The FEBS Journal

FEBS PRESS science publishing by scientists

Transcriptional signature induced by a metastasis-promoting c-Src mutant in a human breast cell line

Felix Broecker[1,2,*], Christopher Hardt[1], Ralf Herwig[1], Bernd Timmermann[1], Martin Kerick[1,†], Andrea Wunderlich[1], Michal R. Schweiger[1,†], Lubor Borsig[3], Mathias Heikenwalder[4,5,6], Hans Lehrach[1,7,8] and Karin Moelling[1,2]

1 Max Planck Institute for Molecular Genetics, Berlin, Germany
2 University of Zurich, Switzerland
3 Institute of Physiology, Zurich Center for Integrative Human Physiology, University of Zurich, Switzerland
4 Institute of Virology, Technische Universität München, Germany
5 Institute of Virology, Helmholtz Zentrum Munich, Germany
6 Department Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
7 Dahlem Centre for Genome Research and Medical Systems Biology, Berlin, Germany
8 Alacris Theranostics GmbH, Berlin, Germany

Keywords
c-Src; metastasis; oncogene; PDZ domain; transcriptome

Correspondence
F. Broecker, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, D-14424 Potsdam, Germany
Fax: +49 331 567 9302
Tel: +49 331 567 9735
E-mail: felixbroecker@gmx.net

Present addresses
*Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, D-14424 Potsdam, Germany
†Cologne Center for Genomics, University of Cologne, Weyertal 115b, D-50931 Cologne, Germany

(Received 17 July 2015, revised 20 January 2016, accepted 23 February 2016)

doi:10.1111/febs.13694

Deletions at the C-terminus of the proto-oncogene protein c-Src kinase are found in the viral oncogene protein v-Src as well as in some advanced human colon cancers. They are associated with increased kinase activity and cellular invasiveness. Here, we analyzed the mRNA expression signature of a constitutively active C-terminal mutant of c-Src, c-Src (mt), in comparison with its wild-type protein, c-Src(wt), in the human non-transformed breast epithelial cell line MCF-10A. We demonstrated previously that the mutant altered migratory and metastatic properties. Genome-wide transcriptome analysis revealed that c-Src(mt) de-regulated the expression levels of approximately 430 mRNAs whose gene products are mainly involved in the cellular processes of migration and adhesion, apoptosis and protein synthesis. 82.9% of these genes have previously been linked to cellular migration, while the others play roles in RNA transport and splicing processes, for instance. Consistent with the transcriptome data, cells expressing c-Src(mt), but not those expressing c-Src(wt), showed the capacity to metastasize into the lungs of mice *in vivo*. The mRNA expression profile of c-Src(mt)-expressing cells shows significant overlap with that of various primary human tumor samples, possibly reflecting elevated Src activity in some cancerous cells. Expression of c-Src(mt) led to elevated migratory potential. We used this model system to analyze the transcriptional changes associated with an invasive cellular phenotype. These genes and pathways de-regulated by c-Src(mt) may provide suitable biomarkers or targets of therapeutic approaches for metastatic cells.

Database
This project was submitted to the National Center for Biotechnology Information BioProject under ID PRJNA288540. The Illumina RNA-Seq reads are available in the National Center for Biotechnology Information Sequence Read Archive under study ID SRP060008 with accession numbers SRS977414 for MCF-10A cells, SRS977717 for mock cells, SRS978053 for c-Src(wt) cells and SRS978046 for c-Src(mt) cells.

Introduction

The c-Src proto-oncogene is the cellular homolog of the viral oncogene v-Src of Rous sarcoma virus (RSV), which was the first identified retroviral oncogene protein. c-Src is a non-receptor tyrosine kinase

Abbreviations
ECM, extracellular matrix; eIF4E, eukaryotic translation initiation factor 4E; FAK, focal adhesion kinase; mTOR, mammalian target of rapamycin; PDZ, postsynaptic density 95/discs large/zonula occludens-1; RSEM, RNA-Seq by Expectation Maximization; qRT-PCR, quantitative real-time PCR; RSV, Rous sarcoma virus; SFK, Src family kinase; SH, Src homology; STAT, signal transducer and activator of transcription; TF, transcription factor.

that is involved in a wide range of cellular functions including migration, invasion and adhesion. Elevated Src activity promotes oncogenic transformation [1]. The *in vitro* transforming capacity of activated Src kinase was originally demonstrated in RSV-infected (and thereby v-Src-expressing) chicken cells [2]. Src kinase-induced transformation also occurs in mammalian cells, and has been demonstrated by ectopic expression of v-Src in mouse fibroblasts [3]. *In vitro* studies with the human non-transformed epithelial breast cell line MCF-10A [4] have shown that activation of Src kinase activity leads to transformation *in vitro*, and tumor formation in nude mice [5–7]. Despite its known role in promoting cellular transformation, many of the functions of Src at the molecular level remain unclear [1].

Known targets of Src kinase include focal adhesion proteins, adaptor proteins, cell-cycle regulators and transcription factors. Src family kinases (SFKs) have been shown to modulate cell–cell and cell–matrix interactions, and to promote the expression of matrix-degrading enzymes. The c-Src protein is regulated by reversible phosphorylation of tyrosine residues Y416 and Y527 (numbering according to v-Src) and by protein–protein interactions through its Src homology (SH) domains SH2 and SH3 [1]. Phosphorylation of Y527 inactivates the kinase activity of c-Src by inducing a compact autoinhibitory protein conformation, while its dephosphorylation leads to an open active conformation [8,9]. Y527 is phosphorylated by the non-receptor tyrosine kinases C-terminal Src kinase and C-terminal Src kinase-homologous kinase [10]. Dephosphorylation of Y527 is catalyzed by several protein tyrosine phosphatases such as protein tyrosine phosphatase α [11] and SH2 domain-containing phosphatases 1 [12] and 2 [13]. Y416 is an autophosphorylation site whose phosphorylation is required for optimal c-Src kinase activity [14]. Moreover, various c-Src-interacting proteins may modulate its kinase activity non-enzymatically, such as focal adhesion kinase (FAK) [15] and SH2 domain-containing phosphatase 2 [16], which bind to the SH2 and SH3 domains, respectively, leading to increased kinase activity.

A conserved C-terminal hydrophobic motif, GENL, is shared by a number of ubiquitously expressed SFKs, including Src, Yes and Fyn. In contrast, SFKs lacking this motif, such as Lyn, Lck, Hck and Blk, are mainly expressed in non-adherent hematopoietic cells [17]. Thus, the C-terminus of c-Src appears to be relevant for adherence, and its absence, for instance in the v-Src protein, appears to contribute to non-adherent phenotypes of cells. Therefore, we suspected a possible role for the absent C-terminus in migration and invasion. The C-terminal GENL motif of c-Src has been described as binding motif for PDZ (postsynaptic density 95/discs large/zonula occludens-1) domains, and the C-terminal hydrophobic leucine residue is crucial for this interaction [17]. By substituting this single amino acid by alanine, to create a protein herein referred to c-Src(mt), we demonstrated loss of binding to PDZ domains, which are characteristic of many tumor suppressor proteins in MCF-10A and other cell lines. c-Src-interacting PDZ proteins include the tumor suppressor AF-6 (Afadin), the Membrane Protein Palmitoylated 2 (MPP2), a homolog of the Drosophila tumor suppressor Discs large (Dlg) and LNX1 (Ligand-of-Numb protein X1) [17–20]. Disruption of these interactions by alteration of the C-terminal motif led to constitutive activation and elevated kinase activity compared with wild-type c-Src protein with the intact GENL sequence [17–20]. We previously noted that expression of c-Src(mt) changed the cellular morphology more than expression of c-Src(wt). Such changes included impaired cell polarization and disorganization of the cytoskeleton in MCF-10A cells, associated with increased cellular transformation, as characterized by *in vitro* assays [17,19]. Ectopic expression of c-Src(mt) in MCF-10A cells also led to increased colony formation in soft agar, and foci formation on normal cell culture plates, as well as elevated invasion of Matrigel test systems [17]. Moreover, c-Src (mt)-expressing cells showed disruption of spheroid cell growth in three-dimensional acinar cultures in extracellular matrix, impaired wound healing and elevated migration on basement membranes *in vitro* [19].

To investigate the underlying mRNA expression changes, we performed genome-wide transcription profiling on the cellular model used previously, i.e. MCF-10A cells expressing transfected c-Src(mt) or c-Src(wt) constructs [17–20]. This cellular model allows for determining the effects of the single C-terminal point mutation in the c-Src protein by comparison with otherwise genetically identically cell lines. We show that c-Src(mt), but not c-Src(wt), significantly de-regulated the mRNA expression levels of 435 genes, herein referred as c-Src(mt) effector genes. These genes were characterized in terms of pathways and biological functions, allowing detailed insights into the mRNA expression profile induced by c-Src(mt). The respective gene products are mainly involved in migration and adhesion, as expected, but also in apoptosis and protein synthesis. An increased migratory potential of c-Src(mt)-expressing cells was suggested by the mRNA expression profile, and corroborated using an *in vivo* mouse model of lung tissue invasion, in line with our previous *in vitro* findings [17–20]. Mining of the

published scientific literature revealed that approximately 17.1% of the c-Src(mt) effector genes have not previously been linked to cellular migration. These include genes that are mainly involved in RNA transport and splicing processes. Comparison of the mRNA expression profile of c-Src(mt)-expressing cells with that of primary human cancer samples revealed significant overlap with 56 of 112 tumor samples, indicating similarities between c-Src(mt)-expressing cells and some cancer cells.

Results

Establishment of a c-Src(mt) cellular model

To study the effects of c-Src(mt), we used the human non-transformed epithelial MCF-10A cell line stably expressing c-Src(mt) or c-Src(wt) as a control [17–19]. The c-Src(mt) protein is identical to c-Src(wt) except for a single L→A substitution in the C-terminal GENL sequence. This hydrophobic motif mediates binding to PDZ domain-containing tumor suppressor proteins, and is shared by the ubiquitously expressed SFKs Src, Yes and Fyn, but not by Fgr, Hck, Lck, Lyn and Blk, which are expressed in non-adherent hematopoietic cells (Fig. 1A). The c-Src protein contains an SH1 (kinase) domain, plus SH2 and SH3 domains (Fig. 1B). The experimental approach comprised generation of stable cell lines with matched protein expression levels, RNA isolation, Illumina RNA sequencing (RNA-Seq), mapping of generated reads to the human genome using Bowtie [21], quantification of transcript expression levels with the RNA-Seq by Expectation Maximization (RSEM) software [22], determination of significantly differentially expressed mRNAs using DEseq [23], and functional characterization of the respective gene products.

We stably transfected MCF-10A cells with the previously described constructs c-Src(wt) or c-Src(mt) [17–20], using lentiviral expression vectors (Fig. 1C). Cells transfected with an otherwise identical expression vector but with a randomized nonsense sequence instead of c-Src are referred to as mock cells. Inclusion of the tetracycline-inducible system allowed for regulating transgene expression by addition of tetracycline (Tet) to the cell culture medium. Expression of the transgene in c-Src(wt)- and c-Src(mt)-transfected cells was verified by quantitative real-time PCR (qRT-PCR) (using the primers listed in Table 1) and Western blotting (Fig. 1D). Only c-Src(mt)-expressing cells exhibited a transformed phenotype upon exposure to Tet, including a stretched morphology (Fig. 1E), but no increased growth (data not shown). This is in accordance with our previous findings that expression of c-Src(mt) alters cellular morphology, but does not affect proliferation of these cells [19]. The RNA preparations displayed characteristic rRNA bands and no signs of degradation (Fig. 1F). We selected RNA preparations obtained from cells exposed to 1 $\mu g \cdot mL^{-1}$ Tet for whole-genome mRNA expression profiling, as the transgenes c-Src(wt) and c-Src(mt) were expressed at similar levels to endogenous c-Src (Fig. 1D). Naive (non-transfected) MCF-10A cells as well as mock cells served as controls. RNA preparations of all four cell lines were subjected to RNA-Seq, resulting in 99 100 544, 46 988 060, 38 071 856 and 56 106 630 quality control-passed reads for naive MCF-10A, mock, c-Src(wt) and c-Src(mt) lines, respectively. Reads were mapped to the human genome using Bowtie [21], yielding 76 735 649 (77.4%), 37 101 761 (79.0%), 30 474 883 (80.0%) and 43 473 640 (77.5%) mapped reads. Gene expression levels were then quantified at the transcript level using RSEM [22].

To assess the global changes in mRNA expression induced through the action of c-Src(wt) and c-Src(mt), transcript levels of both cell lines were compared to those of naive MCF-10A cells using DEseq [23]. Comparison of mock versus naive MCF-10A cells served as a negative control. The resulting logarithmic fold change values (logFC) for significantly differentially expressed mRNAs ($q \leq 0.05$) were plotted as density graphs (Fig. 1G). The profiles of mock and c-Src(wt) cells were almost identical, whereby the differences to naive MCF-10A cells may reflect non-specific effects of the lentiviral transfection procedure. The profile of c-Src(mt)-expressing cells was different, and showed more pronounced mRNA expression changes in both directions (down- and up-regulated genes), higher q values (Fig. 1H) and a higher proportion of up-regulated genes (Fig. 1G). The similar density graphs of mock- and c-Src(wt)-expressing cells suggest that c-Src (wt) is incapable of inducing detectable Src-specific mRNA expression changes. This is in accordance with previous studies where overexpression of the wild-type c-Src protein in mammalian cell lines did not cause major phenotypic alterations [17–20,24].

To verify Src activity, we performed qRT-PCR expression analysis of six known Src targets: p53 (*TP53*) [25] and the related p63 (*TP63*), cyclin D2 (*CCND2*) [26] and mir-205 (*MIR205HG*) [27], all of which are suppressed by Src, as well as matrix metalloproteinase 2 (*MMP2*) [28] and hyaluronan synthase 2 (*HAS2*) [29] that have been shown to be upregulated (Fig. 1I). *GAPDH* served as a control, and, as expected, showed identical expression levels in all four cell lines. While the reported effects on all of the known Src target genes were verified at the mRNA

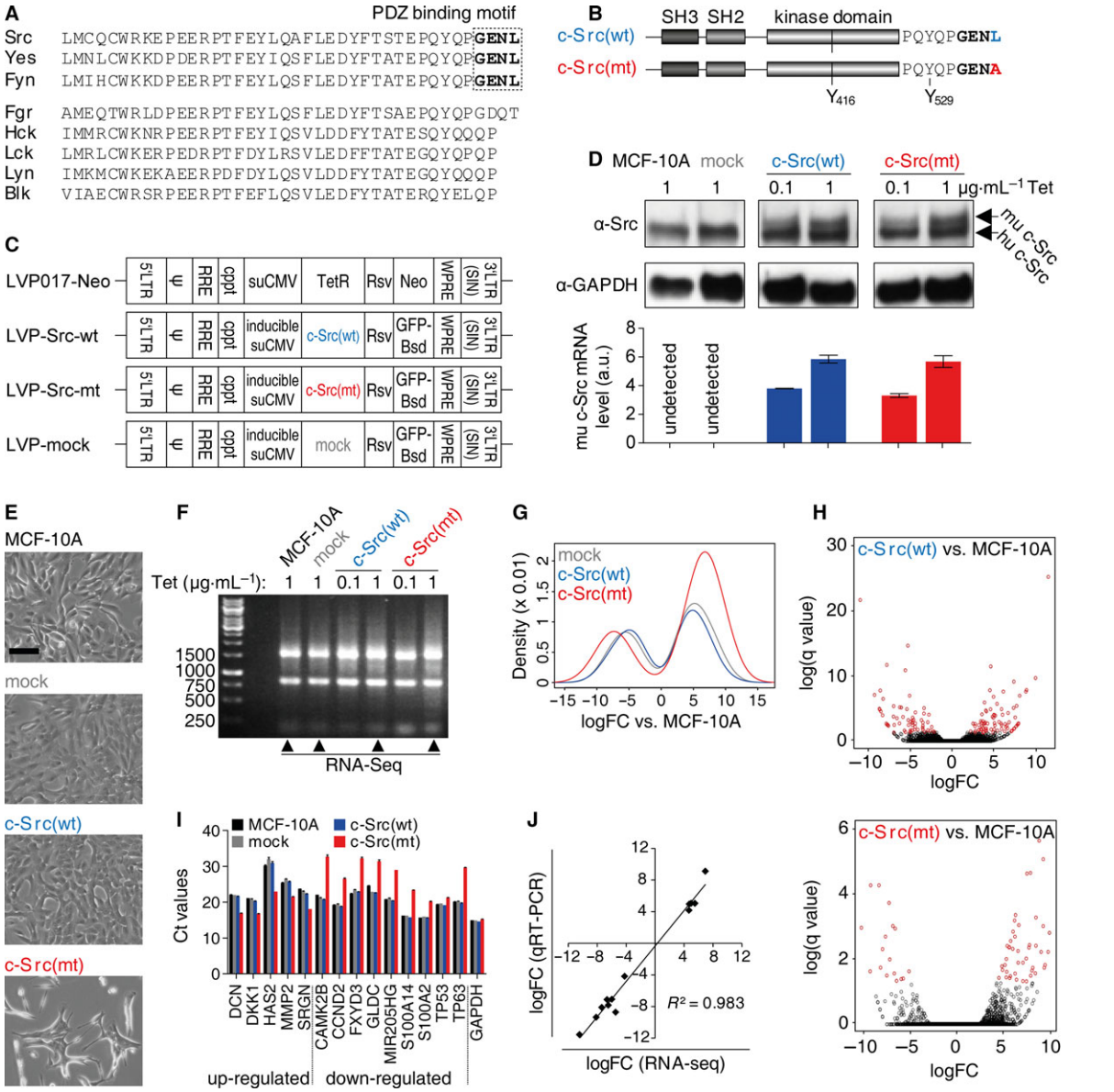


Fig. 1. Establishment of a cells stably expressing wildtype or mutant c-Src. (A) Alignment of C-terminal moieties of human SFKs. The conserved binding motif for PDZ domain-containing tumor suppressor proteins is highlighted. (B) Domain organization of c-Src constructs used in this study. (C) Lentiviral vectors used in this study. The Neo gene confers resistance to G418. The Bsd gene, which is fused to the GFP gene, confers resistance to Blasticidin S. LTR, long terminal repeat; Ψ, psi packaging sequence; RRE, Rev-responsive element; cppt, central polypurine tract; suCMV, 'super' cytomegalovirus promoter; TetR, tetracycline regulatory gene; Rsv, Rous sarcoma virus promoter; WPRE, woodchuck hepatitis virus post-transcriptional regulatory element; SIN, self-inactivating. The suCMV promoter is tetracycline-inducible. The 'mock' construct includes a randomized sequence of approximately 100 nucleotides. (D) Detection of transgene expression in engineered MCF-10A cells by Western blotting (upper panel) and qRT-PCR (lower panel). The α-Src antibody binds to both human (hu) and murine (mu) c-Src proteins. The murine c-Src constructs (61.3 kDa) are bigger than the endogenous human c-Src (59.8 kDa), giving rise to a second band above that of human c-Src in the c-Src(wt) and c-Src(mt)-expressing cells. GAPDH is 36 kDa. The qRT-PCR data are means ± SD of two independent measurements. (E) Inverted light microscopic images of the four cell lines after exposure to 1 μg·mL^{-1} Tet for 24 h. Scale bar = 100 μm. (F) Agarose gel of RNA preparations used for RNA-Seq and qRT-PCR. The sizes of the markers in base pairs are indicated on the left. (G) Density plot showing logarithmic fold change (logFC) values for statistically significantly ($q \leq 0.05$) differentially expressed genes for mock, c-Src(wt) and c-Src(mt) lines compared with naive MCF-10A cells, obtained from RNA-Seq. (H) Volcano plots. Black circles, $q > 0.05$; red circles, $q \leq 0.05$. (I) qRT-PCR analysis of c-Src(mt) effector genes. Values are means and SD of two independent measurements. Most SD values are too small to be visible. (J) Correlation between RNA sequencing data and qRT-PCR measurements for 13 statistically significantly differentially expressed genes for the comparison c-Src(mt) versus mock cells.

Table 1. Primers used in this study.

Gene	Forward primer (5′→3′)	Reverse primer (5′→3′)
TetR	TCAAGTCGCTAAAGAAGAAAG	TCAAGGCCGAATAAGAAG
mu *SRC*_HA	CGCGGGTACCCATACGAC	ACGGTGTCCGAGGAGTTG
CAMK2B	CCACACGACCATCCTGAAC	GGCTCCAAACACCAACTCTG
CCND2	AACTGGAAGTGTGGGAGCAG	GGCAAGCTTTGAGACAATCC
DCN	TGCTGTTGACAATGGCTCTC	ACCGGGTTGCTGAAAAGAC
DKK1	TTTCCGAGGAGAAATTGAGG	TGATGACCGGAGACAAACAG
FXYD3	GCTCTGACATGCAGAAGGTG	GCTCTCACCATAGTAGAAAGGACTG
GAPDH	GTTCCAATATGATTCCACCC	GAAGATGGTGATGGGATTTC
GLDC	TGCGTTACATCTTCCCAC	AGACGCCCTCTTTTGTTC
HAS2	AGAAGATCCCATGGTTGGAG	GGTCCACTAATGCACTGAACAC
MIR205HG	CCATCTTGGAGGGTACGG	CACATTTCTCTCTGGCTG
MMP2	CAATGAGGTGAAGAAGAAAATGG	TATCGAAGGCAGTGGAGAGG
S100A14	GGCCATTGAGACCCTCATC	GCCTCTCCAGCTTCACACTC
S100A2	AAGAGGGCGACAAGTTCAAG	ATCCATGGCAGGAAGTCAAG
SRGN	AGTAATTCTGCAAACTGCCTTG	CCTGTTCCATTTCCGTTAGG
TP63	ACGAAGATCCCCAGATGATG	GAAGTAAGTGCTGGTGCTGCT

level in the c-Src(mt)-expressing cells, cycle threshold values obtained by qRT-PCR in the c-Src(wt)-expressing cells were similar to those of the naive MCF-10A and mock cells. This gave further support to the notion that expression of c-Src(wt) alone is insufficient to induce de-regulation of Src targets in our cellular model, but expression of c-Src(mt) did alter the transcript levels of these genes. This is in accordance with previous findings showing that either activating mutations [17–20,24] or co-factors [1] are required for increased Src activity. This may be because c-Src(wt) is maintained in an inactivated state through binding of PDZ domain-containing tumor suppressor proteins to its C-terminal GENL sequence (Fig. 1A). In contrast, c-Src(mt), due to mutation of this motif, shows impaired binding to PDZ domains, rendering this mutant constitutively active [19].

We next investigated effector mRNAs regulated by c-Src(mt). To account for possible Src-independent gene expression changes resulting from lentiviral transfection, we used comparison with mock cells instead of naive MCF-10A cells as a more stringent control. This comparison, c-Src(mt) versus mock, yielded a total of 435 significantly differentially expressed mRNAs (Table S1), whose respective genes are referred to as c-Src(mt) effector genes. This gene list contained all differentially expressed transcripts with $q \leq 0.05$, irrespective of the fold change value. Of these, 51.5% (224 genes) were up-regulated in c-Src(mt)-expressing cells at the transcript level.

All known c-Src(mt) effector gene transcripts analyzed above by qRT-PCR also showed significant differential expression levels between c-Src(mt) and mock cells by RNA-Seq, except for *TP53*, probably because the corresponding expression levels were too similar (Table S1 and Fig. 1I). mRNAs of three genes mentioned above, *TP63*, *CCND2* and *MIR205HG,* exhibited negative logFC values, whereas those of *MMP2* and *HAS2* showed positive logFC values with statistical significance, confirming our qRT-PCR results as well as previous findings [26–29]. As expected, *GAPDH* showed no altered mRNA expression, as also seen at the protein level (Fig. 1D). Interestingly, the expression level of endogenous c-Src was not altered at the mRNA level (RNA-Seq) or at the protein level (Fig. 1D). Therefore, any effects on mRNA expression levels were due to the activity of ectopically expressed c-Src(mt) and not endogenous c-Src, which was expressed at identical levels in all four cell lines. To further verify the RNA-Seq data, we performed qRT-PCR analysis for eight additional genes with significant mRNA de-regulation in the comparison of c-Src(mt) versus mock cells (Fig. 1I). logFC values were determined using the $2^{-\Delta\Delta CT}$ method [30] and plotted against the logFC values obtained from RNA-Seq (Fig. 1J). The strong correlation ($R^2 = 0.983$) verified the transcript levels inferred by RNA-Seq.

In summary, the c-Src(mt)-expressing cells exhibited alterations in mRNA expression levels that have previously been attributed to increased Src activity, and RNA-Seq data correlated well with the results of qRT-PCR expression analysis. The comparison of c-Src(mt) versus mock cells provided a comprehensive list of 435 c-Src(mt) effector genes for further characterization.

c-Src(mt) de-regulates genes involved in migration, apoptosis and protein synthesis

To obtain further insights into the mRNA expression program mediated by c-Src(mt), we subjected the set

of c-Src(mt) effector genes to bioinformatic analyses. First, characteristics of gene products were assessed by Gene Ontology (GO) category over-representation analysis [31]. Table S2 provides a complete list of GO terms for this analysis and the corresponding gene lists. Analysis of gene products in the GO category 'biological process' (Fig. 2A) showed that c-Src(mt) effector genes play a prominent role in organelle organization (GO:0006996, $q = 6.66 \times 10^{-6}$) (Fig. 2A), indicating changes in cellular morphology. This is in line with observations that c-Src(mt) promotes an invasive phenotype in epithelial cells and impaired cell polarization and cytoskeletal organization [17–20]. Moreover, expression of c-Src(mt) significantly altered the mRNA expression levels of genes involved in cellular response to cytokine stimulation (GO:0034097, $q = 1.31 \times 10^{-3}$). It has been noted that the cytokine profile in tumor microenvironments plays an important role in cellular migration [32]. For instance, c-Src (mt)-expressing cells over-express *WNT5A* mRNA ($q = 9.28 \times 10^{-3}$), whose gene product is a ligand of Frizzled receptors. Wnt/Frizzled signaling is known to be involved in cellular migration [33]. Transcripts of apoptosis-associated genes were also de-regulated in c-Src(mt)-expressing cells (GO:0006915, $q = 1.31 \times 10^{-3}$). Evasion of apoptosis is recognized as a hallmark of cancer, and Src is known to play a key role in this process [1]. Gene products of c-Src(mt) effector genes are mainly localized in the cytosol and nucleus (Fig. 2A, 'cellular component'), perhaps reflecting a predominant involvement in cytosolic signaling and gene regulation processes. An enrichment of proteins binding to structural proteins such as actin (GO:0003779, $q = 5.33 \times 10^{-4}$) suggests that c-Src(mt) effector genes are involved in cytoskeleton remodeling and migration (Fig. 2A, 'molecular function').

To further characterize the roles of c-Src(mt) effector genes, we performed pathway enrichment analysis using ConsensusPathDB [34]. We found that c-Src(mt) effector genes mainly affect pathways involved in cell migration and adhesion (Fig. 2B, yellow), including hemidesmosome assembly and integrin pathways, as well as protein synthesis (Fig. 2B, blue) and apoptosis (Fig. 2B, red).

c-Src(mt) effector genes involved in cellular migration and adhesion included the integrin receptor component genes *ITGA2*, *ITGA6* and *ITGAV*, all of which were transcriptionally down-regulated (Fig. 2C). Integrin receptors are known to attach cells to the extracellular matrix (ECM) by binding to ECM components such as laminin, collagen, fibronectin and E-cadherin, and their de-regulation promotes invasiveness. For instance, *ITGAV* expression has been linked with metastatic activity in colorectal cancer [35], and *ITGA2* genetic variants may be associated with invasiveness of gastric cancer [36]. In addition, a number of components of the ECM itself were de-regulated at the mRNA level by the action of c-Src(mt). These include decorin (*DCN*), fibronectin 1 (*FN1*), fibrillin 1 (*FBN1*), which are up-regulated, and E-cadherin (*CDH1*), which is down-regulated. Various mRNAs encoding collagen chain components were either up- (*COL8A1*, *COL12A1* and *COL28A1*) or down-regulated (*COL17A1*). Of note, over-expression of decorin [37], fibronectin 1 [38] and collagen chains COL8A1 [39] and COL12A1 [40] has recently been shown to promote invasiveness of various cancer types. These results indicate that c-Src(mt) significantly alters cell–matrix interactions, thereby decreasing cell adhesion and promoting cellular migration and invasiveness. Moreover, MMP2, an enzyme that degrades type IV collagen, the major constituent of basement membranes, was transcriptionally up-regulated by c-Src(mt). Increased expression of MMP2 has been shown to correlate with cellular invasion [1]. Except for the collagen chain COL8A1 and fibronectin 1, none of the above mentioned genes involved in cell adhesion were de-regulated by c-Src(wt) at the mRNA level (Fig. 2C), further confirming that the activating mutation of c-Src(mt) promotes Src activity.

c-Src(mt) effector genes involved in apoptosis signaling include those involved in maintaining cytoskeleton integrity, such as adducin 1 (*ADD1*) and plectin (*PLEC*), both of which were transcriptionally up-regulated by c-Src(mt) (Fig. 2C). Adducins are components that are important for the cortical cytoskeleton network, and are subject to caspase-3-mediated cleavage during cisplatin-mediated apoptosis [41]. Plectin serves as a cross-linking molecule for the cytoplasmic filament system, and is thus involved in maintaining cytoskeleton integrity, and its cleavage by caspase-8 occurs early during tumor necrosis factor receptor-mediated apoptosis [42]. Up-regulation of cytoskeletal components such as adducin 1 and plectin by c-Src(mt) may thus interfere with caspase-mediated cytoskeleton instability, thereby suppressing apoptosis. While *ADD1* transcripts were not de-regulated by c-Src(wt), *PLEC* mRNA was up-regulated, but less markedly than by c-Src(mt) (Fig. 2C). We verified the expression of selected ECM and cytoskeletal proteins by immunohistochemistry (see below).

While the involvement of Src in cell adhesion and apoptosis is well-known [1], we also observed over-representation of c-Src(mt) effector genes in a surprisingly large number of pathways involved in translation and protein synthesis (Fig. 2B). These include various steps of protein synthesis, such as mRNA transport

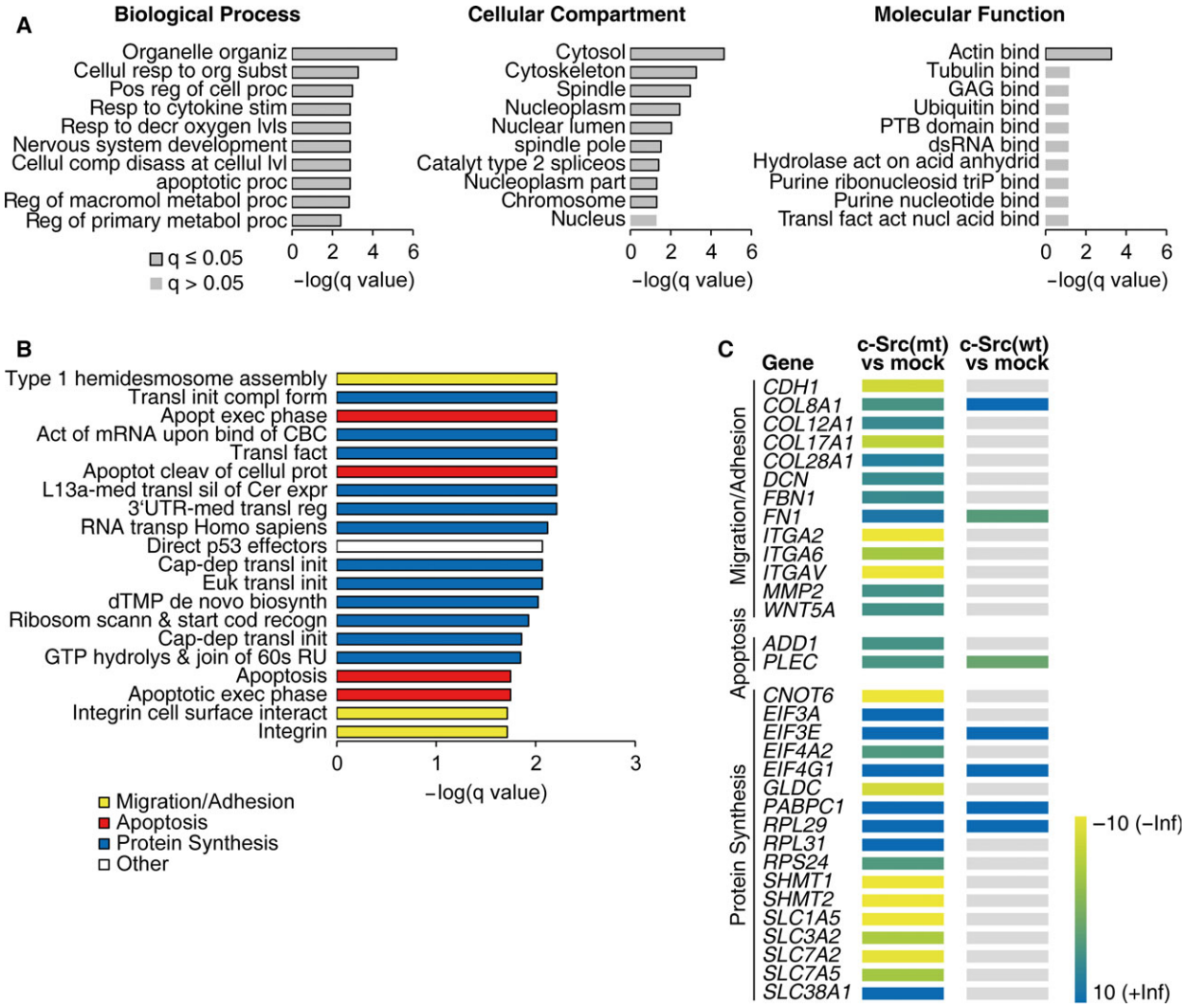


Fig. 2. Functional analysis of c-Src(mt) effector genes. (A) Gene Ontology (GO) enrichment analysis. The ten level 4 GO terms in the indicated categories with the lowest q values are shown. Table S2 provides a complete list of terms and corresponding gene lists. (B) Pathway analysis. The 20 pathways with the lowest q-values (all ≤ 0.05) are shown. Redundant pathways have been removed. The color code indicates pathways associated with different cellular processes. Table S3 provides the complete list of pathways. (C) Selected genes and their relative expression levels in c-Src(mt)- and c-Src(wt)-expressing cells compared with mock cells. The logFC values ($q \leq 0.05$) are indicated by a color code. Gray bars represent genes with a non-significant difference in expression compared with mock cells ($q > 0.05$).

and degradation, and translation initiation, as well as amino acid biosynthesis and transport (Table S3). A number of transcripts of eukaryotic translation initiation factors (EIFs), including *EIF3A*, *EIF3E*, *EIF4A2* and *EIF4G1*, and other proteins involved in translation initiation were up-regulated by c-Src(mt) (Fig. 2C). mRNAs for the small and large subunit ribosomal proteins RPL29, RPL31 and RPS24 were also up-regulated. Amino acid transport across the cell membrane may have been influenced by c-Src(mt), as a number of mRNAs for genes encoding amino acid-transporting solute carrier proteins, such as *SLC1A5*, *SLC3A2*, *SLC7A2*, *SLC7A5* and *SLC38A1*, were de-regulated. In addition, mRNAs of amino acid biosynthesis genes were de-regulated by c-Src(mt), including down-regulation of transcripts of serine hydroxymethyltransferases 1 and 2 (*SHMT1/2*), enzymes that catalyze the conversion of serine to glycine, and down-regulation of the mRNA for the glycine dehydrogenase (decarboxylating) gene *GLDC*, whose gene product degrades glycine. We hypothesize that c-Src(mt) stimulates global protein synthesis by increasing the number of ribosomes and the intracellular levels of translation initiation factors and free

amino acids. Moreover, genes involved in mRNA decay, such as the *CNOT6* cytoplasmic deadenylase, were transcriptionally down-regulated by c-Src(mt), while the polyadenylate-binding protein 1 (*PABPC1*) gene, which is essential for translation initiation, was up-regulated at the mRNA level. Moreover, the overall higher proportion of up-regulated transcripts compared to down-regulated transcripts in c-Src(mt)-expressing cells indicated an elevated pool of available mRNAs, further increasing protein expression, including that of pro-invasion and anti-apoptosis genes (Fig. 2A–C). Only a few of the above-mentioned mRNAs involved in protein synthesis were de-regulated by c-Src(wt) (Fig. 2C). We conclude that activation of protein synthesis-related transcripts required the constitutively active form of c-Src.

In summary, over-expression of c-Src(mt) was sufficient to modulate transcription of genes involved in cell migration/adhesion, apoptosis and protein synthesis. In particular, de-regulation of mRNAs whose gene products are involved in integrin signaling, ECM remodeling, composition of the cytoskeleton, and the translation machinery was observed.

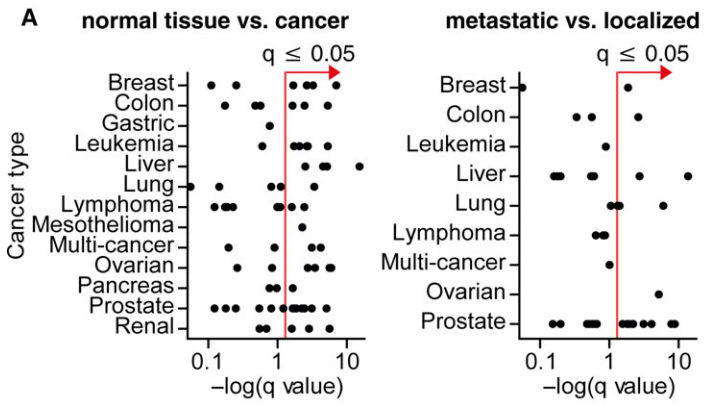
Non-migration-associated c-Src(mt) effector genes are involved in mRNA splicing

The role of activated c-Src in promoting metastasis is well-known [1]. However, the transcriptome data presented here also revealed a prominent effect of c-Src (mt) expression on protein synthesis (Fig. 2B). To identify c-Src(mt) effectors that are not associated with metastasis, we mined PubMed abstracts, using a defined lexicon of terms associated with migration. The majority of c-Src(mt) effector genes, 415 of 435 (95.4%), were referred to in PubMed abstracts. Of these, 344 (82.9%) were linked to migration, which may reflect that previous studies investigating Src activity were biased towards assessment of its migration-promoting activity. On the other hand, 71 genes (17.1%) were not linked to migration (highlighted in green in Table S1). These 71 genes are of particular interest as they may include genes that have not previously been associated with Src activity. To obtain an insight into the role of these 71 genes, they were subjected to Gene Ontology (GO) category over-representation analysis. Interestingly, these genes appeared to be mainly involved in splicing of pre-mRNAs (GO:0005689, 'U12-type spliceosomal complex'; GO:0071013, 'catalytic step 2 spliceosome', both $q = 3.62 \times 10^{-2}$) (Table S4). De-regulated splicing is a known feature of transformed cells [43]. Our data so far suggests that the constitutively active c-Src(mt) mutant, in addition to promoting metastasis, may have an effect on protein synthesis and splicing. To our knowledge, these functions have not previously been attributed to Src.

c-Src(mt)-expressing cells exhibit characteristics of primary human tumors and progression to metastasis

Increased Src activity is a frequent feature of human cancers [1]. Therefore, we determined whether c-Src (mt)-expressing cells show similarities with primary human cancer cells. To analyze whether c-Src(mt)-expressing cells share similar mRNA expression profiles with primary human cancers, we used the ONCOMINE cancer microarray database [44] for comparison with c-Src(mt) effector genes. The gene sets in ONCOMINE consist of genes whose transcripts were differentially expressed when comparing either normal tissue versus cancer or metastatic versus localized tumors. Of the 70 ONCOMINE gene sets for normal tissue versus cancer, 38 (54.3%) showed significant overlap with c-Src(mt) effector genes ($q \leq 0.05$) (Fig. 3A, left). This was the case for most breast cancer (4/6), leukemia (5/6), ovarian (5/7), prostate (7/13) and renal cancer samples (3/5), as well as all liver cancer samples (4/4). Increased Src activity has been identified in breast, ovarian and liver cancers and leukemia [45], as well as prostate cancer [46] and renal cancer [47]. Analysis of ONCOMINE datasets for cancer progression (metastatic versus localized, $n = 42$) revealed that 18 (42.9%) showed significant overlap with c-Src(mt) effector genes (Fig. 3A, right), including most of the lung cancer samples (4/5) and the prostate cancer samples (9/16). Increased Src activity has been shown to be associated with increased metastasis in lung [45] and prostate cancer [48]. Therefore, c-Src (mt)-expressing cells are a valid, albeit simplified, model for some human cancers and their progression to metastasis. These results suggest that the single amino acid mutation in c-Src(mt) does indeed affect similar transcripts to those that are found to be dys-regulated in diverse tumor or metastatic cells.

We next sought to investigate whether the non-migration-associated c-Src(mt) effector genes (Table S1) are associated with human cancer. Therefore, we searched the ONCOMINE database for any evidence of their de-regulation in normal tissue versus cancer pairs. We identified 18 genes that showed significant transcriptional up-regulation (eight genes) or down-regulation (ten genes) ($q \leq 0.05$) in at least half of the analyzed datasets of the respective cancer



B

Cancer	Gene	Over-expressed	Under-expressed
Leukemia	*CDC40*	-	5 of 6
	MARK3	-	4 of 6
	PSRC1	-	2 of 3
	SMCHD1	-	4 of 6
	TTC7A	3 of 3	-
	ZNF655	3 of 3	-
Liver	*SHMT2*	-	4 of 4
Lung	*UACA*	-	2 of 3
Lymphoma	*IQCE*	6 of 8	-
Multicancer	*ANKS1B*	2 of 2	-
	KIAA0664	-	2 of 2
	THSD4	-	2 of 2
	ZNF451	2 of 2	-
Ovarian	*ASXL2*	4 of 5	-
	CDC40	-	4 of 5
	DHX15	-	4 of 5
	TNPO2	3 of 5	-
	YTHDF2	4 of 5	-

Fig. 3. ONCOMINE analysis of c-Src(mt) effector genes. (A) *q* values of overlaps in gene expression between c-Src(mt) effector genes and normal tissue versus cancer pairs (left), and between c-Src(mt) effector genes and metastatic versus localized tumor pairs (right). Table S5 provides a complete list of primary cancer samples and gene lists. (B) Non-migration-associated c-Src(mt) effector genes with significant de-regulation in normal tissue versus cancer pairs.

types (Fig. 3B). All of these genes showed similar behavior in the c-Src(mt) cells (over- or under-expressed with $q \leq 0.05$ compared to mock cells). Of note, most of these genes were de-regulated in leukemia samples. Thus, the non-migration-associated c-Src (mt) effector genes are probably most relevant for this cancer type.

Discussion

Here we have analyzed a cellular model of C-terminally mutated constitutively active Src kinase that was obtained by stable expression of c-Src(mt) at physiological levels in non-transformed human cells. The c-Src (mt)-specific transcriptional changes inferred by RNA sequencing were in good agreement with qRT-PCR measurements (Fig. 1I,J). The substantial changes in cellular morphology induced by expression of c-Src (mt) (Fig. 1E) indicated alterations in the expression levels of functionally active proteins. Despite these findings, mRNA expression changes may not necessarily be reflected at the protein level for every identified gene. However, recent studies demonstrated good correlations between mRNA and protein levels in differential gene expression analyses, suggesting that quantifying mRNA may be more useful than previously acknowledged [49,50]. It should be noted that mRNA turnover may be influenced by c-Src(mt), as suggested by altered transcription of genes involved in mRNA decay (e.g. *CNOT6*) and translation initiation (e.g. *PABPC1*). Altered turnover, in addition to mRNA levels *per se*, may further influence protein expression of c-Src(mt) effector genes.

The expression of c-Src(mt), but not of c-Src(wt) at comparable levels (Fig. 1D), altered the mRNA levels of previously reported Src targets (Fig. 1I). This is in accordance with previous findings that activation of c-Src, either by activating mutations or deletions in the C-terminus [17–20,24], or via the action of co-factors, such as FAK [1], is required for its transforming activity. Here, whole-genome transcriptome analysis revealed that most of the 435 genes de-regulated by c-Src(mt) at the mRNA level are implicated in migration and adhesion, apoptosis and protein synthesis (Fig. 2A–C).

Elevated Src activity frequently correlates with increased tissue invasion [45]. The change in mRNA levels of migration/adhesion-associated genes in the c-Src(mt)-expressing cells (Fig. 2B,C) suggested that these cells may also show increased metastatic potential. To test whether c-Src(mt)-expressing cells exhibit increased capacity to metastasize *in vivo*, we used an established mouse model of metastatic lung tissue invasion [51]. Eight-week-old SCID mice were intravenously injected with 9×10^5 naive MCF-10A, c-Src(wt), c-Src (mt) or mock cell lines. To induce transgene expression in the engineered cell lines, 1 $\mu g \cdot mL^{-1}$ Tet was added to the drinking water of all animals. Mice were sacrificed after 42 days, and the phenotypes of lung tissues were analyzed (Fig. 4A). Macroscopic analysis revealed visible lung metastases only in those mice that received c-Src(mt)-expressing cells, indicating that expression of c-Src(mt) predisposes tumor cells to extravasate and migrate into epithelial tissue. Lung metastases were found in two of five mice injected intravenously with c-Src(mt)-expressing cells, but were absent in the lungs of

mice that received c-Src(wt)-expressing cells (0/4), mock cells or naive MCF-10A cells (both 0/5) (Fig. 4B). Representative histological haematoxylin/eosin-stained sections are shown in Fig. 4C. The difference in metastasis formation in c-Src(mt)-expressing cells compared with mock-expressing cells was not statistically significant ($P = 0.1$). More extensive studies and larger sample sizes are required to corroborate this initial data. The observed trend *in vivo* is in accordance with previous findings showing that this c-Src mutant promoted motility in *in vitro* migration assays [19]. In contrast, over-expression of the wild-type c-Src protein alone has been shown to cause only minimal changes in migratory activity [19,29,54]. Accordingly, we observed no metastases in mice that received c-Src(wt)-expressing cells (Fig. 4B,C).

The lung metastases were subjected to immunohistochemical staining to qualtitatively verify the *in vivo* expression of c-Src(mt) effector genes at the protein level. We used four proteins whose mRNAs were significantly up-regulated by c-Src(mt); the ECM components fibronectin 1 and matrix metalloproteinase 2 (*FN1* and *MMP2* genes), and the cytoskeletal proteins adducin 1 and plectin (*ADD1* and *PLEC* genes) (Fig. 5). In addition, we investigated the expression of Ser235/265-phosphorylated S6 ribosomal protein, a marker of elevated protein synthesis [55]. The results confirmed that adducin 1, MMP2, plectin and Ser235/265-phosphorylated S6 ribosomal protein were expressed in the metastatic cells. Expression of fibronectin 1 appeared to be restricted to the border of the metastases.

Literature mining revealed that more than 82.9% of c-Src(mt) effector genes have previously been linked to cellular migration. The remaining 17.1% of genes showed an enrichment of genes involved in mRNA splicing (Table S4), a post-transcriptional process required for translation. Interestingly, various cellular pathways relevant for protein synthesis showed enrichment of genes that were transcriptionally de-regulated by c-Src(mt) (Fig. 2B). Moreover, immunohistochemistry revealed the expression of phosphorylated S6 ribosomal protein, a marker of increased translation [55], in c-Src(mt)-expressing cells (Fig. 5). To our knowledge, de-regulation of protein synthesis has not yet been linked with Src activity, perhaps because most previous studies mainly focused on investigating its metastasis-promoting potential.

Therefore, we suggest that Src activity is associated with an increase in global protein synthesis. Interestingly, this function has recently been attributed to another proto-oncogene product, c-Myc [56]. The importance of protein synthesis in cancer development has been neglected historically [56]. However, recent evidence underlines the importance of de-regulated translation in various cellular processes, including cancer progression. Using ribosomal profiling, Hsieh *et al.* [57] have shown that oncogenic signaling of the mammalian target of rapamycin (mTOR) kinase heavily influences the translational machinery, and thereby cancer invasion and metastasis, through increased translation of pro-invasion mRNAs. Consequently, components of the translational machinery, such as the translation initiation factor eIF4E, are now being explored as therapeutic targets to treat cancer [56]. In line with these findings, we found that Src activity caused de-regulation of various steps of translational control. Proteins whose mRNAs were up-regulated by c-Src(mt) included translation initiation factors, ribosomal proteins and proteins implicated in mRNA stability, thereby potentially promoting the 'cancerous' translation machinery. Interestingly, Src is known to phosphorylate and hence activate phosphoinositide 3-kinase, whose downstream targets include mTOR [56]. mTOR stimulates protein synthesis by phosphorylating eIF4E-binding protein 1 and ribosomal protein S6 kinase 1/2 (S6K1/2). We therefore propose a dual mechanism of translational activation by Src, through cytosolic activation of the mTOR pathway and transcriptional activation of components of the translational machinery, as shown in this study. This may lead to increased translation of pro-invasive and anti-apoptotic mRNAs, thus potentially promoting cellular migration.

A potential target for cancer therapies is FAK, a cytoplasmic tyrosine kinase that is a key downstream signal transducer of integrin receptors [58,59]. FAK is directly phosphorylated by Src kinase, leading to activation of multiple intracellular signaling pathways that are relevant for apoptosis and cell migration, amongst others [59]. FAK and Src kinase synergize in promoting cell migration and invasion [60,61]. We have recently shown that the expression of c-Src(mt) leads to increased levels of phosphorylated FAK in MCF-10A cells, concomitant with elevated cell motility [19]. The transcriptome analysis in the present study revealed that the mRNA expression level of FAK was not significantly influenced by c-Src(mt), as FAK (*PTK2*, protein tyrosine kinase 2) was not among the c-Src(mt) effector genes (Table S1). Together with our previous findings [19], the results here indicate that FAK activation by c-Src(mt) mainly involves its Src-dependent phosphorylation but transcription levels remain unaltered. Of note, both up-regulation and increased phosphorylation of FAK have been detected in numerous human metastatic tumors compared to

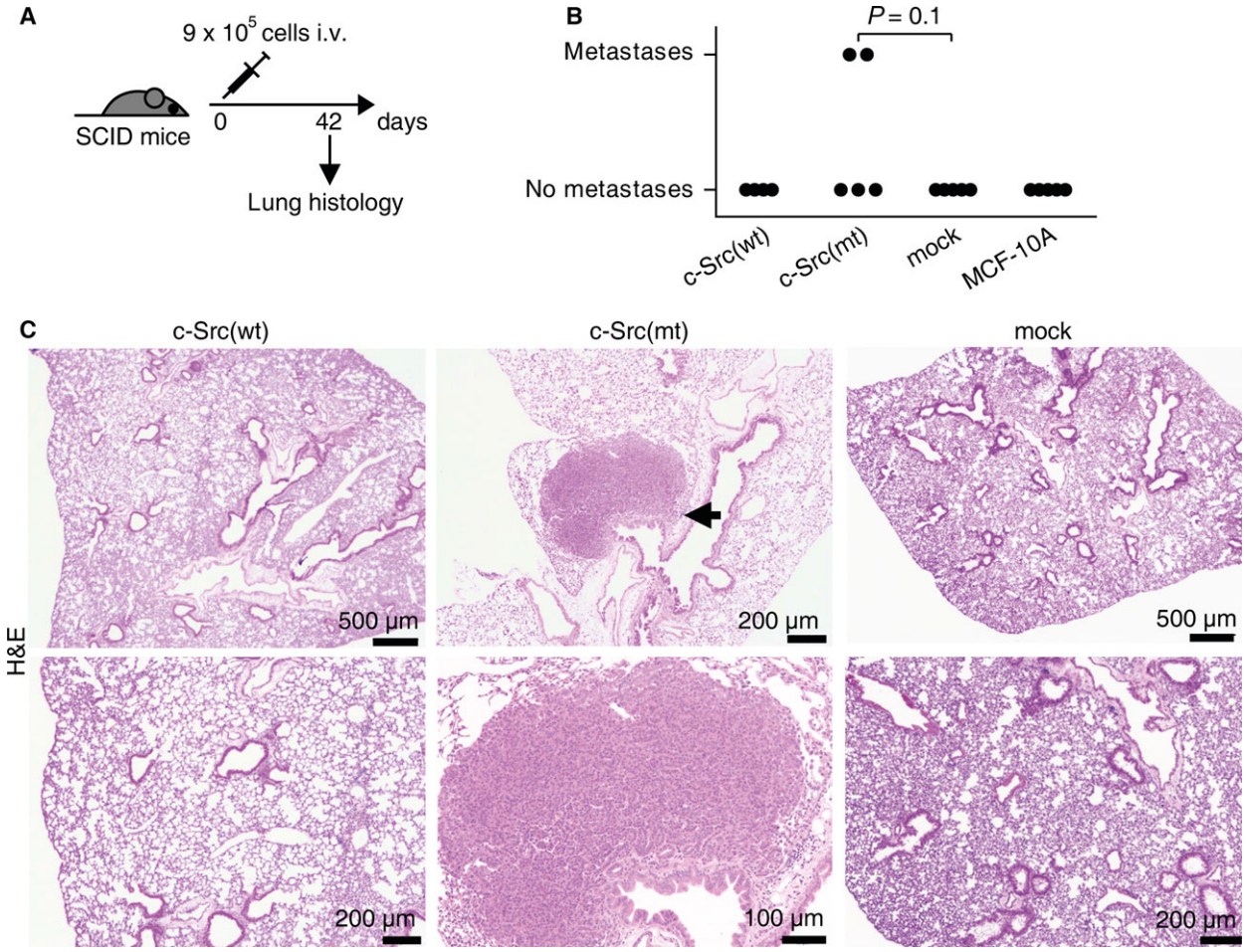


Fig. 4. c-Src(mt)-expressing cells exhibit elevated invasive potential *in vivo*. (A) SCID mice were injected intravenously with cells from one of the four lines and killed 42 days later for histological analyses of their lungs. (B) Determination of lung metastases in mice injected with c-Src(mt)-expressing cells as visualized by hematoxylin/eosin staining [shown in (C)]. The *P* value was determined by Barnard's test [52,53]. (C) Macroscopic analysis of lungs. The arrow indicates a representative metastasis found in a mouse injected with c-Src(mt)-expressing cells.

benign, non-metastatic tumors or normal tissues [59]. The over-expression of FAK in various human neoplastic diseases, such as invasive colorectal carcinoma or esophageal squamous cell carcinoma [59], may be caused by Src kinase-independent mechanisms. Increased FAK phosphorylation has been identified in primary breast cancer specimens and breast cancer cell lines compared to normal tissue, which may be a direct consequence of Src kinase activity [62,63].

In line with these findings, the mRNA expression changes induced by c-Src(mt) showed significant similarities with transcription profiles of many primary human cancers, especially breast cancer, leukemia, ovarian, prostate and renal cancers (Fig. 3A). We observed significant overlap with mRNA profiles of some colon cancers (three of six comparisons of normal tissue versus cancer and one of three comparisons of metastatic versus localized), for which activating C-terminal mutations of the c-Src gene have been reported previously [24]. Thus, c-Src(mt)-expressing cells not only exert increased migratory potential, but also show a transcription signature with similarities to some primary tumors that may not yet have progressed to metastatic behavior. It should be noted that tumorigenesis and progression to metastasis *in vivo* is a multi-step process that involves inactivation of tumor suppressor genes and acquisition of oncogenic mutations [64,65]. While the cellular model studied herein certainly does not reflect the complex mechanisms of tumorigenesis *in vivo*, it provides insights into Src-dependent progression to metastasis.

Previous genome-wide analyses of the effect of Src activity on gene transcription include microarray

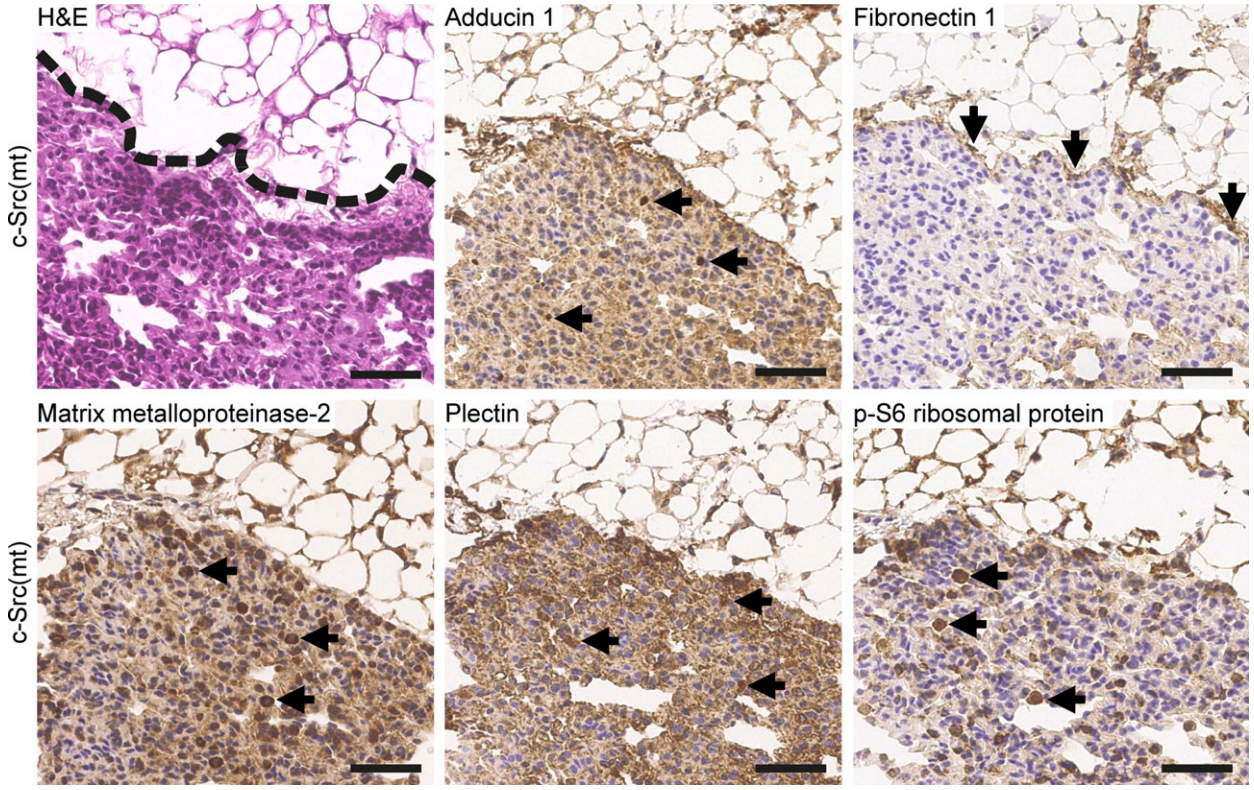


Fig. 5. Protein expression of selected effector genes in lung metastases detected by immunohistochemistry. The border between the metastasis and normal lung tissue is indicated in the hematoxylin/eosin-stained panel by a dashed line. Arrows indicate representative cells that stained positive for the indicated proteins. Scale bars = 50 μm.

analyses of v-Src-transformed mouse cells [3,66] or RSV-transformed chicken cells [67,68]. One previous study analyzed the gene transcription profile of MCF-10A cells transformed with ER-Src, a derivative of v-Src that is fused to the ligand-binding domain of the estrogen receptor [7]. In contrast, the cellular model described herein reduces the induction of metastatic behavior to a single point mutation in the c-Src protein resulting in loss of binding to tumor suppressors, leading to increased kinase activity. This may explain why there is little agreement between the c-Src(mt) effector genes identified in this study and the v-Src-regulated genes identified by others [3,7,66–68]. Interactions with PDZ domain-containing tumor suppressor proteins not only influence the kinase activity of c-Src, but also its intracellular localization [17]. It would therefore be interesting to compare c-Src(mt) to other activated Src mutants, for instance substituted at tyrosine residue 529 (Fig. 1B) but with an intact C-terminal GENL peptide. This may reveal which effects are directly attributed to interactions with PDZ domain-containing proteins.

We identified 71 c-Src(mt) effector genes that have not been previously associated with metastasis, which were mainly involved in mRNA splicing, and thus influence protein synthesis (Table S4). A number of these genes also showed transcriptional de-regulation in human primary cancer samples, especially in leukemia (Fig. 3B). SFK members such as Fgr or Hck that, similar to c-Src(mt), do not bind to PDZ proteins, are mainly present in non-adherent hematopoietic cells including leukocytes [17]. This may explain why we observed such a high degree of overlap between transcriptional signatures of leukemic and c-Src(mt)-expressing cells.

One way by which c-Src(mt) may alter transcription levels is by activating specific transcription factors (TFs). To investigate this, we identified DNA sequence motifs that were enriched within the promoter regions of c-Src(mt) effector genes, using the Amadeus motif discovery platform [69]. A number of these motifs showed overlap with known binding sites of 17 human TFs (Fig. 6) [70–84]. These TFs may be directly or indirectly activated by c-Src(mt), thereby contributing to the mRNA expression profile observed in c-Src(mt)-

expressing cells. For instance, one of the identified TFs, the signal transducer and activator of transcription 3 (STAT3), has been previously identified as a downstream target of c-Src [1]. DNA binding of STAT3 is activated in Src kinase-transformed cells through tyrosine phosphorylation [85]. Similarly, STAT5A may be activated by Src kinase [86]. Moreover, Src signaling has been shown to regulate the activity of the TF p300 that shows histone acetyltransferase activity, thereby contributing to epigenetic gene regulation [87]. Src kinase positively regulates the histone acetyltransferase activity of p300, and thereby its ability to increase the transcription of target genes [87]. Of note, neither the three TFs mentioned nor the other identified TFs were differentially expressed in c-Src(mt)-expressing cells (Fig. 6 and Table S1). This indicates that their activity was probably altered through post-translational modifications involving c-Src(mt), or increased mRNA turnover.

Overall, this study provides detailed insights into the transcriptional program induced by a specific C-terminally mutated form of c-Src kinase, c-Src(mt), which is impaired in binding to tumor suppressors. Cells expressing this mutant extravasated and migrated into lung tissue of mice *in vivo* to form metastases. We identified a number of c-Src(mt)-regulated genes that have no yet been associated with metastatic behavior. They may deserve further investigation as putative biomarkers of invasive cells or as targets of therapeutic approaches, especially in the context of Src-driven progression to metastasis.

Experimental procedures

Ethics statement

Animals were maintained under specific pathogen-free conditions. All experiments and procedures involving animals were approved by the Cantonal Veterinary Authority of Zurich, in accordance with the guidelines of the Swiss Animal Protection Law, which strictly follows the principles declared in the Basel Declaration, namely that research in animals is performed according to the recommendations of the International Council for Laboratory Animal Science. All expression data for human tumors and tissues analyzed in this study were retrieved from the ONCOMINE public domain database [44].

Cells

MCF-10A cells (ATCC-CRL-10317) were obtained from the American Type Culture Collection (Manassas, VA, USA) and grown in complete growth medium (Dulbecco's modified Eagle's medium/F12; Gibco, Grand Island, NY, USA) containing 2 mM L-glutamine and supplemented with 20 $ng{\cdot}mL^{-1}$ epidermal growth factor (Sigma-Aldrich, St Louis, MO, USA), 100 $ng{\cdot}mL^{-1}$ cholera toxin, 10 $ng{\cdot}mL^{-1}$ insulin and 500 $ng{\cdot}mL^{-1}$ hydrocortisone (all Sigma-Aldrich) plus 5% horse serum (Gibco). Cells were maintained at 37 °C with 5% CO_2.

Lentiviral constructs

The hemagglutinin-tagged Src constructs have been described elsewhere [17–19]. The lentiviral particles for Tet-inducible expression of c-Src(wt) (LVP-Src-wt) and c-Src (mt) (LVP-Src-mt), as well that encoding as a short nonsense construct (LVP-mock), conferred resistance to Blasticidin S. The construct LVP017-Neo allows constitutive expression of TetR and confers resistance to G418 (Fig. 1C). Lentiviral particles were prepared by AMS Biotechnology (Abingdon, UK). Sequences were confirmed by capillary sequencing.

Generation of stably expressing cells

MCF-10A cells were grown to 50% confluence in complete growth medium. Cells were incubated with LVP017-Neo (multiplicity of infection = 10) and lentiviral expression particles (multiplicity of infection = 3) for 72 h. Then the medium was replaced with fresh complete growth medium containing 60 $\mu g{\cdot}mL^{-1}$ G418 (Gibco) and 3.5 $\mu g{\cdot}mL^{-1}$ Blasticidin S (Bsd) (Invitrogen, Carlsbad, CA, USA) for 14 days to select doubly positive cells. Non-transfected control cells were killed within 14 days when exposed to either 60 $\mu g{\cdot}mL^{-1}$ G418 or 3.5 $\mu g{\cdot}mL^{-1}$ Bsd. Doubly positive cells were maintained in complete growth medium containing 30 $\mu g{\cdot}mL^{-1}$ G418 and 1.75 $\mu g{\cdot}mL^{-1}$ Bsd. Expression of TetR mRNA was confirmed by qRT-PCR, and was at comparable levels in the three engineered cell lines. The tetracycline-inducible system allowed for regulation of transgene expression by adding tetracycline (Tet) to the cell culture medium. Inducible expression of the transgene in c-Src(wt)- and c-Src(mt)-transfected cells was verified by qRT-PCR and Western blotting. For the subsequent gene expression analyses, we selected cell populations that showed transgene expression at a similar level to endogenous c-Src expression (Fig. 1D).

Preparation of total RNA

Cells were grown to approximately 70% confluence in complete medium supplemented with 30 $\mu g{\cdot}mL^{-1}$ G418 and 1.75 $\mu g{\cdot}mL^{-1}$ Bsd. Tetracycline (Sigma-Aldrich) was added to the culture medium, and cells were incubated for 24 h. Then total RNA was extracted using Trizol (Invitrogen) according to the manufacturer's instructions. The RNA

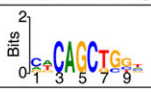
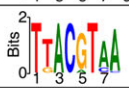
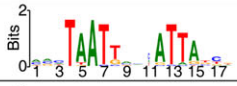
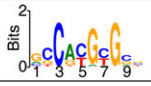
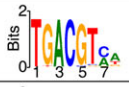
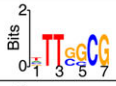
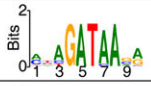
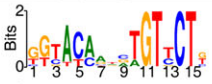
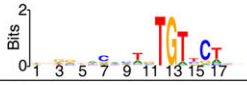
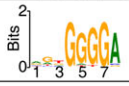
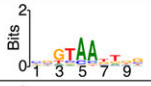
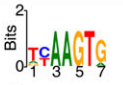
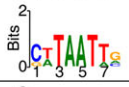
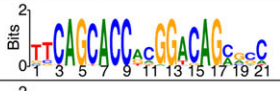
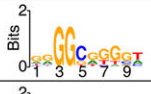
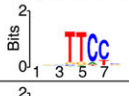
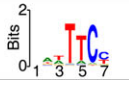
Transcription factor	Gene	Matrix ID	Sequence logo	Reference
AP-4 (Activator protein protein 4)	*TFAP*	M00176-V$AP4_Q6		[70]
ATF2 (Activating transcription factor 2)	*ATF2*	M00040-V$CREBP1_01		[71]
CART1 (Cartilage homeoprotein 1)	*ALX1*	M00416-V$CART1_01		[72]
c-Myc (Myelocystomatosis viral oncogene homolog)	*MYC*	M00322-V$MYCMAX_B		[73]
CREB (cAMP-responsive element binding protein)	*CREB1*	M00039-V$CREB_01		[69]
E2F1 (E2F transcription factor 1)	*E2F1*	M00430-V$E2F1_Q4		[74]
GATA-1 (GATA-binding factor 1)	*GATA1*	M00347-V$GATA1_06		[75]
GR (Glucocorticoid receptor)	*NR3C1*	M00205-V$GRE_C		[76]
		M00192-V$GR_Q6		
HNF-1 (Hepatic nuclear factor 1)	*HNF1A*	M00132-V$HNF1_01		[77]
MZF1 (Myeloid zinc finger protein 1)	*MZF1*	M00083-V$MZF1_01		[78]
NCX (Neural crest homeobox protein)	*TLX2*	M00484-V$NCX_01		[79]
NKX2-5 (Nkx2-5 homeo domain factor)	*NKX2-5*	M00240-V$NKX25_01		[80]
		M00241-V$NKX25_02		
NRSF (Neural-restrictive silencer factor)	*REST*	M00256-V$NRSF_01		[81]
p300	*EP300*	M00033-V$P300_01		[82]
Sp1 (Stimulating protein 1)	*SP1*	M00008-V$SP1_01		[83]
STAT3 (Signal transducer and activator of transcription 3)	*STAT3*	M00497-V$STAT3_02		[84]
STAT5A (Signal transducer and activator of transcription 5a)	*STAT5A*	M00493-V$STAT5A_03		[85]

Fig. 6. Human transcription factor binding sites enriched in the promoter regions of c-Src(mt) effector genes. Further details are provided in Fig. S1.

was precipitated at −80 °C and pelleted by centrifugation at 12 000 ***g*** for 10 min. The pellets were washed with 75% ethanol, air-dried, and solubilized in nuclease-free water. Remaining DNA was removed using an RNase-free DNase set (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Then the solution was subjected to phenol/chloroform extraction. RNA integrity was assessed by gel electrophoresis using 300 ng total RNA per lane in a 1× TAE gel containing 1% w/v agarose. The gel was run at 130 V for 45 min, and stained with ethidium bromide. The GeneRuler 1 kb DNA ladder (Fermentas, Waltham, MA, USA) was used as a size marker.

Primer design and synthesis

Primers for qRT-PCR were designed such that PCR products were smaller than 300 bp and were intron-spanning, using the primer3 program (http://primer3.wi.mit.edu) [88] and cDNA sequences retrieved from the UCSC Genome Browser (https://genome.ucsc.edu) and the current release of the human genome [89]. The GAPDH primer pair has been described previously [90]. The forward primer mu SRC_HA is situated within the hemagglutinin tag of the c-Src constructs to avoid amplification of endogenous c-Src expressed by MCF-10A cells. Primers were synthesized by Metabion (Martinsried, Germany). Table 1 provides a complete list of primer sequences used in this study.

Quantitative real-time PCR

Total RNA preparations were reverse-transcribed using a high-capacity cDNA reverse transcription kit (Applied Biosciences, Foster City, CA, USA) according to the manufacturer's instructions. Reverse transcription products were quantified using Fast SYBR Green Master Mix (Applied Biosystems) on an Applied Biosystems 7900HT fast real-time PCR system. Cycle conditions were: 10 min incubation at 95 °C, followed by 50 amplification cycles of 95 °C for 15 s and 58 °C for 1 min.

Western blotting

Cells were resuspended in buffer A (10 mM HEPES, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1 mM dithiothreitol, protease inhibitor cocktail) and lyzed mechanically. After centrifugation for 5 min at 2500 ***g***, the pellet was subjected to DNase I digestion in 10 mM HEPES, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.2U DNase I, 1 mM dithiothreitol, protease inhibitor cocktail, and centrifuged as described above. The resulting supernatant was combined with the first one, and the total protein concentration was determined using Bradford reagent [91] (Sigma-Aldrich) according to the manufacturer's instructions. Then 30 μg lysate were subjected to SDS/PAGE and immunoblotting. The primary antibodies used in the experiment were anti-Src clone GD11 (Merck Millipore, Billerica, MA, USA) and anti-GAPDH (AM4300, Ambion, Carlsbad, CA, USA). The secondary antibody was peroxidase-conjugated AffiniPure goat anti-mouse IgG (H + L) (Jackson Immuno Research, West Grove, PA, USA).

RNA sequencing

Aliquots (1 μg) of total RNA preparations from the four cell lines were subjected to cDNA library preparation using the NEBNext Ultra RNA library prep kit (New England Biolabs, Ipswich, MA, USA) and subsequently to DNA sequencing on an Illumina Genome Analyzer IIx sequencer (Illumina, San Diego, CA, USA) using a 100 bp paired-end multiplex run, yielding 99 100 500, 46 988 060, 38 071 856 and 56 106 530 quality control-passed reads for naive MCF-10A, mock, c-Src(wt) and c-Src(mt) lines, respectively.

Sequence read alignment

The Illumina paired-end sequence reads were mapped to the University of California Santa Cruz hg19 human reference genome (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/) using the Bowtie read aligner (0.12.8 release, default parameters) [21].

Gene expression and determination of differential expression

Gene expression was measured in non-transfected cells (naive MCF-10A), cells transfected with the mock lentiviral vector, cells transfected with the c-Src(wt) lentiviral vector and cells transfected with the c-Src(mt) lentiviral vector using Illumina RNA-Seq technology. RNA expression levels were determined using a standardized analysis pipeline based on RSEM (RNA-Seq by expectation maximization) [22] and DESeq (differential expression of RNA-Seq data) [23]. RSEM version 1.1.20 was used to quantify transcripts from RNA-Seq using the following parameters for 'rsem-prepare-reference': reference genome hg19; annotation Ensembl 67; polyA length 15; Bowtie version 0.12.8. The following parameters were used for 'rsem-calculate-expression': calculate 95% credibility intervals; paired-end reads. The retrieved transcript count data were then fed into DESeq, which normalizes the count data between samples and calculates expression fold changes (FC) as well as *P* values and adjusted *P* values to determine differential expression of transcripts between two samples (cell lines). The default settings were used for normalization and variance estimation: estimateSizeFactors(cds) and estimateVarianceFunctions(cds), respectively. Differential expression between cell lines A and B was tested using nbinomTest(cds, 'cell line A', 'cell line B'). Corresponding genes with a false discovery rate-adjusted *q* value ≤ 0.05 for their fold change were considered differentially expressed. For any given gene of the comparison between cell line A and cell line B, if transcripts were detect-

able in A but not detectable in B, the reported FC value is 'Inf'; if transcripts were detectable in B but not detectable in A, the reported FC value is '–Inf'.

Gene Ontology over-representation

The ConsensusPathDB [34] was used to identify predominantly occurring GO terms among the candidate genes in order to determine functional commonalities between them. *P* values were derived from a Fisher's exact test performed for the candidate genes against the background set of all genes in the human genome. Correction for multiple testing was performed using the false discovery rate method [92], to calculate *q* values.

Pathway over-representation

Over-represented pathways involving the candidate genes were identified using ConsensusPathDB (release 25) [34], a pathway database that currently integrates interaction data from 31 public resources. *P* and *q* values were calculated as for the above GO analysis.

Mouse experiments

For the tumor growth studies, severe combined immune-deficient (SCID) mice were obtained from Charles River Laboratories (Wilmington, MA, USA). Eight-week-old female mice were intravenously injected with 9×10^5 MCF-10A cells or c-Src(wt), c-Src(mt) or mock cells. Lung tissues were evaluated after 42 days. To induce transgene expression in engineered cell lines, 1 $\mu g \cdot mL^{-1}$ Tet was added to the drinking water of all mice.

Immunohistochemistry

Lung tissues were fixed with 4% paraformaldehyde and embedded in paraffin. Paraffin sections (2 μm) were either stained with hematoxylin/eosin or by automated immunohistochemistry staining using rabbit anti-MMP2 (Abcam, Cambridge, UK), rabbit anti-fibronectin (Abcam), mouse anti-α-adducin (Abcam), rabbit anti-plectin (Abcam), rabbit anti-phospho-S6 ribosomal protein (Ser235/236) (Cell Signaling Technology, Danvers, MA, USA), and polymeric horseradish peroxidase (Poly-HRP)-conjugated anti-Rabbit IgG and anti-Mouse IgG (Leica, Wetzlar, Germany) for detection. Image acquisition was performed using DotSlide BX51 (Olympus, Tokyo, Japan), SCN400 (Leica), Axio Z1 (Zeiss, Oberkochen, Germany) or BX53 (Olympus) microscopes.

Literature mining

We aimed to find genes that may be associated with cellular migration but have so far not been reported in this context. Thus, we used an automated literature mining approach to scan PubMed abstracts for information on the candidate genes with respect to migration. For that purpose, we defined a lexicon of migration-related terms (migration, metastasis, pseudopodia, wound healing, adherence, adhesion, homing, pathfinding, motility, angiogenesis, invasion, cancer progression, PDZ domain, cytoskeleton), and searched for co-occurrence of the gene names with those terms. We used an evidence score defined as

$$s_{ij} = \log_2(P_{ij}/(P_i P_j))$$

to quantify literature evidence for a gene, where P_{ij} is the frequency of co-occurrence of the lexicon term i and the gene j, and P_i and P_j are their marginal frequencies. The evidence score for each gene was computed as the sum of co-associations over the entire lexicon:

$$s_j = \Sigma_i s_{ij}$$

A high score indicates that a gene has often been related to metastasis and other terms of the lexicon. Genes with low or negative evidence scores were therefore selected for further analysis. Seventy-one genes showed no evidence.

Cancer gene expression profiles

Gene set enrichment analysis was performed using publicly available cancer gene expression datasets from the ONCOMINE database [44]. We looked for cancer datasets that showed similar expression profiles to our c-Src(mt)-expressing cells versus mock cells, focusing on studies that compared different stages of cancer tissue, i.e. sets with the prefix NC for 'normal tissue versus cancer' and CP for 'cancer progression versus primary tumor' (referred to as 'metastatic versus localized tumor' throughout the manuscript). At the time of writing, 112 such datasets were available, all of which were considered. To calculate the significance of overlap of c-Src(mt) effector genes with the gene sets from the ONCOMINE database, we performed Fisher's exact tests and corrected for multiple testing using the false discovery rate method.

Identification of transcription factor binding sites

DNA sequence motifs that are enriched in the promoter regions of c-Src(mt) effector genes were determined using the Amadeus platform [69]. Standard parameters were used, with motif lengths of 8–12 bp and the hg19 release of the human genome (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/). The output from Amadeus is shown in Fig. S1. Human transcription factor binding sites from TRANSFAC [93] that showed similarity to the identified

motifs were chosen for further investigation. The respective position-specific weight matrices were retrieved from the -Cistrome database at cistrome.dfci.harvard.edu [94]. Sequence logos were computed from position-specific weight matrices using WebLogo3 at weblogo.threeplusone.com [95].

Acknowledgments

We thank Jochen Heinrich (University of Zurich, Switzerland) for help with generating the c-Src expression constructs, Nada Kumer (Max Planck Institute for Molecular Genetics) for excellent technical assistance, Algirdas Ziogas (University Hospital Zurich) for help with some of the experiments, and Peter H. Seeberger (Max Planck Institute of Colloids and Interfaces, Potsdam, Germany) and Martin Vingron (Max Planck Institute for Molecular Genetics) for their generous support. The excellent technical assistance of Ruth Hillermann (Helmholtz Zentrum Munich) is gratefully acknowledged. We thank the Bundesministerium für Bildung und Forschung (grants 0315428A and 0316190A to RH) and the Max Planck Society for financial support. This work was also supported by private funding (KM). MH was supported by the Helmholtz Alliance Preclinical Comprehensive Cancer Center, a European Research Council starting grant (Liver Cancer Mechanism), the Helmholtz Foundation, the SFBTR36 and the Hofschneider Foundation. MRS was supported by the Volkswagenstiftung (Lichtenberg program).

Author contributions

FB performed the *in vitro* experiments, generated stable cell lines and performed qRT-PCR measurements. CH, RH and MK performed bioinformatic evaluations. BT, AW and MRS performed Illumina sequencing, quality control and Western blotting. LB and MH designed and performed the *in vivo* mouse studies and performed the histological analyses. HL and KM designed and initiated the study. FB and KM wrote the manuscript. MH and LB helped with writing the manuscript. KM organized and coordinated the study. All authors read and approved the final manuscript.

References

1 Yeatman TJ (2004) A renaissance for SRC. *Nat Rev Cancer* **4**, 470–480.
2 Martin GS (2004) The road to Src. *Oncogene* **23**, 7910–7917.
3 Paz K, Socci ND, van Nimwegen E, Viale A & Darnell JE (2004) Transformation fingerprint: induced STAT3-C, v-Src and Ha-Ras cause small initial changes but similar established profiles in mRNA. *Oncogene* **23**, 8455–8463.
4 Soule HD, Maloney TM, Wolman SR, Peterson WD Jr, Brenz R, McGrath CM, Russo J, Pauley RJ, Jones RF & Brooks SC (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res* **50**, 6075–6086.
5 Hirsch HA, Iliopoulos D, Tsichlis PN & Struhl K (2009) Metformin selectively targets cancer stem cells, and acts together with chemotherapy to block tumor growth and prolong remission. *Cancer Res* **69**, 7507–7511.
6 Iliopoulos D, Hirsch HA & Struhl K (2009) An epigenetic switch involving NF-κB, Lin28, Let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell* **139**, 693–706.
7 Hirsch HA, Iliopoulos D, Joshi A, Zhang Y, Jaeger SA, Bulyk M, Tsichlis PN, Shirley Liu X & Struhl K (2010) A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. *Cancer Cell* **17**, 348–361.
8 Cooper JA, Gould KL, Cartwright CA & Hunter T (1985) Tyr527 is phosphorylated in pp60c-src: implications for regulation. *Science* **231**, 1431–1434.
9 Kmiecik TE & Shalloway D (1987) Activation and suppression of pp60c-src transforming ability by mutation of its primary sites of tyrosine phosphorylation. *Cell* **49**, 65–73.
10 Chong YP, Mulhern TD & Cheng HC (2005) C-terminal Src kinase (CSK) and CSK-homologous kinase (CHK) – endogenous negative regulators of Src-family protein kinases. *Growth Factors* **23**, 233–244.
11 Zheng XM, Wang Y & Pallen CJ (1992) Cell transformation and activation of pp60c-src by overexpression of a protein tyrosine phosphatase. *Nature* **359**, 336–339.
12 Somani AK, Bignon JS, Mills GB, Siminovitch KA & Branch DR (1997) Src kinase activity is regulated by the SHP-1 protein-tyrosine phosphatase. *J Biol Chem* **272**, 21113–21119.
13 Peng ZY & Cartwright CA (1995) Regulation of the Src tyrosine kinase and Syp tyrosine phosphatase by their cellular association. *Oncogene* **11**, 1955–1962.
14 Frame MC (2002) Src in cancer: deregulation and consequences for cell behaviour. *Biochim Biophys Acta* **1602**, 114–130.
15 Xing Z, Chen HC, Nowlen JK, Taylor SJ, Shalloway D & Guan JL (1994) Direct interaction of v-Src with the focal adhesion kinase mediated by the Src SH2 domain. *Mol Biol Cell* **5**, 413–421.
16 Walter AO, Peng ZY & Cartwright CA (1999) The Shp-2 tyrosine phosphatase activates the Src tyrosine

kinase by a non-enzymatic mechanism. *Oncogene* **18**, 1911–1920.
17 Radziwill G, Weiss A, Heinrich J, Baumgartner M, Boisguerin P, Owada K & Moelling K (2007) Regulation of c-Src by binding to the PDZ domain of AF-6. *EMBO J* **26**, 2633–2644.
18 Baumgartner M, Weiss A, Fritzius T, Heinrich J & Moelling K (2009) The PDZ protein MPP2 interacts with c-Src in epithelial cells. *Exp Cell Res* **315**, 2888–2898.
19 Baumgartner M, Radziwill G, Lorger M, Weiss A & Moelling K (2008) c-Src-mediated epithelial cell migration and invasion regulated by PDZ binding site. *Mol Cell Biol* **28**, 642–655.
20 Weiss A, Baumgartner M, Radziwill G, Dennler J & Moelling K (2007) c-Src is a PDZ interaction partner and substrate of the E3 ubiquitin ligase Ligand-of-Numb protein X1. *FEBS Lett* **581**, 5131–5136.
21 Langmead B, Trapnell C, Pop M & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.
22 Li B & Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.
23 Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**, R106.
24 Irby RB, Mao W, Coppola D, Kang J, Loubeau JM, Trudeau W, Karl R, Fujita DJ, Jove R & Yeatman TJ (1999) Activating SRC mutation in a subset of advanced human colon cancers. *Nat Genet* **21**, 187–190.
25 Lu Y, Zhang K, Li C, Yao Y, Tao D, Liu Y, Zhang S & Ma Y (2012) Piwil2 suppresses p53 by inducing phosphorylation of signal transducer and activator of transcription 3 in tumor cells. *PLoS One* **7**, e30999.
26 Matsui T, Kinoshita T, Hirano T, Yokota T & Miyajima A (2002) STAT3 down-regulates the expression of cyclin D during liver development. *J Biol Chem* **277**, 36167–36173.
27 Majid S, Saini S, Dar AA, Hirata H, Shahryari V, Tanaka Y, Yamamura S, Ueno K, Zaman MS, Singh K *et al.* (2011) MicroRNA-205 inhibits Src-mediated oncogenic pathways in renal cancer. *Cancer Res* **71**, 2611–2621.
28 Luo Y, Liang F & Zhang ZY (2009) PRL1 promotes cell migration and invasion by increasing MMP2 and MMP9 expression through Src and ERK1/2 pathways. *Biochemistry* **48**, 1838–1846.
29 Cermák V, Kosla J, Plachý J, Trejbalová K, Hejnar J & Dvorák M (2010) The transcription factor EGR1 regulates metastatic potential of v-src transformed sarcoma cells. *Cell Mol Life Sci* **67**, 3557–3568.
30 Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408.
31 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29.
32 Dranoff G (2004) Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer* **4**, 11–22.
33 Lai SL, Chien AJ & Moon RT (2009) Wnt/Fz signaling and the cytoskeleton: potential roles in tumorigenesis. *Cell Res* **19**, 532–545.
34 Kamburov A, Stelzl U, Lehrach H & Herwig R (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res* **41**, D793–D800.
35 Viana Lde S, Affonso RJ Jr, Silva SR, Denadai MV, Matos D, Salinas de Souza C & Waisberg J (2013) Relationship between the expression of the extracellular matrix genes SPARC, SPP1, FN1, ITGA5 and ITGAV and clinicopathological parameters of tumor progression and colorectal cancer dissemination. *Oncology* **84**, 81–91.
36 Chen J, Liu NN, Li JQ, Yang L, Zeng Y, Zhao XM, Xu LL, Luo X, Wang B & Wang XR (2011) Association between ITGA2 C807T polymorphism and gastric cancer risk. *World J Gastroenterol* **17**, 2860–2866.
37 El Behi M, Krumeich S, Lodillinsky C, Kamoun A, Tibaldi L, Sugano G, De Reynies A, Chapeaublanc E, Laplanche A, Lebret T *et al.* (2013) An essential role for decorin in bladder cancer invasiveness. *EMBO Mol Med* **5**, 1835–1851.
38 Yousif NG (2014) Fibronectin promotes migration and invasion of ovarian cancer cells through up-regulation of FAK-PI3K/Akt pathway. *Cell Biol Int* **38**, 85–91.
39 Zhao Y, Jia L, Mao X, Xu H, Wang B & Liu Y (2009) siRNA-targeted COL8A1 inhibits proliferation, reduces invasion and enhances sensitivity to D-limonence treatment in hepatocarcinoma cells. *IUBMB Life* **61**, 74–79.
40 Karagiannis GS, Petraki C, Prassas I, Saraon P, Musrap N, Dimitromanolakis A & Diamandis EP (2012) Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis. *Oncotarget* **3**, 267–285.
41 van de Water B, Tijdens IB, Verbrugge A, Huigsloot M, Dihal AA, Stevens JL, Jaken S & Mulder GJ (2000) Cleavage of the actin-capping protein α-adducin at Asp-Asp-Ser-Asp633-Ala by caspase-3 is preceded by its phosphorylation on serine 726 in cisplatin-induced apoptosis of renal epithelial cells. *J Biol Chem* **275**, 25805–25813.
42 Stegh AH, Herrmann H, Lampel S, Weisenberger D, Andrä K, Seper M, Wiche G, Krammer PH & Peter ME (2000) Identification of the cytolinker plectin as a major early *in vivo* substrate for caspase 8 during

CD95- and tumor necrosis factor receptor-mediated apoptosis. *Mol Cell Biol* **20**, 5665–5679.
43 David CJ & Manley JL (2010) Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* **24**, 2343–2364.
44 Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A & Chinnaiyan AM (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6.
45 Irby RB & Yeatman TJ (2000) Role of Src expression and activation in human cancer. *Oncogene* **19**, 5636–5642.
46 Edwards J (2010) Src kinase inhibitors: an emerging therapeutic treatment option for prostate cancer. *Expert Opin Investig Drugs* **19**, 605–614.
47 Qayyum T, McArdle PA, Lamb GW, Jordan F, Orange C, Seywright M, Horgan PG, Jones RJ, Oades G, Aitchison MA *et al.* (2012) Expression and prognostic significance of Src family members in renal clear cell carcinoma. *Br J Cancer* **107**, 856–863.
48 Aleshin A & Finn RS (2010) SRC: a century of science brought to the clinic. *Neoplasia* **12**, 599–607.
49 Koussounadis A, Langdon SP, Um IH, Harrison DJ & Smith VA (2015) Relationship between differentially expressed mRNA and mRNA–protein correlations in a xenograft model system. *Sci Rep* **5**, 10775.
50 Csárdi G, Franks A, Choi DS, Airoldi EM & Drummond DA (2015) Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genet* **11**, e1005206.
51 Yang S, Zhang JJ & Huang XY (2012) Mouse models for tumor metastasis. *Methods Mol Biol* **928**, 221–228.
52 Barnard GA (1945) A new test for 2×2 tables. *Nature* **156**, 177.
53 Barnard GA (1947) Significance tests for 2×2 tables. *Biometrika* **34**, 123–138.
54 Malek RL, Irby RB, Guo QM, Lee K, Wong S, He M, Tsai J, Frank B, Liu ET, Quackenbush J *et al.* (2002) Identification of Src transformation fingerprint in human colon cancer. *Oncogene* **21**, 7256–7265.
55 Dufner A & Thomas G (1999) Ribosomal S6 kinase signaling and the control of translation. *Exp Cell Res* **253**, 100–109.
56 Ruggero D (2013) Translational control in cancer etiology. *Cold Spring Harb Perspect Biol* **5**, a012336.
57 Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ *et al.* (2012) The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61.
58 Schlaepfer DD, Mitra SK & Ilic D (2004) Control of motile and invasive cell phenotypes by focal adhesion kinase. *Biochim Biophys Acta* **1692**, 77–102.
59 Zhao J & Guan JL (2009) Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev* **28**, 35–49.
60 Carragher NO, Westhoff MA, Fincham VJ, Schaller MD & Frame MC (2003) A novel role for FAK as a protease-targeting adaptor protein: regulation by p42 ERK and Src. *Curr Biol* **13**, 1442–1450.
61 Hsia DA, Mitra SK, Hauck CR, Streblow DN, Nelson JA, Ilic D, Huang S, Li E, Nemerow GR, Leng J *et al.* (2003) Differential regulation of cell motility and invasion by FAK. *J Cell Biol* **160**, 753–767.
62 Nishimura M, Machida K, Imaizumi M, Abe T, Umeda T, Takeshima E, Watanabe T, Ohnishi Y, Takagi K & Hamaguchi M (1996) Tyrosine phosphorylation of 100–130 kDa proteins in lung cancer correlates with poor prognosis. *Br J Cancer* **74**, 780–787.
63 Imaizumi M, Nishimura M, Takeuchi S, Murase M & Hamaguchi M (1997) Role of tyrosine specific phosphorylation of cellular proteins, especially EGF receptor and p125FAK in human lung cancer cells. *Lung Cancer* **17**, 69–84.
64 Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646–674.
65 Vogelstein B & Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* **10**, 789–799.
66 Liu Y, Gao L & Gelman ICH (2006) SSeCKS/Gravin/AKAP12 attenuates expression of proliferative and angiogenic genes during suppression of v-Src-induced oncogenesis. *BMC Cancer* **6**, 105.
67 Masker K, Golden A, Gaffney CJ, Mazack V, Schwindinger WF, Zhang W, Wang LH, Carey DJ & Sudol M (2007) Transcriptional profile of Rous Sarcoma Virus transformed chicken embryo fibroblasts reveals new signaling targets of viral-src. *Virology* **364**, 10–20.
68 Maślikowski BM, Néel BD, Wu Y, Wang L, Rodrigues NA, Gillet G & Bédard PA (2010) Cellular processes of v-Src transformation revealed by gene profiling of primary cells – implications for human cancer. *BMC Cancer* **10**, 41.
69 Linhart C, Halperin Y & Shamir R (2008) Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. *Genome Res* **18**, 1180–1189.
70 Mermod N, Williams TJ & Tijan R (1988) Enhancer binding factors AP-4 and AP-1 act in concert to activate SV40 late transcription *in vitro*. *Nature* **332**, 557–561.
71 Benbrook DM & Jones NC (1994) Different binding specificities and transactivation of variant CRE's by CREB complexes. *Nucleic Acids Res* **22**, 1463–1469.
72 Cai RL (1998) Human CART1, a paired-class homeodomain protein, activates transcription through palindromic binding sites. *Biochem Biophys Res Commun* **250**, 305–311.

73 Grandori C & Eisenman RN (1997) Myc target genes. *Trends Biochem Sci* **22**, 177–181.
74 Kirch HC, Pützer B, Schwabe G, Gnauck HK & Schulte Holthausen H (1993) Regulation of adenovirus 12 E1A transcription: E2F and ATF motifs in the E1A promoter bind nuclear protein complexes including E2F1, DP-1, cyclin A and/or RB and mediate transcriptional (auto)activation. *Cell Mol Biol Res* **39**, 705–716.
75 Ko LJ & Engel JD (1993) DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol* **13**, 4011–4022.
76 Chandler VL, Maler BA & Yamamoto KR (1983) DNA sequences bound specifically by glucocorticoid receptor *in vitro* render a heterologous promoter hormone responsive *in vivo*. *Cell* **33**, 489–499.
77 Tronche F & Yaniv M (1992) HNF1, a homeoprotein member of the hepatic transcription regulatory network. *BioEssays* **14**, 579–587.
78 Morris JF, Hromas R & Rauscher FJ 3rd (1994) Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol Cell Biol* **14**, 1786–1795.
79 Shimizu H, Kang M, Iitsuka Y, Ichinose M, Tokuhisa T & Hatano M (2000) Identification of an optimal Ncx binding sequence required for transcriptional activation. *FEBS Lett* **475**, 170–174.
80 Chen CY & Schwartz RJ (1995) Identification of novel DNA binding targets and regulatory domains of a murine *tinman* homeodomain factor, *nkx-2.5*. *J Biol Chem* **270**, 15628–15633.
81 Schoenherr CJ, Paquette AJ & Anderson DJ (1996) Identification of potential target genes for the neuron-restrictive silencer factor. *Proc Natl Acad Sci USA* **93**, 9881–9886.
82 Rikitake Y & Moran E (1992) DNA-binding properties of the E1A-associated 300-kilodalton protein. *Mol Cell Biol* **12**, 2826–2836.
83 Thiesen HJ & Bach C (1990) Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res* **18**, 3203–3209.
84 Ehret GB, Reichenbach P, Schindler U, Horvath CM, Fritz S, Nabholz M & Bucher P (2001) DNA binding specificity of different STAT proteins. Comparison of *in vitro* specificity with natural target sites. *J Biol Chem* **276**, 6675–6688.
85 Yu CL, Meyer DJ, Campbell GS, Larner AC, Carter-Su C, Schwartz J & Jove R (1995) Enhanced DNA-binding activity of a Stat3-related protein in cells transformed by the Src oncoprotein. *Science* **269**, 81–83.
86 Olayioye MA, Beuvink I, Horsch K, Daly JM & Hynes NE (1999) ErbB receptor-induced activation of stat transcription factors is mediated by Src tyrosine kinases. *J Biol Chem* **274**, 17209–17218.
87 Gong J, Zhu J, Goodman OB Jr, Pestell RG, Schlegel PN, Nanus DM & Shen R (2006) Activation of p300 histone acetyltransferase activity and acetylation of the androgen receptor by bombesin in protate cancer cells. *Oncogene* **25**, 2011–2021.
88 Rozen S & Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365–386.
89 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM & Haussler D (2002) The human genome browser at UCSC. *Genome Res* **12**, 996–1006.
90 Kwok T, Helfer H, Alam MI, Heinrich J, Pavlovic J & Moelling K (2009) Inhibition of influenza A virus replication by short double-stranded oligodeoxynucleotides. *Arch Virol* **154**, 109–114.
91 Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254.
92 Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate-A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* **57**, 289–300.
93 Wingender E, Dietze P, Karas H & Knuppel R (1996) TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**, 238–241.
94 Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y *et al.* (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* **12**, R83.
95 Crooks GE, Hon G, Chandonia JM & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. List of 435 differentially expressed genes obtained by comparing c-Src(mt) versus mock cells by RNA-Seq.

Table S2. Gene Ontology over-representation analysis for c-Src(mt) effector genes.

Table S3. Pathway over-representation analysis for c-Src(mt) effector genes.

Table S4. Gene Ontology over-representation analysis for non-metastasis-associated c-Src(mt) effector genes.

Table S5. ONCOMINE analysis.

Fig. S1. Motif discovery in the promoter regions of c-Src(mt) effector genes.