Oncogene (2013), 1–12

www.nature.com/onc

ORIGINAL ARTICLE

An integrative analysis reveals functional targets of GATA6 transcriptional regulation in gastric cancer

R Sulahian[1,2], F Casey[3], J Shen[3], ZR Qian[1,4], H Shin[5], S Ogino[1,4], BA Weir[1], F Vazquez[1], XS Liu[5], WC Hahn[1,2,6], AJ Bass[1,2,6], V Chan[3] and RA Shivdasani[1,2,6]

Lineage-restricted transcription factors (TFs) are frequently mutated or overexpressed in cancer and contribute toward malignant behaviors; however, the molecular bases of their oncogenic properties are largely unknown. As TF activities are difficult to inhibit directly with small molecules, the genes and pathways they regulate might represent more tractable targets for drug therapy. We studied *GATA6*, a TF gene that is frequently amplified or overexpressed in gastric, esophageal and pancreatic adenocarcinomas. GATA6-overexpressing gastric cancer cell lines cluster in gene expression space, separate from non-overexpressing lines. This expression clustering signifies a shared pathogenic group of genes that GATA6 may regulate through direct *cis*-element binding. We used chromatin immunoprecipitation and sequencing (ChIP-seq) to identify GATA6-bound genes and considered TF occupancy in relation to genes that respond to GATA6 depletion in cell lines and track with *GATA6* mRNA (synexpression groups) in primary gastric cancers. Among other cellular functions, GATA6-occupied genes control apoptosis and govern the M-phase of the cell cycle. Depletion of GATA6 reduced the levels of the latter transcripts and arrested cells in G2 and M phases of the cell cycle. Synexpression in human tumor samples identified likely direct transcriptional targets substantially better than consideration only of transcripts that respond to GATA6 loss in cultured cells. Candidate target genes responded to the loss of GATA6 or its homolog GATA4 and even more to the depletion of both proteins. Many GATA6-dependent genes lacked nearby binding sites but several strongly dependent, synexpressed and GATA6-bound genes encode TFs such as *MYC*, *HES1*, *RARB* and *CDX2*. Thus, many downstream effects occur indirectly through other TFs and GATA6 activity in gastric cancer is partially redundant with GATA4. This integrative analysis of locus occupancy, gene dependency and synexpression provides a functional signature of GATA6-overexpressing gastric cancers, revealing both limits and new therapeutic directions for a challenging and frequently fatal disease.

Oncogene advance online publication, 9 December 2013; doi:10.1038/onc.2013.517

Keywords: transcriptional control of cancer; synexpression groups; somatic copy number alterations; ChIP-seq; GATA transcription factors

INTRODUCTION

Some lineage-restricted transcription factors (TFs) specify developing tissues and regulate cell-specific genes in adults. Cancers often amplify such TF genes, including *MITF* in melanoma,[1] *NKX2–1* in lung adenocarcinoma,[2] *SOX2* in squamous esophageal cancer[3] and *AR* in prostate cancer.[4] As tumors may depend on amplified TF genes,[2,3] they are potential targets for cancer therapy. However, TFs other than nuclear hormone receptors are notoriously difficult drug targets.[5,6] Therefore, core downstream genes and pathways might suggest alternative targets that are more sensitive to small molecules. *GATA4* and *GATA6* are amplified in up to 30% of gastric and esophageal adenocarcinomas,[7] and GATA6 depletion in the latter specifically impairs anchorage-independent cell growth.[8,9] We studied the dependencies and transcriptional functions of this TF.

Worldwide, stomach cancer is the second leading cause of cancer death.[10,11] Somatic copy number amplifications (SCNAs) or mutations of *ERBB2*, *EGFR*, *MET* and *FGFR2* offer avenues for targeted therapy in few patients.[12–14] Esophageal adenocarcinomas, which are closely related, frequently amplify *GATA6* and *GATA4*,[7] TF gene loci that show especially high expression in gastric and duodenal epithelia.[15,16] In mouse intestine, GATA6 levels are highest in the crypts, where cell proliferation is reduced in conditional *Gata6*$^{-/-}$ mice.[17] TF co-occupancy, determined by chromatin immunoprecipitation (ChIP), further suggests that GATA6 mediates crypt functions together with CDX2, a master intestinal regulator.[18] As gastric cancer frequently arises in a background of intestinal metaplasia,[19] this partnership suggests that GATA gene amplifications may promote proliferative, crypt progenitor-like properties in stomach epithelial cells.

GATA6 is also amplified in pancreas cancer;[20,21] however, interference with its functions is hampered by limited information about the targets of transcriptional control. To delineate core downstream genes, pathways and functions in gastric cancer, we examined genome-wide GATA6 occupancy in relation to GATA6-dependent gene expression in cell lines and GATA6-associated gene expression (synexpression) in human tumor samples. This approach revealed features, consequences and core transcriptional targets of GATA6 in gastric cancer.

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; [2]Department of Medicine, Harvard Medical School, Boston, MA, USA; [3]Novartis Institutes for Biomedical Research, Emeryville, CA, USA; [4]Department of Pathology, Harvard Medical School and Brigham & Women's Hospital, Boston, MA, USA; [5]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA and [6]Department of Medicine, Brigham & Women's Hospital, Boston, MA, USA. Correspondence: Dr RA Shivdasani, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA.
E-mail: ramesh_shivdasani@dfci.harvard.edu
Received 29 April 2013; revised 11 October 2013; accepted 18 October 2013

RESULTS

Amplification and expression of GATA genes in upper digestive tract cancers

Small regions on chromosomes 8p and 18q are focally amplified in 17–22% of stomach and gastro-esophageal junction adenocarcinomas.[7] GISTIC analysis[22] of these cases and public SCNA data from 321 additional primary stomach cancers identified *GATA6* and *GATA4* as the only genes within the minimal common areas of amplification (Figure 1a). Among hundreds of diverse cancers, high-level *GATA4* amplifications were largely confined to gastric cancer and *GATA6* amplifications to stomach and pancreas adenocarcinomas (Supplementary Figure S1A). GATA4 and GATA6 are homologous TFs that recognize the same DNA sequence and have overlapping functions in some mouse tissues,[23,24] suggesting that they may serve similar roles in gastric cancer. As *GATA6* amplifications are more common and GATA6 antibodies (Abs) perform well in tissue and chromatin studies, we concentrated on this TF.

GATA6 is expressed in normal human stomach epithelium, intestinal metaplasia and carcinoma (Figure 1b, Supplementary Figure S1B). Both *GATA6*-amplified and -unamplified gastric cancers showed strong immunostaining in most tumor cells (Supplementary Figure S1C), consistently stronger than in colorectal cancer (CRC, Figure 1b, Supplementary Figure S1B), where *GATA6* amplification is uncommon (Supplementary Figure S1A). Many gastric cancers also express the intestine-restricted factor CDX2, consistent with their likely origin in areas of intestinal metaplasia;[19] however, GATA6 levels showed no association with CDX2 expression or tumor cell differentiation (Supplementary Figure S1C). mRNA analysis of 290 gastric cancers in The Cancer Genome Atlas (https://tcga-data.nci.nih.gov/tcga/) showed high *GATA6* expression in a significant fraction of cases and *GATA4* overexpression in fewer cases (Figure 1c). Tumors with high *GATA6* amplification showed significantly higher mRNA levels than diploid samples and *GATA6* mRNA or protein (Supplementary Figure S1C) were rarely lost. In summary, *GATA6* amplification is common in gastric cancers and elevates expression, although tumors lacking amplification also may overexpress GATA6.

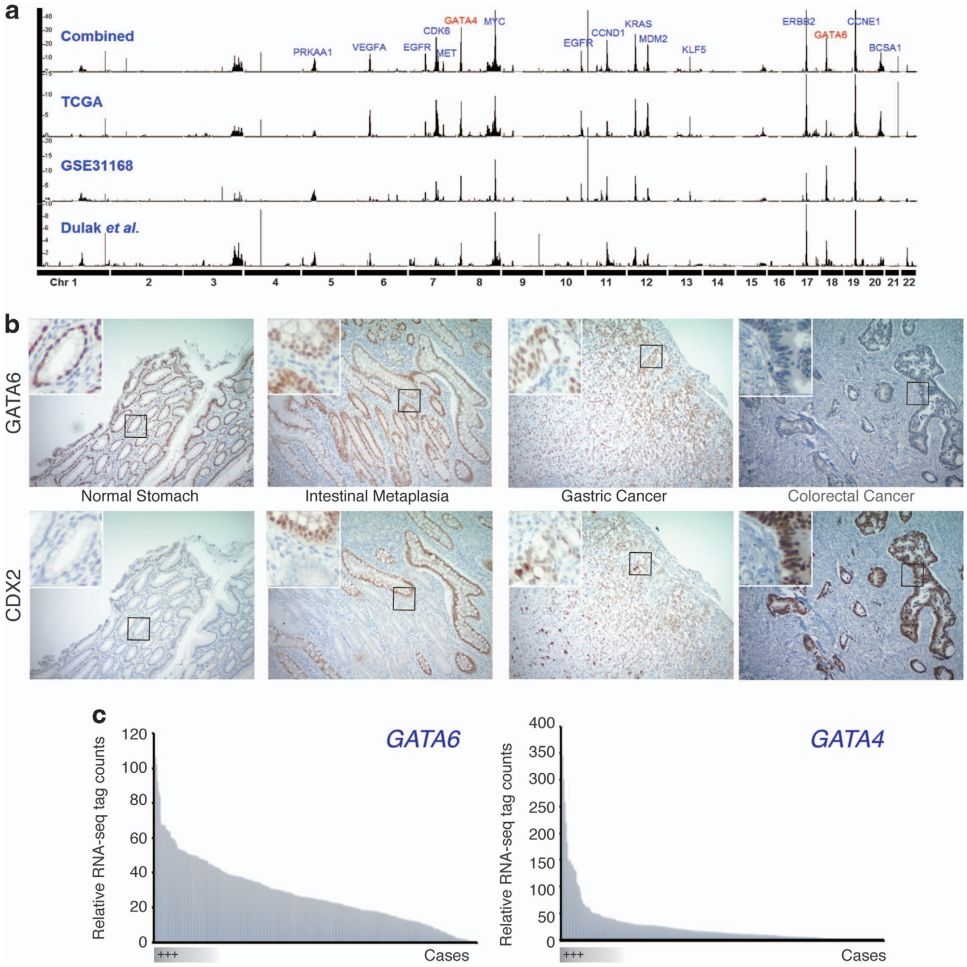

Figure 1. SCNAs in adenocarcinomas of the upper digestive tract. (**a**) Summary of SCNA data from SNP array analysis in studies to date: Dulak *et al.*,[7] GSE31168 in the Gene Expression Omnibus, and stomach adenocarcinoma STAD in The Cancer Genome Atlas (TCGA). Results of GISTIC analysis are represented by chromosome and show that the minimal common regions of amplification on chromosomes 8p and 18q encompass the single gene loci *GATA4* and *GATA6*, respectively. Also see Supplementary Figure S1A. (**b**) Representative immunohistochemical analysis of GATA6 and CDX2 in human tissue, showing abundance of both proteins in a gastric cancer with *GATA6* amplification. GATA6 is expressed in normal stomach epithelium and often increased in cancer, whereas CDX2 is expressed only in intestinal metaplasia and cancer; additional examples appear in Supplementary Figure S1B. The right panels display GATA6 and CDX2 expression in CRC. Original magnification $\times 10$, inserts show $\times 6$ additional digital magnification. (**c**) Collated *GATA* gene expression from RNA-seq data in the TCGA_stomach adenocarcinoma database, showing *GATA6* overexpression in a significant fraction of cases and that, although *GATA4* mRNA loss is common, primary tumors rarely lose *GATA6* expression. Color scales interpret the degree of GATA factor mRNA overexpression.

Expression of GATA6 and possible downstream genes in gastric cancer cell lines

Focal *GATA* gene amplifications were uncommon in gastric cancer cell lines (Supplementary Figures S1A and S2A). Only 2 of 26 lines, HUG1N and GC1Y, showed *GATA6*, and none showed *GATA4*, amplification (Supplementary Figure S2B). Protein expression among nine gastric cancer lines was highest in HUG1N, at levels similar to those found in the CRC cell line Caco2 (Figure 2a). Immunoblotting detected no GATA6 in GC1Y cells and two lines lacking amplification, AGS and SNU16, expressed almost the same GATA6 levels as HUG1N. Thus, gene amplification does not invariably produce excess protein and is only one basis for overexpression. Indeed, despite the paucity of SCNAs at chromosome 18p, gastric cancer cell lines commonly express abundant *GATA6* mRNA, and hierarchical clustering of gene expression similarities (Euclidean distance, 1000 most variant probe sets) across 36 lines revealed *GATA*-overexpressing lines as a distinct subclass (Figure 2b). The 15 cell lines in this group included all GATA6-expressing lines we detected with the help of immunoblotting: AGS, HUG1N, NCIN87 and SNU16. *GATA4* mRNA was overexpressed in fewer lines, many of which showed high *GATA6*. The presence of a characteristic mRNA profile in GATA6-overexpressing gastric cancers suggests that some of these genes represent its transcriptional targets. We chose HUG1N and AGS cells to study GATA6 functions because these lines express more GATA6 than GATA4 (Figure 2a) and HUG1N, in particular, carries multiple gene copies.

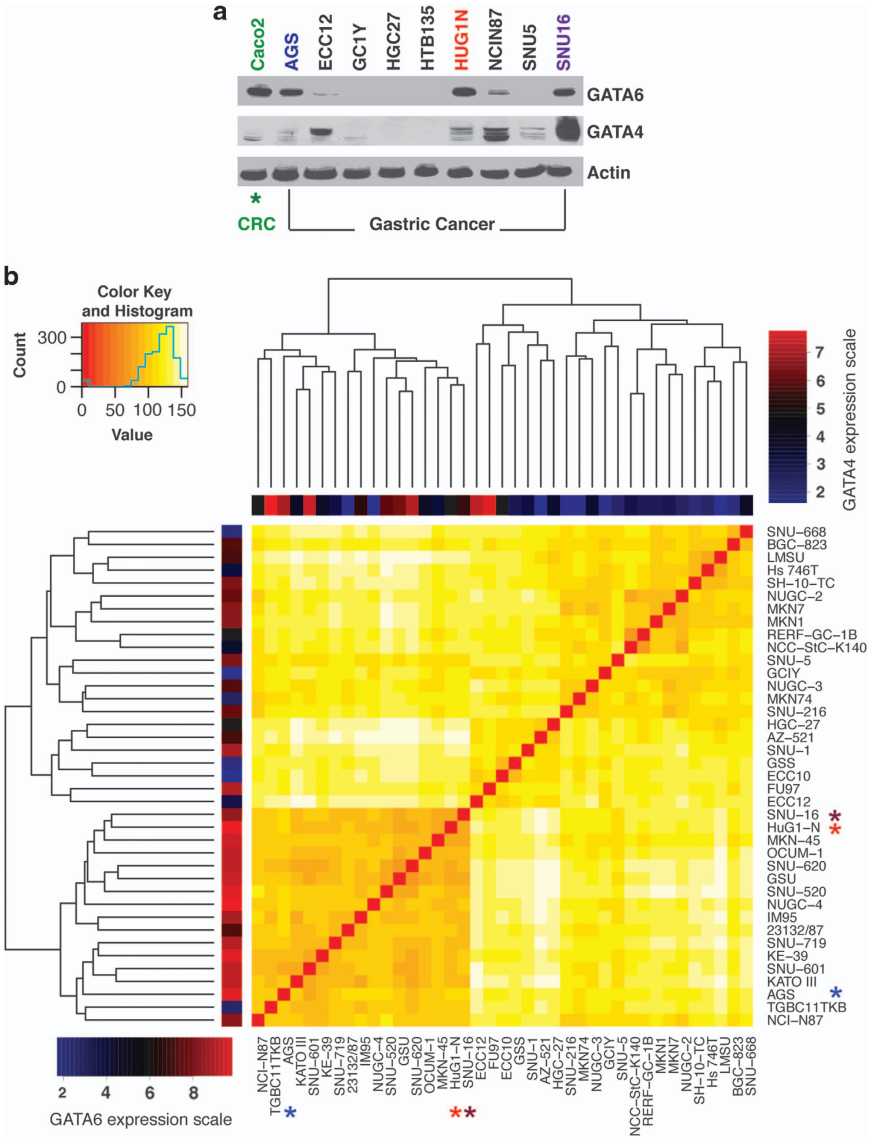

Figure 2. GATA factor expression and synexpression in gastric cancer cell lines. (**a**) Immunoblot analysis of GATA6 and GATA4 in Caco2 CRC cells and nine gastric cancer lines chosen for presence (HUG1N, GC1Y) or absence of *GATA6* gene amplification. (**b**) Hierarchical clustering of mRNA expression in 36 gastric cancer cell lines, based on the 1000 probe sets that vary the most across this cell line collection (variance calculated by Euclidean distance). Bars with a red–blue scale represent relative expression levels of *GATA6* (vertical bar) and *GATA4* (horizontal bar) mRNAs in each cell line. This analysis revealed two distinct tumor classes. One class encompasses *GATA6* (and/or *GATA4*) mRNA-overexpressing cell lines (bottom left), including all GATA6-expressing lines from the immunoblot analysis (**a**, which are highlighted by asterisks). Gene expression in these 17 cell lines clusters separately from all others, revealing that GATA factors associate specifically with a sizable body of genes and putative transcriptional targets in gastric cancer. The data also indicate that GATA6 is often (but not always) highly expressed, even without gene amplification.

Effects of GATA6 depletion on cell replication and gene expression

Using lentivirus-delivered short hairpin RNAs (shRNAs) to deplete GATA6, we confirmed that cells selected for viral integration were efficiently depleted compared with those receiving a control shRNA (Figure 3a). AGS, SNU16 and Caco2 cells all proliferated slower after GATA6 depletion (Figure 3a). Proliferation of HUG1N was unaffected by short-term depletion ($\leqslant$6 cell doublings), suggesting possible adaptation in culture to escape dependency. By contrast, in a pooled-shRNA screen of 89 cancer cell lines,[25] GATA6 dependency was strongly correlated with its expression level: over $\geqslant$16 cell doublings, HUG1N and other high GATA6-expressing gastric and CRC cell lines were among the most susceptible ones to GATA6 depletion (box no. 3 in Figure 3b). Moreover, transcripts affected by GATA6 depletion in HUG1N and Caco2 cells were significantly correlated with *GATA6*-synexpressed mRNAs in primary tumors and cell lines (Figure 3c). Thus, transcripts that respond to GATA6 deficiency closely match *in vivo* gene expression and are likely enriched for *bona fide* target genes. Transcripts altered in GATA6-deficient HUG1N cells were correlated significantly with synexpressed genes in CRC and even better with synexpressed genes in gastric cell lines and tumors (Figure 3c), suggesting that GATA6 may control some genes specifically in the stomach or colon and other genes in both organs.

Compared with cells treated with a control shRNA, 805 transcripts dropped in levels in GATA6-deficient HUG1N cells and 595 transcripts were increased. Similar to its homolog GATA1,[26,27] GATA6 may both activate and repress genes directly or one function could be direct and the other indirect. A non-parametric test of all $\log_2$ fold-changes (FCs) in transcript levels in GATA6-depleted HUG1N cells revealed that the mean changes were significantly enriched for reduced expression (one-sided Wilcoxon-signed rank test, $P = 2 \times 10^{-8}$), suggesting that in unperturbed cells GATA6 activates many more genes than it might repress. In line with this idea, no functional Gene Ontology category was enriched among 595 transcripts that increased in

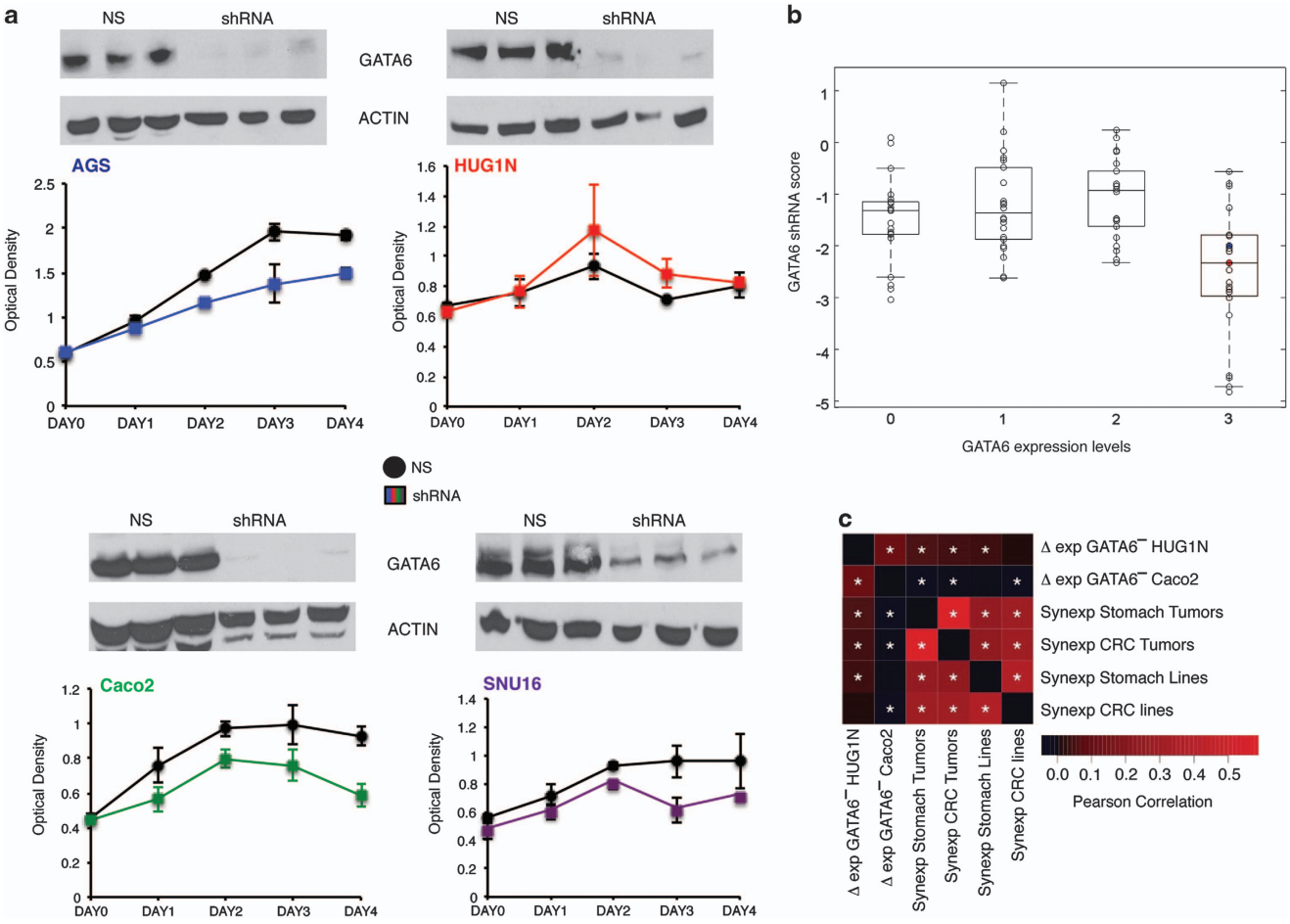

Figure 3. Loss of GATA6 impairs growth and affects gene expression in gastrointestinal cancer cells. (**a**) GATA6-overexpressing cell lines were infected with lentiviral vectors carrying *GATA6*-specific or a nonspecific (NS) 21-bp shRNA, selected in puromycin for viral integration and assessed daily in triplicate for cell viability using MTT assays. Immunoblots verified efficient GATA6 depletion in all cell lines for each replicate, using actin as a loading control, and each plotted value represents the mean ($\pm$ s.d.) optical density. (**b**) Correlation of GATA6 expression levels with GATA6 dependency in an unbiased, pooled-shRNA screen of >11 000 genes in human cancer cell lines (Cheung *et al.*[25]). Eighty-nine cell lines were grouped into four bins according to the relative level of *GATA6* mRNA, 0 = absence, 3 = high. Relative response in growth of these lines to GATA6 shRNA is shown in box plots with the centers indicating median sensitivity, edges at the 25th and 75th percentiles, and whiskers extending to the extreme non-outlier data points. Red and blue dots represent HUG1N and AGS, respectively, and many high-expressing, shRNA-sensitive cell lines were from CRCs. (**c**) Concordance of GATA6 expression correlations in primary tumors and cell lines and transcripts differentially expressed in GATA6-deficient cells. Pearson's correlations with GATA6 expression were computed from primary tumor (125 stomach, 1,732 CRC) and cell line (36 gastric, 59 CRC) data sets. Correlation coefficients for every gene were then correlated with fold-changes (Δ exp) from shRNA experiments in HUG1N and Caco2 cells. The resulting Pearson's correlations are displayed in a heat map, with zeroing of the diagonal. Asterisks denote significant concordance based on a *t*-test and $P < 0.001$.

HUG1N cells or 144 transcripts that increased in both Caco2 and HUG1N. In contrast, functional categories were readily apparent among transcripts that declined upon GATA6 loss (Supplementary Figures S3A and B), providing clues to its cellular functions.

The 212 transcripts reduced in both Caco2 and HUG1N cells depleted of GATA6 represent a highly significant, non-random overlap (Figure 4a; odds ratio 9.0, $P < 4 \times 10^{-103}$ by Fisher's exact test) and were enriched for Gene Ontology terms related to the cell cycle, particularly the M-phase (Figure 4b). GATA6 knockdown in Caco2, AGS and SNU16 cells resulted in smaller G1- and S-phase fractions and higher G2/M-phase fractions than control shRNA-treated cells, indicating M-phase dysfunction (Figure 4c). Phospho-histone H3 immunofluorescence and immunoblot analyses verified this G2/M-phase arrest (Figures 4d, e). Although the cell

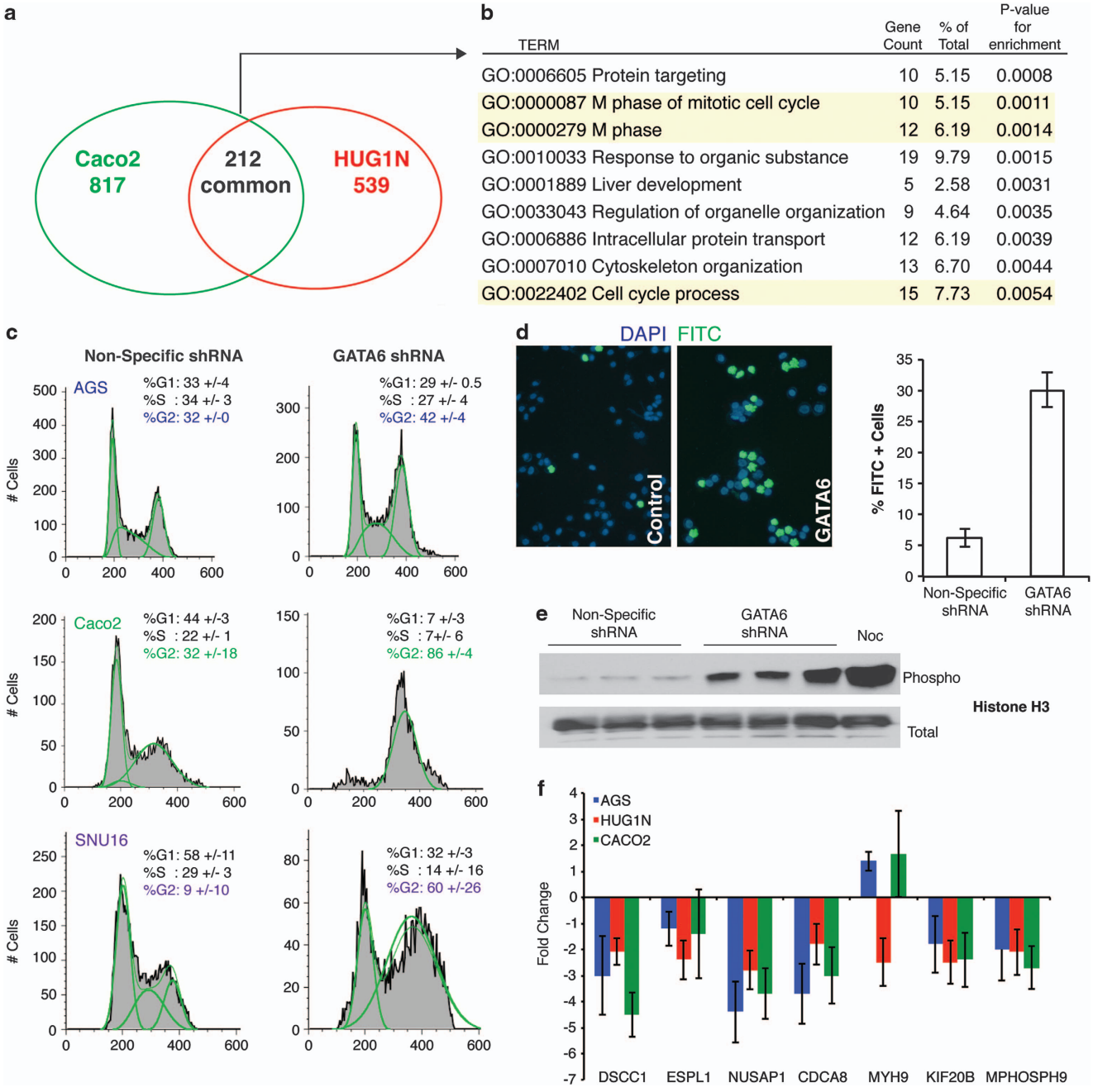

Figure 4. Depletion of GATA6 impairs cell growth and cell cycle progression in gastrointestinal cancer cells. (**a**) Venn diagram showing the overlap in transcripts reduced in GATA6-depleted Caco2 and HUG1N cells. (**b**) Table of functional categories enriched among the 212 genes reduced in both cell lines and statistical significance of this enrichment over the genome background. Shaded Gene Ontology terms highlight the cell cycle, particularly the M-phase. (**c**) Flow cytometry analysis of the cell cycle in GATA6-depleted AGS, Caco2 and SNU16 cells. Propidium iodide staining of cell cycle phases is shown; each value represents the mean ± s.d. of three replicates. (**d**, **e**) Immunofluorescence (**d**), merged images of green Phospho-histone H3 (PH3) immunostaining and blue DAPI and immunoblot (E, 3 replicate lanes, Noc = 100 ng/ml nocodazole-positive control) analysis of phospho-histone H3. GATA6-depleted AGS cells show G2/M-phase arrest, compared with cells treated with a control shRNA. (**f**) Relative expression in AGS, HUG1N and Caco2 cells of M-phase-implicated transcripts (derived from the Gene Ontology classification in (**b**) that showed reduced expression in GATA6-depleted cells. qRT–PCR results are expressed as the average fold-change ± s.d. relative to cells treated with a nonspecific shRNA in three independent samples.

cycle seemed overtly undisturbed in HUG1N cells, qRT–PCR analysis showed reduced M-phase-related transcripts in GATA6-depleted cell lines, including HUG1N (Figure 4f). These data implicate late cell cycle control as a core GATA6 function.

Delineation of GATA6 cistromes

To identify primary transcriptional targets, we used chromatin immunoprecipitation and sequencing (ChIP-seq) to localize GATA6 binding in HUG1N cells. After adjusting for results from mock immunoglobulin ChIP, we mapped nearly 7000 high-confidence GATA6-binding sites in HUG1N cells and considered the data together with occupancy previously mapped in Caco2 cells.[18] High sequence conservation within ChIP fragments and greatest enrichment for the consensus WGATAA motif, present in at least 70% of bound regions, implied direct occupancy at most sites (Figure 5a). Although GATA6 occupied many regions within 3 kb of transcriptional start sites, it predominantly bound DNA far from promoters (Figure 5a). For example, binding near *FGFR2*, which is implicated and occasionally amplified in gastric cancer,[28,29] occurs 8 and 13 kb from the promoter (Figure 5b). Moreover, GATA6 sometimes occupied exactly the same sites in HUG1N and Caco2 cells (for example, *ANKRD30BL* locus, Figure 5c) or different sites near the same locus (for example, *FGFR2*, Figure 5b) but often bound sites in only 1 of the 2 cell lines (Figure 5d). Binding unique to each line did not trivially reflect deletion of the corresponding

Figure 5. Genome-wide GATA6 occupancy in gastrointestinal cancer cell lines. (**a**) High sequence conservation (left) and significant enrichment of the canonical WGATAR sequence motif (middle) among GATA6 ChIP sites, whose pie-chart distribution across the genome (right) reveals localization mainly in introns and intergenic regions far from promoters. (**b**–**d**) Wiggle tracks from GATA6 ChIP-seq in HUG1N and Caco2 cells illustrate diverse binding patterns: same locus, different sites (**b**), *FGFR2*; same locus, overlapping sites (**c**), *ANKRD30BL*; and exclusive occupancy in one line (**d**), *KCNJ5*. (**e**) Venn diagram representing the overlap of GATA6 occupancy in HUG1N, AGS and Caco2 cells. Representative wiggle traces from the 1007 sites common to the two stomach cell lines demonstrate co-occupancy in all three lines at the *FDPS* promoter (top) and exclusive binding in the two gastric cancer cell lines in a *BTRC* intron (bottom).

genomic region in the other. To determine whether these differences reflect tissue-specific binding in the stomach and colon, we mapped GATA6 occupancy in overexpressing AGS gastric cancer cells and observed modest overlap among the three cell lines (Figure 5e). Sites confined to one or two lines seemed not to reflect technical artifacts in the others but *bona fide* absence of GATA6 at those regions (Figures 5b, d and e). In aggregate, ChIP-seq revealed diverse GATA6 binding to cellular DNA, with cell line-specific, tissue-specific and shared binding sites. We reasoned that common binding sites and transcriptional targets might represent the most pertinent outcome of frequent gene amplification.

GATA6-dependent genes, synexpression and DNA occupancy together impute target genes

To determine roles of GATA6 binding in gene regulation, we measured concordance of FCs in gene expression in GATA6-deficient HUG1N cells with its occupancy near affected genes. Figure 6a plots the Pearson correlations between binding sites and genes affected by GATA6 depletion; negative correlations denote higher concordance. Altered gene expression in HUG1N cells correlated with GATA6 binding in HUG1N cells better than binding in Caco2 cells, and ChIP-seq in AGS gastric cancer cells did not improve this correlation. Figure 6b shows heat maps of ordered average FCs per gene, from most repressed (left, blue) to most activated (right, red), and the corresponding average GATA6 binding near those genes (<20 kb, brighter yellow denotes higher occupancy). GATA6 binding in HUG1N cells alone (Figure 6b) or in both HUG1N and AGS cells (data not shown) correlated with altered gene expression. To identify direct transcriptional targets, we first considered genes that bind GATA6 within 20 kb of the transcriptional start site and also reduce expression in GATA6-depleted cells. Gene Ontology analysis of such genes in HUG1N (Supplementary Figure S3C) suggests that GATA6 directly regulates 75 genes that control cell replication; 41 of these genes promote cell proliferation, and GATA6 also binds DNA near dependent genes implicated in cell death. Among genes that bind GATA6 and depend on it in Caco2 cells, 86 genes regulate cell proliferation (Supplementary Figure S3D). 'Digestion' and 'Response to hormone stimulus', function strongly linked to gastrointestinal physiology, were also highly enriched among genes near GATA6-binding sites in HUG1N and Caco2 cells, respectively, suggesting a breadth of cellular roles.

Reasoning that *bona fide* transcriptional targets might additionally co-express with *GATA6* RNA in human cancer tissues, we integrated these cell line data with information on synexpressed genes—that is, correlation scores for every gene's expression with *GATA6* expression. We used Principal Component Analysis (PCA)[30] to determine directions of maximal variation in the combined space of altered gene expression in GATA6-deficient HUG1N and Caco2 cells (axes 1 and 2) and *GATA6* synexpression across four large data sets of colon and gastric cancer tissues and cell lines (axes 3–6); the first two PCA axes accounted for 53% of the variation. A linear combination of normalized PCA scores for each gene, incorporating cell line and tumor sample data, yielded a ranked list of candidate GATA6 transcriptional targets (Supplementary Table S1). This integrated PCA ranking (Figure 6c) gave far stronger concordance with individual binding sites than did GATA6-dependent gene expression in cell lines (Figure 6a). HUG1N binding showed the highest concordance; however, GATA6 occupancy in Caco2 and AGS cells also was significantly correlated. PCA ranking (Figure 6d) also performed significantly better than GATA6-dependent gene expression (Figure 6b) at assigning probable regulatory functions to GATA6-binding sites.

We compared the genes best correlated in PCA analysis (using z less than -4 as a stringent measure; Figure 6e shows the 20 highest scoring genes) with the most anticorrelated genes (z more than $+4$, Supplementary Table S1). Among the 109 genes highly associated with *GATA6* in expression space, 11 genes showed GATA6 occupancy within 20 kb of the transcriptional start site in all three cell lines (shaded yellow in Supplementary Table S1) and 31 genes bound GATA6 in two or more lines (highlighted in Figures 6e and g, shaded green in Supplementary Table S1). Thus, within a reasonable distance for TF regulation of nearby genes,[31,32] the data reveal a high likelihood that GATA6 activates synexpressed genes. If GATA6 also represses transcription, we might expect to detect its binding near the 97 genes most anticorrelated with *GATA6* expression (z more than $+4$). However, none of these genes showed nearby GATA6 binding in all three lines ($P=9e^{-4}$, Fisher's exact test) and only one gene bound GATA6 in two cell lines ($P=6.5e^{-9}$, Figure 6f). The disparity in GATA6 binding near genes that are synexpressed and those that are anticorrelated in expression persisted for weaker associations (z less than -3 or more than $+3$, Figure 6F, Supplementary Table S1). Together, these data argue for a predominantly activating function for GATA6. Moreover, as most GATA6-dependent, co-expressed genes do not bind GATA6 within 20 kb of the promoter, they may represent mainly indirect transcriptional targets.

Integrative analysis of GATA6 occupancy, gene dependencies and tumor synexpression

At least 30% of GATA6-bound regions detected in only one cell line lacked canonical GATA motifs and may therefore represent indirect binding or spurious signals. By contrast, all of the hundreds of sites common to two or three cell lines carried a consensus GATA motif. Thus, shared sites reflect occupancy best, and although HUG1N and Caco2 cells shared only a fraction of binding sites (Figure 5e), these might denote authentic, common target genes. Genes that control hormonal responses, cell proliferation and lipid metabolism were significantly enriched near such shared binding sites (Figure 7a). Indeed, 18 genes associated with cell proliferation showed similar GATA6 dependency and GATA6 binding in both cell lines (for example, *HES1* promoter, Figure 7a). Eight of these likely *bona fide* GATA6 transcriptional targets encode TFs.

To identify targets most relevant to gastric cancer, we captured genes with high PCA rank and nearby GATA6 binding in HUG1N and AGS cells, irrespective of occupancy in Caco2 (Figure 7b). Target genes determined by integrative analysis of expression and GATA6 binding encode diverse transcriptional, secretory, survival, cytoskeletal and metabolic factors and, as such, fail to deliver a unifying view of GATA6 function. As binding in gastric cancer cells generally matched that in CRC cells (yellow shading in Figure 7b, illustrated in Figure 7d; see Venn diagram in Figure 5e), we considered 733 sites present in all three cell lines, noting that 49 of the sites lie within 20 kb of genes that depend on GATA6 in both Caco2 and HUG1N cells. The Gene Ontology biological process associated with the highest statistical significance was Regulation of Transcription (Figure 7c), as represented by TF genes *CDX2, CEBPG, HES1, IRF8, LRRFIP1, MLXIPL, MYC*, *NR5A2* and *RARB*. All nine of these transcripts were reduced in GATA6-depleted AGS, HUG1N and Caco2 cells (Figure 7d) and GATA6 bound the same sites in each cell line (for example, 8 kb upstream of *CEBPG*, Figure 7e). Five of these nine TF genes showed very little expression in gastric cancer cell lines that lack both GATA6 and GATA4 (Supplementary Figure S3E), and four of them fell within the highest percentile of PCA ranks, with z-scores less than -3.36 (Figure 7d). These data reveal a strong correlation with *GATA6* in expression space. Other TFs (NR0B2, MYCN and KLF5) also were strongly correlated in expression but bound GATA6 in one or two cell lines, and the gene with the highest PCA rank, *FOXE1* ($z=-23$), also encodes a TF. Identification of these candidate target genes might explain indirect GATA6 effects on gene regulation and suggests that it governs many genes through these TFs.

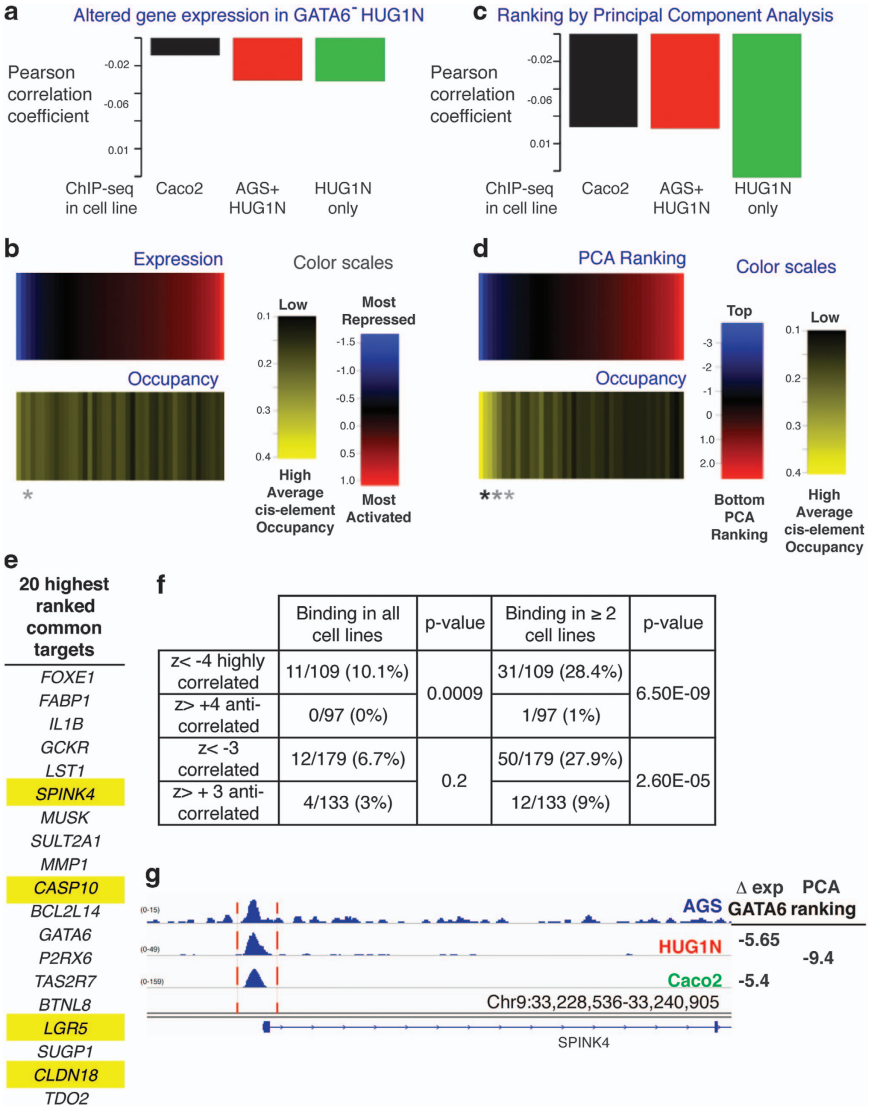

Figure 6. Identification of GATA6 target genes from the analysis of DNA binding and effects on gene expression. (**a**) Pearson's correlation coefficients of GATA6 binding (ChIP-seq data) and genes affected by GATA6 depletion (microarray analysis following shRNA treatment). GATA6 binding was encoded as 0 or 1; negative correlations denote higher concordance. (**b**) Top: heat map of sorted transcript fold-changes after GATA6 depletion, averaged over bins of 500 genes and ordered from most reduced (left, blue) to most increased (right, red). Bottom: heat map of average GATA6 occupancy within 20 kb of the corresponding 500-gene bins. (**c**) Pearson correlation coefficients of nearby GATA6 binding and the PCA ranks of genes based on reduced expression in GATA6-deficient cells and synexpression in human cancer specimens and cell lines; negative correlations denote higher concordance. (**d**) Top: heat map of combined PCA ranking score, averaged over windows of 500 genes and ordered from best correlated (left, blue, high negative *z*-scores) to the least correlated (right, red, high positive *z*-scores). Bottom: heat map of GATA6 occupancy within 20 kb of the corresponding bins of 500 PCA-ranked genes. (**e**) The 20 best-scoring genes by PCA rank, extracted from Supplementary Table S1; genes showing nearby GATA6 binding in at least two cell lines are highlighted. (**f**) GATA6 occupancy in cell lines within 20 kb of the transcriptional start site (TSS) of PCA-ranked gene groups. (**g**) Representative wiggle trace showing GATA6 occupancy at the *SPINK4* locus, representing a high-ranking gene from **e**.

Functional redundancy among GATA factors

As GATA6 and GATA4 recognizes the same DNA sequence, are often co-expressed and have overlapping functions in some tissues,[23,24] we asked next whether the two homologous TFs might share transcriptional targets in human gastric cancer. Lentiviral delivery of specific shRNAs depleted GATA4 efficiently in AGS and HUG1N gastric cancer cells (Figure 8a), which we selected because they express both proteins (Figure 2a). The same representative cell cycle and TF genes that respond to GATA6 deficiency and bind GATA6 showed significantly reduced levels in GATA4-depleted cells (Figure 8b). To test whether the two factors are truly redundant, we examined the expression of the same genes following shRNA-mediated depletion of GATA4, GATA6 or both in HUG1N cells. Some target genes showed similar effects upon the loss of one or both TFs; however, most transcripts showed at least an additive effect (Figure 8c). The heightened sensitivity to dual TF depletion indicates that both GATA4 and GATA6 contribute toward the regulation of a wide array of gastric cancer genes (Figure 8d).

DISCUSSION

GATA4 and GATA6 regulate diverse digestive epithelia.[17,23,33,34] Their genes are amplified in a high fraction of gastric and gastro-esophageal junction adenocarcinomas and frequently overexpressed, even in the absence of gene amplification, and

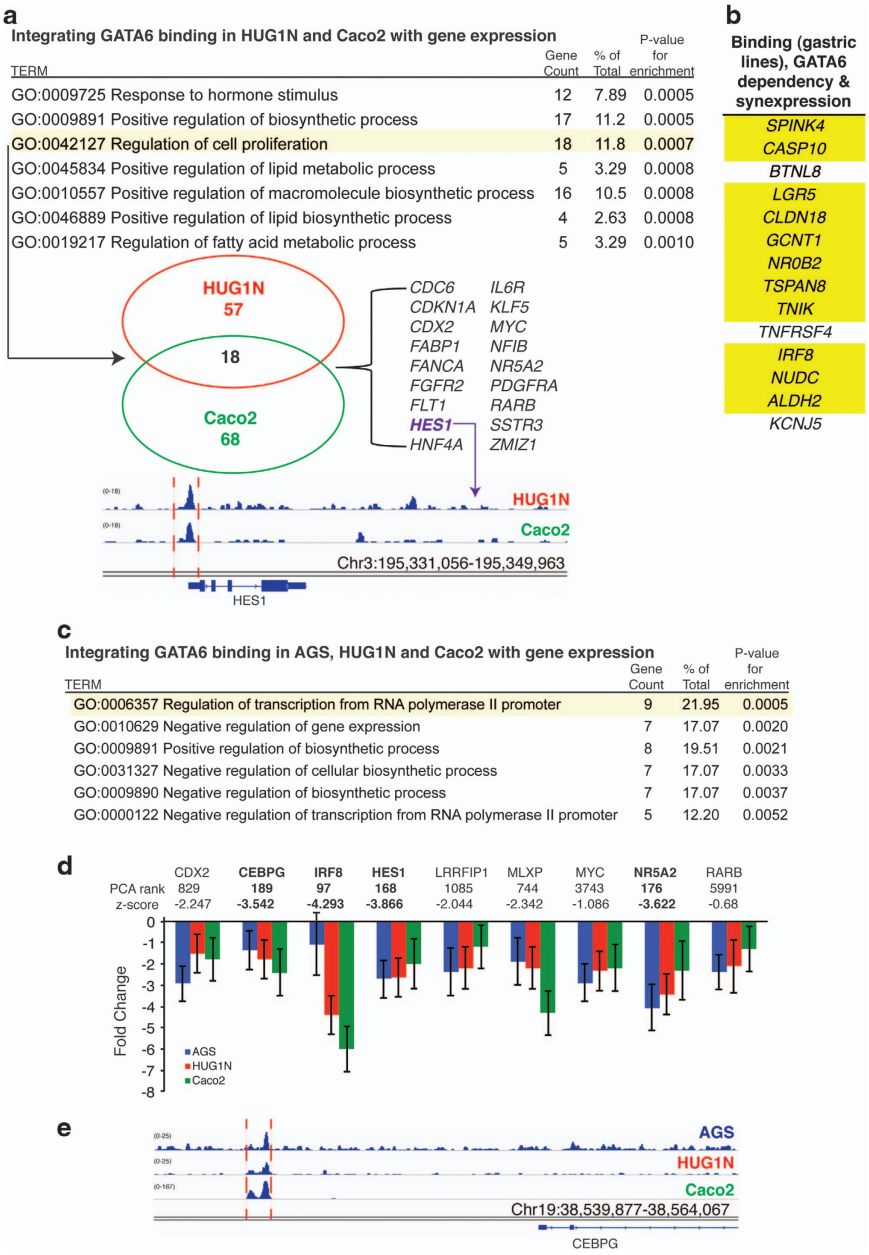

Figure 7. Integrative analysis of GATA6 occupancy, dependent genes and synexpression groups to delineate primary transcriptional targets. (**a**) Biological process Gene Ontology (GO) terms enriched among genes that bind GATA6 and require it for expression in both HUG1N and Caco2 cells. A Venn diagram illustrates the 18-gene overlap between two cell lines of GATA6-regulated and GATA6-occupied genes associated with cell proliferation and a representative data trace shows GATA6 binding at the *HES1* promoter. (**b**) Genes with a significant *z*-score in PCA and nearby GATA6 occupancy in gastric cell lines. Yellow shading marks genes that also bind GATA6 in Caco2 CRC cells. (**c**) Transcripts ascribed to the GO term 'Transcriptional regulation' represent the most enriched category of genes with common GATA6-binding sites and levels that respond to GATA6 depletion in AGS, HUG1N and Caco2 cells. (**d**) qRT–PCR for these nine TF genes confirmed reduced expression in HUG1N and Caco2 and showed similarly reduced levels in GATA6-depleted AGS gastric cancer cells. Values represent the average fold-change ± s.d. in three independent experiments. PCA rank and *z*-score for each gene are listed, and genes showing extreme association with *GATA6* in expression space are marked in bold. (**e**) GATA6 occupancy at one of these nine TF loci, *CEPBG*.

GATA6 is also amplified in pancreatic cancer—a related neoplasm.[20,21] Cell replication and colony assays suggest GATA6 dependence in esophageal and pancreas cancers;[8,9,21] however, the transcriptional basis of this dependence is unknown. Among 39 gastric cancer cell lines, GATA6 overexpression delineates a disease subset that expresses many genes in common, likely reflecting GATA6 regulatory functions. Genome-wide GATA6 occupancy, considered in relation to GATA6-dependent gene expression, provided one layer of insight into target genes. Transcripts that increased in GATA6-depleted cells, possibly owing to relief of GATA6-mediated repression, rarely showed GATA6 occupancy higher than the genome background. In contrast, genes with reduced expression in GATA6-deficient cells were enriched for nearby GATA6 binding and for functions related to cell cycle M-phase control, digestion, metabolism and hormonal responses. Thus, in gastrointestinal cells, GATA6 coordinately activates genes of related function, with perhaps little direct transcriptional repression, distinct from its homolog GATA1, which

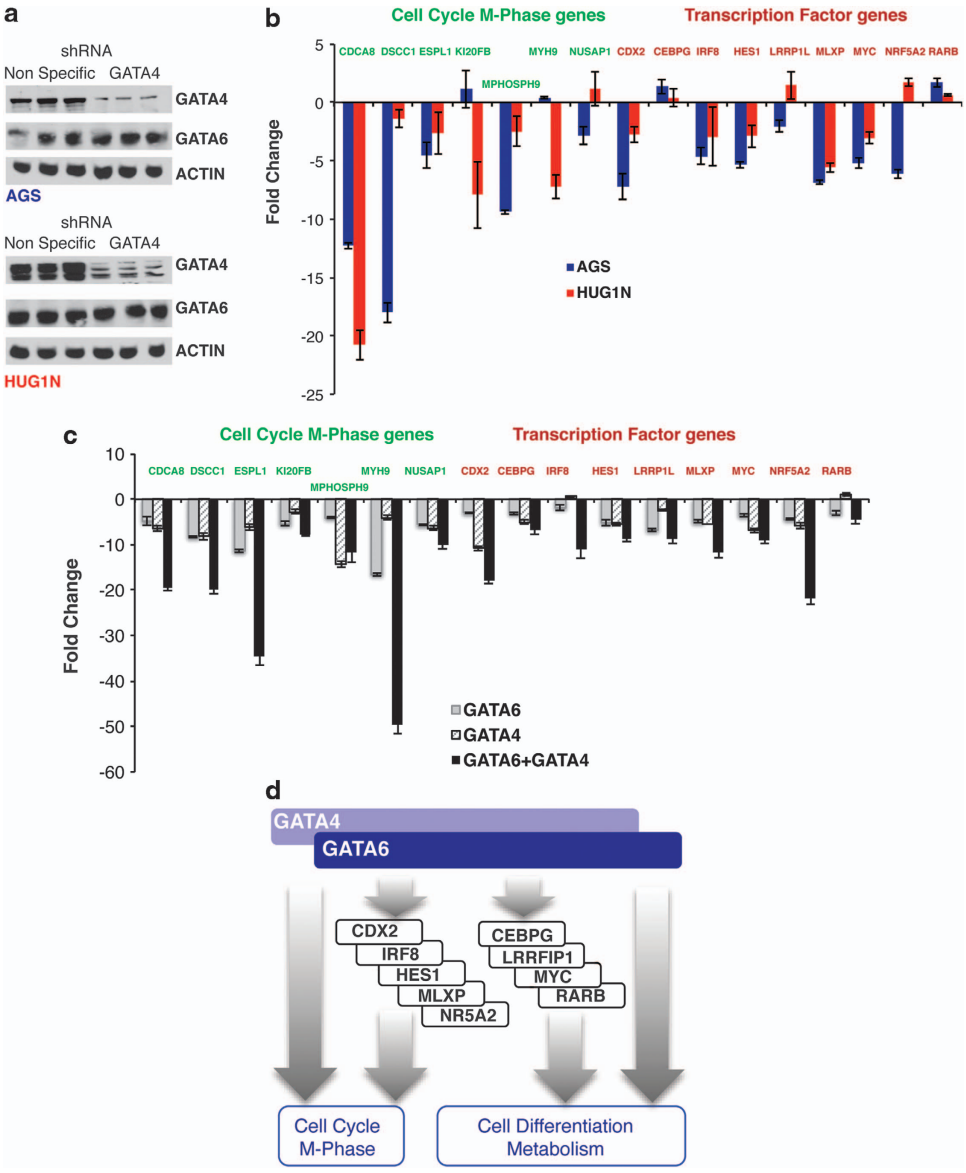

Figure 8. Additive and redundant effects of GATA6 and GATA4. (**a**) Immunoblot evidence that *GATA4* shRNA depletes GATA4 but not GATA6 protein in AGS (top) and HUG1N (bottom) gastric cancer cells. The lanes represent triplicate samples from one of the three experiments. (**b**) qRT–PCR analysis showing GATA4 dependency of most GATA6 target genes in HUG1N and AGS cells. GATA4 shRNA reduced *GATA4* mRNA levels 17- (AGS) to 19- (HUG1N) fold and relative levels of candidate cell cycle (green labels) and TF (brown labels) target transcripts of GATA regulation are shown from three independent replicates. (**c**) qRT–PCR in HUG1N cells, showing that most target genes respond to GATA4 or GATA6 deficiency and more significantly to loss of both, revealing redundant activities. shRNAs reduced *GATA6* and *GATA4* mRNA levels about 30-fold and relative mRNA levels of target genes were measured in three replicates. (**d**) Model for GATA function in gastric cancer. GATA6, in partial redundancy with GATA4, controls diverse genes, including those involved in cell differentiation and late phases of the cell cycle. Although some genes are direct transcriptional targets, others are controlled indirectly through GATA-factor regulation of intermediary TF genes, such as *CDX2*, *HES1* and *NR5A2*.

activates and represses genes in blood cells.[27] There is no obvious mechanism for this apparent difference, other than the transcriptional contexts in epithelial and blood cells.

Our consideration of gene synexpression across primary gastric and colorectal cancers significantly improved imputation of likely direct targets and functions. Most GATA6-dependent, synexpressed genes, however, lack GATA6 occupancy within $\geqslant$20 kb, implying indirect regulation through other TFs. Sequence-specific TFs that may fulfill this role are readily found among transcriptional targets that bind GATA6 in multiple cell lines and show extreme associations in expression space: CEBPG, IRF8, HES1, NR5A2 and CDX2; we propose that these TFs execute many cellular functions downstream of GATA6 (Figure 8d). As TFs generally make poor drug targets, our observations have implications for the selective treatment of GATA6-overexpressing gastric cancers. Suitable targets for new drugs might include enzymes that modify components of this TF network or nuclear hormone receptors such as NR5A2.

GATA6 overexpression begins early in esophageal neoplasia[8] and the spectrum of transcriptional targets we identify supports early biologic effects. Control of the M-phase genes and arrest of GATA6-depleted cells late in the cell cycle, for example, may explain accelerated cell replication. Target genes include *CDX2*, a TF that controls intestinal differentiation[35,36] and partners with GATA6 to regulate intestinal cell growth.[18] Presence of both CDX2 and GATA6 in many gastric cancers reflects the prevalence of

intestinal differentiation in this disease[37] and their partnership may promote tissue-specific proliferation. Among other intestinal gene targets (for example, *FABP1*, *SPINK4*, *GCNT1* and *VILL*), *LGR5* in particular marks intestinal[38] and some gastric[39] stem cells. At least two TF target genes, *MYC* and *NR5A2*, are components of self-renewal and pluripotency networks[40] and we implicate GATA6 in regulating *HES1*, a transcriptional target of Notch signals.[41] GATA6 activity at the *HES1* promoter is noteworthy because Notch signaling has an important role in mouse stomach epithelial stem and progenitor cell activity[42] and because HES1 and other GATA factors regulate each other's expression in other tissues.[43,44] Thus, GATA6 targets in gastric cancer place its activity within pathways for cell replication and self-renewal, differentiation and transcriptional control.

Epigenetic and signaling determinants of the GATA6 cistrome in each tissue and tumor are likely complex. Although mutations and other features may influence TF binding, we reason that a shared transcriptional program mediates GATA6-dependent oncogenic functions. We propose a core set of target genes that GATA6 occupies in multiple cell lines and that associate tightly with *GATA6* expression in primary gastric cancers. The homologous protein GATA4 seems to regulate the same target genes redundantly with GATA6.

MATERIALS AND METHODS

Tissue samples and copy number analysis

SCNAs in gastric and G-E junction adenocarcinomas were reported recently[7] and supplanted with extensive public data from The Cancer Genome Atlas and GSE31168 series in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). Immunohistochemistry was performed on tumor specimens archived after fixation in formalin and embedding in paraffin. Gastric cancer samples ($n = 125$) used for synexpression analysis were derived from four GEO series, GSE19826, GSE2109 and GSE13911, GSE22377, and 24 additional tumors assayed on Affymetrix HG U133 Plus 2.0 arrays. Colon cancer samples ($n = 1732$) are from GEO series GSE10714, GSE13059, GSE13067, GSE13294, GSE13471, GSE14333, GSE17536, GSE17537, GSE17538, GSE18088, GSE18105, GSE20916, GSE2109, GSE21510, GSE23878, GSE26682, GSE26906, GSE28702, GSE31595, GSE33113, GSE33114, GSE4107, GSE4183 and GSE9348, and 11 additional samples. Transcript data in gastric and colon cancer cell lines are from the Cancer Cell Line Encyclopedia (http://www.broadinstitute.org/ccle/home), normalized by the GCRMA method.[45] PCR primers used to confirm copy number are listed in Supplementary Table S2.

Cell lines

All cell lines were obtained from the American Type Culture Collection. Caco2, AGS, HTB135, NCIN87 and GC1Y were cultured in Dulbecco's Modified Eagle Medium containing 10% fetal bovine serum. SNU5 and HGC27 cells were maintained in Iscove's Modified Dulbecco Medium containing 20% fetal bovine serum. HUG1N, SNU16 and ECC12 were cultured in Roswell Park Memorial Institute-1640 medium containing 15% fetal bovine serum. All media were purchased from Life Technologies (Grand Island, NY, USA) and supplemented with penicillin and streptomycin (Life Technologies).

RNA and protein expression

Total RNA was isolated using Trizol (Invitrogen, Grand Island, NY, USA), treated with the RNeasy Mini Kit (Qiagen, Valencia, CA, USA), and DNA was removed using Turbo DNA Free (Ambion, Grand Island, NY, USA). For qRT–PCR analysis, 2 μg of total RNA was reverse transcribed with Superscript III First Strand Synthesis System (Invitrogen) and complementary DNA was amplified using SYBRGreen PCR Master Mix (Applied Biosystems, Grand Island, NY, USA) and primers are listed in Supplementary Table S2. For transcript profiling, 1 μg of total RNA was processed for hybridization to U133A 2.0 oligonucleotide arrays for human genes (Affymetrix, Santa Clara, CA, USA). To prepare nuclear lysates, cells were suspended in lysis buffer (20 mM Tris pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl), followed by nuclear lysis buffer (20 mM Tris pH 7.9, 1.5 mM $MgCl_2$, 0.42 M NaCl, 0.2 mM EDTA) in the presence of protease and phosphatase inhibitors. Whole-cell extracts were prepared in parallel to assess equal inputs for immunoblotting. Proteins resolved by SDS–PAGE were transferred to nitrocellulose membranes and probed with GATA6 (Cell Signaling (Danvers, MA, USA), no. 4253), GATA4 (Millipore, no. AB4132) and Actin (Santa Cruz (Santa Cruz, CA, USA) no. 47778) Abs. Binding was detected by chemiluminescence after incubation with horseradish peroxidase-coupled secondary Ab (Santa Cruz, no. 2048). Immunohistochemistry was performed as described previously[42] using CDX2 (Biogenex, Fremont, CA, USA, no. 10M MU392A-UC) and GATA6 Ab.

GATA6 and GATA4 knockdown, proliferation and cell cycle analyses

Cells were infected with lentiviruses in the pLKO.1 vector (Open Biosystems, Pittsburgh, PA, USA) carrying GATA6 (5′-AGAACAGCGAGCTCAAGTATT-3′), GATA4 (5′-CCAGAGATTCTGCAACACGAA-3′) or a control shRNA not complementary to any human gene and known to lack cytotoxicity (NS, 5′-CCTAAGGTTAAGTCGCCCTCG-3′). Stable clones were selected in 2 μg/ml puromycin and GATA6 or GATA4 depletion assessed 10 days later by immunoblotting. Triplicate samples of 5000 cells maintained in puromycin-free medium for 48 h were seeded in 96-well plates, and proliferation was assessed using CellTiter 96 (Promega, Madison, WI, USA, G4000) at 570 nm absorbance using a Synergy-2 multi-detection plate reader (BioTek, Winooski, VT, USA). For cell cycle analysis, cells were fixed in 70% ethanol, incubated overnight at −20 °C, washed in phosphate-buffered saline, stained with 10 μg/ml propidium iodide for 30 min and analyzed on a FACScan instrument (Becton-Dickinson, Franklin Lakes, NJ, USA) at 488 nm.

Phospho-histone H3 analyses

Histones were extracted as described,[46] with substitution of 0.2 N HCl for 0.4 N H_2SO_4, neutralized with three volumes of Trizma base (pH 11), resolved by SDS–PAGE and immunoblotted as described above with Histone H3 (Cell Signaling, cat. no. 9715) and Phospho-histone H3 (Cell Signaling, 3377) Ab. shRNA-treated AGS cells were also seeded on glass cover slips, fixed, incubated sequentially with Phospho-histone H3 Ab and fluorescein isothiocyanate-conjugated goat anti-rabbit IgG (Jackson Immunoresearch, West Grove, PA, USA, 111–096–144) and mounted in medium containing DAPI (Vectashield, Burlingame, CA, USA, H-1200).

ChIP and ChIP-seq

ChIP and ChIP-seq were performed as described previously,[18] using 3 μg GATA6 Ab (Cell Signaling 4253S). We used Cistrome tools (www.cistrome.org/) to call and annotate peaks, generate wiggle files and conservation plots, identify enriched sequence motifs and linked genes, and compare data across ChIP-seq libraries.[47] Wiggle traces were generated using the Integrative Genome Viewer.[48] Data associated with this study are available in the GEO, GSE51936.

Data analysis

Integrative expression analysis was performed in the R statistical computing environment (http://cran.r-project.org). Hierarchical clustering of the 1000 most varying probe sets in mRNA expression across gastric cancer cell lines was generated with an Euclidean distance metric and complete linkage clustering. A distinct cluster containing the HUG1N and AGS cell lines corresponded to those that express high levels of *GATA6* (probe set 210002_at) and *GATA4* (probe set 205517_at) mRNAs, independent of the presence of GATA4 or GATA6 probe sets in the distance matrix calculation. Functional categories enriched among groups of genes were determined using the DAVID software (http://david.abcc.ncifcrf.gov/). PCA was applied to the following six dimensions of measurements across all genes: average FCs in gene expression in Caco2 and HUG1N cells following shRA-mediated GATA6 depletion (measured by dChip, http://biosun1.harvard.edu/complab/dchip), and Pearson correlations of *GATA6* mRNA levels with all mRNAs in primary stomach and colorectal cancers and cell lines. FCs in transcript levels have a positive or negative sign to denote increases or decreases; for PCA analysis we centered FC values at 0 by the transformation FC −> sign(FC) (abs(FC) − 1.0). PCA axes are ordered by decreasing amount of variance explained in the six-dimensional space,[30] and the first two principal axes were used to derive scores for each gene, by projection. PCA scores for each axis were converted into *z*-scores and added to yield a single ranking that integrated shRNA and synexpression data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by the Dana-Farber Cancer Institute-Novartis Drug Discovery Program and National Institutes of Health awards R01CA151993 (SO), R01HG4069 (XSL), K08CA134931 (AJB) and the SPORE program in GI cancers P50 CA127003. Data associated with this study are available in the GEO, GSE51936.

REFERENCES

1 Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 2005; **436**: 117–122.
2 Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007; **450**: 893–898.
3 Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG *et al.* SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* 2009; **41**: 1238–1242.
4 Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinanen R, Palmberg C *et al.* In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* 1995; **9**: 401–406.
5 Darnell Jr. JE. Transcription factors as targets for cancer therapy. *Nat Rev Cancer* 2002; **2**: 740–749.
6 Redell MS, Tweardy DJ. Targeting transcription factors for cancer therapy. *Curr Pharm Des* 2005; **11**: 2873–2887.
7 Dulak AM, Schumacher S, van Lieshout J, Imamura Y, Fox C, Shim B *et al.* Gastrointestinal adenocarcinomas of the esophagus, stomach and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res* 2012; **72**: 4383–4393.
8 Alvarez H, Opalinska J, Zhou L, Sohal D, Fazzari MJ, Yu Y *et al.* Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet* 2011; **7**: e1001356.
9 Lin L, Bass AJ, Lockwood WW, Wang Z, Silvers AL, Thomas DG *et al.* Activation of GATA binding protein 6 (GATA6) sustains oncogenic lineage-survival in esophageal adenocarcinoma. *Proc Natl Acad Sci USA* 2012; **109**: 4251–4256.
10 Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; **127**: 2893–2917.
11 Matsuda A, Machii R. Trends in stomach cancer mortality rates in Japan, USA, UK, France and Korea based on the WHO mortality database. *Jpn J Clin Oncol* 2012; **42**: 154.
12 Nakajima M, Sawada H, Yamada Y, Watanabe A, Tatsumi M, Yamashita J *et al.* The prognostic significance of amplification and overexpression of c-met and c-erb B-2 in human gastric carcinomas. *Cancer* 1999; **85**: 1894–1902.
13 Lennerz JK, Kwak EL, Ackerman A, Michael M, Fox SB, Bergethon K *et al.* MET amplification identifies a small and aggressive subgroup of esophagogastric adenocarcinoma with evidence of responsiveness to crizotinib. *J Clin Oncol* 2011; **29**: 4803–4810.
14 Okines A, Cunningham D, Chau I. Targeting the human EGFR family in esophagogastric cancer. *Nat Rev Clin Oncol* 2011; **8**: 492–503.
15 Laverriere AC, MacNeill C, Mueller C, Poelmann RE, Burch JB, Evans T. GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *J Biol Chem* 1994; **269**: 23177–23184.
16 Fang R, Olds LC, Sibley E. Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development. *Gene Expr Patterns* 2006; **6**: 426–432.
17 Beuling E, Baffour-Awuah NY, Stapleton KA, Aronson BE, Noah TK, Shroyer NF *et al.* GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice. *Gastroenterology* 2011; **140**: 1219–1229.
18 Verzi MP, Shin H, He HH, Sulahian R, Meyer CA, Montgomery RK *et al.* Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev Cell* 2010; **19**: 713–726.
19 Schmidt PH, Lee JR, Joshi V, Playford RJ, Poulsom R, Wright NA *et al.* Identification of a metaplastic cell lineage associated with human gastric adenocarcinoma. *Lab Invest* 1999; **79**: 639–646.
20 Fu B, Luo M, Lakkur S, Lucito R, Iacobuzio-Donahue CA. Frequent genomic copy number gain and overexpression of GATA-6 in pancreatic carcinoma. *Cancer Biol Ther* 2008; **7**: 1593–1601.
21 Kwei KA, Bashyam MD, Kao J, Ratheesh R, Reddy EC, Kim YH *et al.* Genomic profiling identifies GATA6 as a candidate oncogene amplified in pancreatobiliary cancer. *PLoS Genet* 2008; **4**: e1000081.
22 Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci USA* 2007; **104**: 20007–20012.
23 Holtzinger A, Evans T. Gata4 regulates the formation of multiple organs. *Development* 2005; **132**: 4005–4014.
24 Zhao R, Watt AJ, Battle MA, Li J, Bondow BJ, Duncan SA. Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev Biol* 2008; **317**: 614–619.
25 Cheung HW, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA *et al.* Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc Natl Acad Sci USA* 2011; **108**: 12372–12377.
26 Welch JJ, Watts JA, Vakoc CR, Yao Y, Wang H, Hardison RC *et al.* Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* 2004; **104**: 3136–3147.
27 Yu M, Riva L, Xie H, Schindler Y, Moran TB, Cheng Y *et al.* Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell* 2009; **36**: 682–695.
28 Jang JH, Shin KH, Park JG. Mutations in fibroblast growth factor receptor 2 and fibroblast growth factor receptor 3 genes associated with human gastric and colorectal cancers. *Cancer Res* 2001; **61**: 3541–3543.
29 Kunii K, Davis L, Gorenstein J, Hatch H, Yashiro M, Di Bacco A *et al.* FGFR2-amplified gastric cancer cell lines require FGFR2 and Erbb3 signaling for growth and survival. *Cancer Res* 2008; **68**: 2340–2348.
30 Abdi H, Williams LJ. Principal component analysis. *CompStat* 2010; **2**: 433–459.
31 Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA *et al.* Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. *Genes Dev* 2007; **21**: 385–390.
32 Eeckhoute J, Lupien M, Meyer CA, Verzi MP, Shivdasani RA, Liu XS *et al.* Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res* 2009; **19**: 372–380.
33 Tamura S, Wang XH, Maeda M, Futai M. Gastric DNA-binding proteins recognize upstream sequence motifs of parietal cell-specific genes. *Proc Natl Acad Sci USA* 1994; **91**: 4609 1993;90:10876-80.
34 Gao X, Sedgwick T, Shi YB, Evans T. Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. *Mol Cell Biol* 1998; **18**: 2901–2911.
35 Beck F, Chawengsaksophak K, Waring P, Playford RJ, Furness JB. Reprogramming of intestinal differentiation and intercalary regeneration in Cdx2 mutant mice. *Proc Natl Acad Sci USA* 1999; **96**: 7318–7323.
36 Verzi MP, Shin H, Ho LL, Liu XS, Shivdasani RA. Essential and redundant functions of caudal family proteins in activating adult intestinal genes. *Mol Cell Biol* 2011; **31**: 2026–2039.
37 Chen X, Leung SY, Yuen ST, Chu KM, Ji J, Li R *et al.* Variation in gene expression patterns in human gastric cancers. *Mol Biol Cell* 2003; **14**: 3208–3215.
38 Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 2007; **449**: 1003–1007.
39 Barker N, Huch M, Kujala P, van de Wetering M, Snippert HJ, van Es JH *et al.* Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units *in vitro*. *Cell Stem Cell* 2010; **6**: 25–36.
40 Heng JC, Feng B, Han J, Jiang J, Kraus P, Ng JH *et al.* The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* 2010; **6**: 167–174.
41 Kageyama R, Ohtsuka T, Kobayashi T. The Hes gene family: repressors and oscillators that orchestrate embryogenesis. *Development* 2007; **134**: 1243–1251.
42 Kim TH, Shivdasani RA. Notch signaling in stomach epithelial stem cell homeostasis. *J Exp Med* 2011; **208**: 677–688.
43 Ross J, Mavoungou L, Bresnick EH, Milot E. GATA-1 utilizes Ikaros and polycomb repressive complex 2 to suppress Hes1 and to promote erythropoiesis. *Mol Cell Biol* 2012; **32**: 3624–3638.
44 Guiu J, Shimizu R, D'Altri T, Fraser ST, Hatakeyama J, Bresnick EH *et al.* Hes repressors are essential regulators of hematopoietic stem cell development downstream of Notch signaling. *J Exp Med* 2013; **210**: 71–84.
45 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostat* 2003; **4**: 249–264.
46 Shechter D, Dormann HL, Allis CD, Hake SB. Extraction, purification and analysis of histones. *Nat Protoc* 2007; **2**: 1445–1457.
47 Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y *et al.* Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* 2011; **12**: R83.
48 Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G *et al.* Integrative genomics viewer. *Nat Biotechnol.* 2011; **29**: 24–26.

Supplementary Information accompanies this paper on the Oncogene website (http://www.nature.com/onc)