total RNA of our samples was evaluated using Agilent 2100 Bioanalyzer (Agilent Technologies). For gene expression profiling, biotinylated antisense RNA of each sample was amplified using Illumina TotalPrep RNA Amplification Kit (Ambion), and hybridized onto HumanHT-12 v4 Expression BeadChip (Illumina) as per manufacturer's instruction. Raw data (with background subtraction) was exported using Bead Studio (Illumina) and analyzed using LUMI[50] and LIMMA[51], which were executed in R statistical software[52]. The microarray data was normalized using the quantile method[53]. All spots that were considered background were coerced to the maxima of background signal. The processed data was then used to generate the hierarchical clustering of samples as well as pair-wise comparison using 2-sample t-test and inference made through empirical Bayesian. Genes that were differentially regulated must fulfill two criteria: $\geq$2-fold change and FDR-adjusted P-value (FDR) $\leq$0.05.

**Genome-wide DNA methylation profiling.** Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen), and bisulfite-treated using the EZ DNA Methylation Kit (Zymo Research), as per the instructions in the kit. Bisulfite-treated DNA was amplified and hybridized to the Infinium HumanMethylation27 BeadChip (Illumina), in accordance with the manufacturer's recommendations. The chips were scanned and the raw intensities for both methylated and unmethylated DNA were exported using Genome Studio (Illumina). The microarray data was analyzed using methyLUMI[50] and LIMMA[51] executed on R software[52]. The two-color channel data was preprocessed separately. The intensity signal was background subtracted before normalizing by the quantile method[53]. This is followed by coercing all negative values to 0. The processed data were used to generate hierarchical clustering as well as the M-value and β-value. Each interrogated locus was probed by a methylated-specific probe (Met) and an unmethylated-specific probe (Unmet). We represent the degree of methylation using the M-value and β-value. The M-value=$\log_2$(Met/Unmet) while the β-value=Met/(Unmet+Met)[54]. M-value was utilized for statistical testing while β-value was used for a more intuitive interpretation of the degree of methylation. We categorized β-value as methylated if β-value$\in$(0.8, 1.0], partially methylated if β-value$\in$[0.2, 0.8] and unmethylated if β-value$\in$[0, 0.2). We declare a promoter was methylated if at least half the probes have β-value$\geq$0.2; a promoter was claimed hypermethylated if at least half the probes were in higher methylation category compared to the reference (i.e. from unmethylated to partially methylated or methylated) and if FDR$\leq$0.05 (calculated using M-values). Hypomethylated was the reverse. All hypomethylated promoters in iPC compared to cancer parents were analyzed using Database for Annotation, Visualization and Integrated Discovery (DAVID)[55,56] by 'Gene Ontology biological process' enrichment.

**Gene set analysis.** Four sets of genes were generated i.e. aberrantly methylated promoters (AMP), commonly upregulated genes in NSCLC (UR), oncogenes (OG) and tumor suppressors (TS). AMP was generated by literature search. UR was generated by literature search as well as using a data set from GSE19188 (taken from GEO database)[22]. OG and TS were obtained from the Memorial Sloan-Kettering Cancer Center[28].

**Methylation-specific PCR assay.** 1 μL bisulfite-treated DNA was amplified using the methylation-specific PCR primers designed using MethPrimer. All primer sequences were supplied in Table S7. For each PCR reaction, a total of 25 μL reaction volume contained 1 μM forward and reverse primers, as well as HotStarTaq Plus Master Mix (Qiagen). PCR products were resolved by agarose gel electrophoresis and stained with ethidium bromide. The bands were scored as methylated or unmethylated according to the presence or absence of a PCR product respectively.

**Immunofluorescence staining.** For immunofluorescence assay, cells were first fixed with 4% paraformaldehyde for 15 minutes followed by permeabilization step for 10 minutes with 0.5% NP-40 in PBS. The cells were then blocked with 5% BSA for one hour and incubated with primary antibodies for overnight, and later followed with fluorescence-conjugated secondary antibodies for an hour. Primary antibodies used in this assay include TRA-1-60 (Cell Signaling) and Nanog (Cell Signaling). All images were captured using Olympus Fluoview FV1000 microscope and one representative image of at least three repeats/triplicates is shown in results section.

**Bisulfite sequencing.** 3 μL bisulfite-treated DNA was amplified and for each PCR reaction, 0.5 μM forward and reverse primers, HotStarTaq Plus Master Mix (Qiagen) were used in a 50-μL total reaction volume. The PCR cycles were as follows: initial denaturation at 95°C for 10 minutes; 37 cycles of 30 seconds at 94°C, 30 seconds at 55°C and 60 seconds at 72°C; followed by a final extension for 10 minutes at 72°C. The PCR products were purified with QIAquick column (Qiagen) before subcloning into pGEM-T vector (Promega). The ligation reactions were transformed into JM109 competent cells (Promega), as described in kit procedures, and blue/white screening was used to randomly select a minimum of ten bacterial clones. Plasmid DNA was then isolated from each clone using QIAprep Miniprep Kit (Qiagen). Clones were further screened by amplifying with universal primers (T7 and SP6) and were resolved by agarose gel electrophoresis to verify insert and plasmid size. Four clones of each sample were verified by sequencing with the T7 promoter universal primers.

**Accession number.** All microarray gene expression and methylation data have been deposited at NCBI GEO database under accession number GSE35913.


1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J Clin* **61**, 69–90 (2011).
2. Duesberg, P., Fabarius, A. & Hehlmann, R. Aneuploidy, the primary cause of the multilateral genomic instability of neoplastic and preneoplastic cells. *IUBMB Life* **56**, 65–81 (2004).
3. Esteller, M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* **16 Spec No 1**, R50–59 (2007).
4. Adjei, A. A. Blocking oncogenic Ras signaling for cancer therapy. *J Natl Cancer Inst* **93**, 1062–1074 (2001).
5. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
6. Boveri, T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J Cell Sci* **121 Suppl 1**, 1–84 (2008).
7. Natarajan, A. T. *DNA repair and human disease* 61 (Landes Bioscience 2006).
8. Duesberg, P., Mandrioli, D., McCormack, A. & Nicholson, J. M. Is carcinogenesis a form of speciation? *Cell Cycle* **10**, 2100–2114 (2011).
9. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
10. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
11. Colman, A. & Dreesen, O. Pluripotent stem cells and disease modeling. *Cell Stem Cell* **5**, 244–247 (2009).
12. Devine, M. J. *et al.* Parkinson's disease induced pluripotent stem cells with triplication of the alpha-synuclein locus. *Nat Commun* **2**, 440 (2011).
13. Allegrucci, C. *et al.* Epigenetic reprogramming of breast cancer cells with oocyte extracts. *Mol Cancer* **10**, 7 (2011).
14. Miyoshi, N. *et al.* Defined factors induce reprogramming of gastrointestinal cancer cells. *Proc Natl Acad Sci U S A* **107**, 40–45 (2010).
15. Utikal, J., Maherali, N., Kulalert, W. & Hochedlinger, K. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci* **122**, 3502–3510 (2009).
16. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28**, 848–855 (2010).
17. Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
18. Bock, C. *et al.* Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).
19. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010–1022 (2011).
20. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**, 755–766 (2008).
21. Suzuki, M. *et al.* Aberrant methylation of Reprimo in lung cancer. *Lung Cancer* **47**, 309–314 (2005).
22. Hou, J. *et al.* Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One* **5**, e10312 (2010).
23. Hirashima, T. *et al.* Prognostic significance of CYFRA 21-1 in non-small cell lung cancer. *Anticancer Res* **18**, 4713–4716 (1998).
24. Arumugam, T. & Logsdon, C. D. S100P: a novel therapeutic target for cancer. *Amino Acids* **41**, 893–899 (2011).
25. Flanagan, J. M. *et al.* Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADARB1 as putative anticancer drug targets. *Mol Cancer Ther* **8**, 249–260 (2009).
26. Wang, K. K. *et al.* Novel candidate tumor marker genes for lung adenocarcinoma. *Oncogene* **21**, 7598–7604 (2002).

27. Camilo, R., Capelozzi, V. L., Siqueira, S. A. & Del Carlo Bernardi, F. Expression of p63, keratin 5/6, keratin 7, and surfactant-A in non-small cell lung carcinomas. *Hum Pathol* **37**, 542–546 (2006).
28. Higgins, M. E., Claremont, M., Major, J. E., Sander, C. & Lash, A. E. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res* **35**, D721–726 (2007).
29. Nojiri, K. *et al.* The proangiogenic factor ephrin-A1 is up-regulated in radioresistant murine tumor by irradiation. *Exp Biol Med (Maywood)* **234**, 112–122 (2009).
30. Hatfield, K. J., Bedringsaas, S. L., Ryningen, A., Gjertsen, B. T. & Bruserud, O. Hypoxia increases HIF-1alpha expression and constitutive cytokine release by primary human acute myeloid leukaemia cells. *Eur Cytokine Netw* **21**, 154–164 (2010).
31. Keeley, E. C., Mehrad, B. & Strieter, R. M. CXC chemokines in cancer angiogenesis and metastases. *Adv Cancer Res* **106**, 91–111 (2010).
32. Cheng, Y. J. *et al.* Id1 promotes lung cancer cell proliferation and tumor growth through Akt-related pathway. *Cancer Lett* **307**, 191–199 (2011).
33. Nowacki, S. *et al.* Expression of the tumour suppressor gene CADM1 is associated with favourable outcome and inhibits cell survival in neuroblastoma. *Oncogene* **27**, 3329–3338 (2008).
34. Jarmalaite, S. *et al.* Tumor suppressor gene ZAC/PLAGL1: altered expression and loss of the nonimprinted allele in pheochromocytomas. *Cancer Genet* **204**, 398–404 (2011).
35. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460**, 1132–1135 (2009).
36. Hochedlinger, K. *et al.* Reprogramming of a melanoma genome by nuclear transplantation. *Genes Dev* **18**, 1875–1885 (2004).
37. Lin, S. L. *et al.* Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. *RNA* **14**, 2115–2124 (2008).
38. Carette, J. E. *et al.* Generation of iPSCs from cultured human malignant cells. *Blood* **115**, 4039–4042 (2010).
39. Mathieu, J. *et al.* HIF induces human embryonic stem cell markers in cancer cells. *Cancer Res* **71**, 4640–4652 (2011).
40. Ron-Bigger, S. *et al.* Aberrant epigenetic silencing of tumor suppressor genes is reversed by direct reprogramming. *Stem Cells* **28**, 1349–1354 (2010).
41. Helman, E., Naxerova, K. & Kohane, I. S. DNA hypermethylation in lung cancer is targeted at differentiation-associated genes. *Oncogene* **31**, 1181–1188 (2012).
42. Liu, H., Kho, A. T., Kohane, I. S. & Sun, Y. Predicting survival within the lung cancer histopathological hierarchy using a multi-scale genomic model of development. *PLoS Med* **3**, e232 (2006).
43. Naxerova, K. *et al.* Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. *Genome Biol* **9**, R108 (2008).
44. Landi, M. T. *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* **3**, e1651 (2008).
45. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).
46. Fingleton, B. Matrix metalloproteinase inhibitors for cancer therapy:the current situation and future prospects. *Expert Opin Ther Targets* **7**, 385–397 (2003).
47. Tweardy, D. & Chang, J. C. Stat5: from breast development to cancer prognosis, prediction, and progression. *J Clin Oncol* **29**, 2443–2444 (2011).
48. Kuhn, T. S. *The structure of scientific revolutions* (University of Chicago Press, 1996).
49. Yoo, C. B. & Jones, P. A. Epigenetic therapy of cancer: past, present and future. *Nat Rev Drug Discov* **5**, 37–50 (2006).
50. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
51. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
52. team, R. d. c. *R: A language and environment for statistical computing. R foundation for statistical computing.* (2011).
53. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
54. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
55. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13 (2009).
56. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).


## Acknowledgements

We sincerely thank Siew Chin Hoong for his critical input in data analysis and Dr. Yu Qiang for reading the manuscript. This work was supported by the funding from Academic Research Fund (AcRF) Tier 1 Faculty Research Committee (FRC) grant, National University of Singapore.

## Authors contributions

D. M., C. M. K. and J. L. designed and performed the experiments, analyzed the results and wrote the manuscript. L. L. T. carried out some of the experiments. H. Y. provided critical comments on the analysis and manuscript. X. W. supervised the project and contributed with reagents.

## Additional information

**Supplementary information** accompanies this paper at http://www.nature.com/scientificreports

**Competing financial interests:** The authors declare no competing financial interests.



**How to cite this article:** Mahalingam, D. *et al.* Reversal of Aberrant Cancer Methylome and Transcriptome upon Direct Reprogramming of Lung Cancer Cells. *Sci. Rep.* **2**, 592; DOI:10.1038/srep00592 (2012).