Transcriptional signature induced by a C-terminal c-Src mutant in a human breast cell line

Felix Broecker[1,2,3*], Christopher Hardt[1], Ralf Herwig[1], Bernd Timmermann[1], Martin Kerick[1,4], Andrea Wunderlich[1], Michal R. Schweiger[1,4], Lubor Borsig[5], Mathias Heikenwalder[6,7], Hans Lehrach[1,8,9] and Karin Moelling[1,2]

1 Max Planck Institute for molecular Genetics, Ihnestr. 63, D-14195 Berlin, Germany
2 University of Zurich, Gloriastr. 30/32, CH-8006 Zurich, Switzerland
3 Present address: Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, D-14424 Potsdam, Germany
4 Present address: Cologne Center for Genomics, University of Cologne, Weyertal 115b, D-50931 Cologne, Germany
5 Institute of Physiology & Zurich Center for Integrative Human Physiology, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland
6 Institute of Virology, Technische Universität München, Schneckenburgerstr. 8, and Helmholtz Zentrum Munich, Trogerstr. 30, D-81675 Munich, Germany
7 Department Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, D-69121 Heidelberg, Germany
8 Dahlem Centre for Genome Research and Medical Systems Biology, Fabeckstr. 60-62, D-14195 Berlin, Germany
9 Alacris Theranostics GmbH, Fabeckstr. 60-62, D-14195 Berlin, Germany

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

Running title: c-Src transcriptional signature

Article type : Original Article

Abbreviations: DE, differentially expressed; ECM, extracellular matrix; FAK, Focal Adhesion Kinase; GO, Gene Ontology; logFC, logarithmic fold change; ORA, overrepresentation analysis; PDZ, postsynaptic density-95/discs large/zona occludens-1); RSEM, RNA-Seq by Expectation Maximization; RSV, Rous Sarcoma Virus; SFK, Src family kinase; TF, transcription factor

Databases: This project was submitted to NCBI BioProject with ID: PRJNA288540. The Illumina RNA-Seq reads are available in the NCBI SRA database under study ID SRP060008 with accessions SRS977414 for MCF-10A, SRS977717 for mock, SRS978053 for c-Src(wt), and SRS978046 for c-Src(mt) cells.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/febs.13694

Keywords: c-Src, metastasis, migration, PDZ domain, oncogene, transcriptome

Abstract

Deletions at the C-terminus of the proto-oncogene protein c-Src kinase are characteristic of the viral oncogene protein v-Src and are present in some advanced human colon cancers. They are associated with increased kinase activity and elevated cellular invasiveness. Here, we analyzed the mRNA expression signature of a constitutively active C-terminal mutant of c-Src, c-Src(mt), in comparison to its wildtype protein, c-Src(wt), expressed in the human non-transformed breast epithelial cell line MCF-10A. We demonstrated previously that the mutant changed migratory and metastatic properties. Genome-wide transcriptome analysis revealed that c-Src(mt) deregulated the expression levels of about 430 mRNAs whose gene products are mainly involved in the cellular processes of migration and adhesion, apoptosis and protein synthesis. More than 80% of these genes have previously been linked to cellular migration, while the others play roles, for instance, in RNA transport and splicing processes. Consistent with the transcriptome data, c-Src(mt)-, but not c-Src(wt)-expressing cells showed the capacity to metastasize into mouse lung tissue *in vivo*. The mRNA expression profile of c-Src(mt)-expressing cells shows significant overlap with that of various primary human tumor samples, perhaps reflecting elevated Src activity in some cancerous cells. Expression of c-Src(mt) lead to elevated migratory potential. We used this model system to analyze the transcriptional changes associated with an invasive cellular phenotype. We identified genes and pathways deregulated by c-Src(mt) as biomarkers with potential interest for diagnostics or therapy of metastatic cells.

Introduction

The c-Src proto-oncogene is the cellular homologue of the viral oncogene v-Src of the Rous Sarcoma Virus (RSV), which was the first retroviral oncogene protein. c-Src is a non-receptor tyrosine kinase 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 has originally been shown in RSV-infected and thereby v-Src-expressing chicken cells [2]. Src kinase-induced transformation can also occur in mammalian cells. This has been shown, for instance, 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 the 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 on 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 Y_{416} and Y_{527} (numbering according to v-Src) and by protein-protein interactions through its Src-Homology 2 and 3 (SH2 and SH3) domains [1]. Phosphorylation of Y_{527} inactivates the kinase activity of c-Src by inducing a compact repressed autoinhibitory protein conformation, while its dephosphorylation leads to an open, active conformation [8,9]. Y_{527} is phosphorylated by non-receptor tyrosine kinases, the C-terminal Src kinase (CSK) and the CSK-homologous kinase (CHK) [10]. Dephosphorylation of Y_{527} is catalyzed by several protein tyrosine phosphatases (PTPs) such as PTPα [11], SH2 domain-containing

phosphatase-1 (SHP-1) [12] and SHP-2 [13]. Y_{416} is an autophosphorylation site whose phosphorylation is required for optimal c-Src kinase activity [14]. Moreover, various c-Src-interacting proteins can modulate its kinase activity non-enzymatically, including, for instance, Focal Adhesion kinase (FAK) [15] and SHP-2 [16] that 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, Lyn, Lck, Hck and Blk, are mainly expressed in non-adherent hematopoietic cells [17]. We noticed that the C-terminus of c-Src was relevant for adherence and that its absence, for instance in the v-Src protein, seemed to contribute to non-adherent phenotypes of cells. Therefore, we suspected a possible role of the deleted 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/zona occludens-1) domains, whereby the very C-terminal hydrophobic Leucine is crucial for this interaction. By substitution of this single amino acid from Leucine to Alanine, herein referred to c-Src(mt), we have recently demonstrated the loss of binding to PDZ domains, which are characteristic of many tumor suppressor proteins in MCF-10A and other cell lines. Some of the c-Src-interacting PDZ proteins were the tumor suppressor AF-6 (ALL-1 fused gene from chromosome 6), the membrane protein palmitoylated 2, MPP2, a protein related to the *Drosophila* tumor suppressor discs large-A, dlg-A, and LNX1 (Ligand-of-Numb protein X1) [17-20]. Disruption of these interactions *via* alteration of the C-terminal motif led to constitutive activation and elevated kinase activity compared to the wildtype c-Src protein with the intact GENL sequence [17-20]. We have previously noticed that the expression of c-Src(mt) changed the cellular morphology more than c-Src(wt). This included impaired cell polarization and disorganization of the cytoskeleton in MCF-10A cells associated with increased cellular transformation, which has been characterized by several *in*

vitro assays [17,19]. The 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 into 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 here subjected the identical cellular model we used previously, MCF-10A cells expressing transfected c-Src(mt) or wildtype c-Src(wt) constructs [17-20] to genome-wide transcription profiling. This cellular model allows for delineating the effects of the single C-terminal point mutation in the c-Src protein by comparison of otherwise genetically identically cell lines. We show that c-Src(mt), but not c-Src(wt), significantly deregulated the mRNA expression levels of 435 genes, herein referred as c-Src(mt) effector genes. These genes were characterized by means of pathways and biological functions, allowing detailed insights into the mRNA expression program of c-Src(mt). The respective gene products are mainly involved in migration and adhesion, as expected, but also in apoptosis and protein synthesis. Increased migratory potential of c-Src(mt)-expressing cells was suggested by the mRNA expression profile and corroborated in 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 about 20% of the c-Src(mt) effector genes have previously not been associated with cellular migration. These comprise genes 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 of this set of mRNA transcripts with about 60% of the 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 selected the human non-transformed epithelial MCF-10A cell line stably expressing c-Src(mt) or c-Src(wt) as control [17-19]. The c-Src(mt) protein is identical to c-Src(wt) except for a single L-to-A substitution of 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 expressed in non-adherent hematopoietic cells (Fig. 1A). The c-Src protein contains Src homology 1 (SH1) or kinase, SH2 and SH3 domains (Fig. 1B). The experimental approach comprised the 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 with Bowtie [21], the quantification of expression levels at the transcript level through RNA-Seq by Expectation Maximization (RSEM) [22], determination of significantly differentially expressed (DE) mRNAs with DEseq [23] and the functional characterization of the respective gene products.

We stably transfected MCF-10A cells with the previously described constructs [17-20], c-Src(wt) or c-Src(mt), using lentiviral expression vectors (Fig. 1C). Cells transfected with an otherwise identical expression vector but with a randomized nonsense sequence instead of c-Src were designated as mock cells. A tetracycline-inducible system allowed regulation of transgene expression by adding 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) (primer sequences can be found in Table 1) and Western Blot (Fig. 1D). Of note, 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]. RNA preparations displayed characteristic rRNA bands and no signs of degradation (Fig. 1F). We selected those RNA preparations obtained from cells exposed to 1 µg/mL Tet for whole-genome mRNA expression profiling, as transgene expression was at physiological levels (Fig. 1D). 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 (naïve MCF-10A, mock, c-Src(wt) and c-Src(mt), respectively) quality control-passed reads. Reads were mapped to the human genome with Bowtie [21], which yielded 76,735,649 (77.4%), 37,101,761 (79.0%), 30,474,883 (80.1%) and 43,473,640 (77.5%) mapped reads. Gene expression levels were then quantified on the transcript level with RSEM [22].

To learn about the global changes in mRNA expression induced through action of c-Src(wt) and c-Src(mt), transcript levels of both cell lines were compared to those of naïve MCF-10A cells, using DEseq [23]. The comparison of mock vs. naïve MCF-10A served as negative control. Resulting logarithmic fold change values (logFC) of significantly DE mRNAs ($q \leq 0.05$) were plotted as density graphs (Fig. 1G). While the profiles of mock and c-Src(wt) cells were almost identical and might reflect unspecific 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 upregulated genes), higher q values (Fig. 1H) and a higher proportion of upregulated genes (Fig. 1G). The similar density graphs of mock- and c-Src(wt)-expressing cells gave a first indication that c-Src(wt) was incapable of inducing detectable Src-specific mRNA expression changes, as noted before [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 suppressed by Src, as well as matrix metalloproteinase-2

(*MMP2*) [28] and hyaluronan synthase 2 (*HAS2*) [29] (Fig. 1I). *GAPDH* served as control, which as expected showed identical expression levels in all four cell lines. While the reported effects of Src could be verified for all of these known Src target genes in the c-Src(mt)-expressing cells at the mRNA level, cycle threshold (Ct) values obtained by qRT-PCR in the c-Src(wt)-expressing cells were similar to those of the naïve MCF-10A and mock cells. This gave further credence to the notion that expression of c-Src(wt) alone was insufficient to induce deregulation of Src targets in our cellular model – in contrast, expression of c-Src(mt) did suffice to alter the transcript levels of these genes. This is in accordance with previous findings that either activating mutations [17-20,24] or co-factors [1] are required for increased Src activity. This might well 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, has been shown to be impaired in binding to PDZ domains, rendering this mutant constitutively active [19].

We next sought to investigate effector mRNAs regulated by c-Src(mt). To account for possible Src-independent gene expression changes by the lentiviral transfection we selected the comparison to mock cells instead of naïve MCF-10A cells as a more stringent control. This comparison, c-Src(mt) vs. mock, yielded a total of 435 significantly DE mRNAs (Table S1), whose respective genes are being referred to as c-Src(mt) effector genes. This gene list contained all DE transcripts with $q \leq 0.05$ irrespective of the fold change value. Of these, 52% (224 genes) were upregulated in the 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 inferred by RNA-Seq except for *TP53*, likely 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]. Expectedly, *GAPDH* showed no altered mRNA expression, which was also seen at the protein level (Fig. 1D). Interestingly, the expression level of endogenous c-Src was not altered at the mRNA (RNA-Seq) as well as at the protein level (Fig. 1D). Therefore any effects on mRNA expression levels were due to activity of ectopically expressed c-Src(mt) and not of endogenous c-Src, which was expressed at identical levels in all four cell lines. To further verify the RNA-Seq data, we selected eight additional genes with significant mRNA deregulation in the comparison of c-Src(mt) vs. mock for qRT-PCR analysis (Fig. 1I). logFC values were determined with the $2^{-\Delta\Delta CT}$ method and plotted against the logFC values obtained from RNA-Seq (Fig. 1J). The strong correlation of R^2=0.983 verified the transcript levels inferred by RNA-Seq.

In summary, the c-Src(mt)-expressing cells showed increased Src activity with previously reported effects on mRNA expression, and RNA-Seq data correlated well with qRT-PCR expression analysis. The comparison c-Src(mt) vs. mock provided a comprehensive list of 435 c-Src(mt) effector genes for further characterization.

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

To gain 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 through Gene Ontology (GO) overrepresentation analysis (ORA) [30]. Please see Table S2 for a complete list of GO terms of this analysis and the corresponding gene lists. Analysis of gene products falling into the GO category of 'Biological Process' (Fig. 2A) showed that c-Src(mt) effector genes played 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 the observation that c-Src(mt) promoted an invasive phenotype in epithelial cells and impaired cell polarization and cytoskeletal organization [17-20]. Moreover, expression of c-Src(mt) significantly altered mRNA expression levels of genes involved in cellular responsiveness to cytokine stimulation (GO:0034097, $q = 1.31 \times 10^{-3}$). It has been noted that the composition of cytokines produced in tumor microenvironments play an important role in cellular migration [31]. For instance, c-Src(mt)-expressing cells overexpress the mRNA of *WNT5A* ($q = 9.28 \times 10^{-3}$), whose gene product is a ligand of Frizzled receptors. Wnt/Frizzled signaling is known to be implicated in cellular migration [32]. Transcripts of apoptosis-associated genes were deregulated in c-Src(mt)-expressing cells as well (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, $q = 5.33 \times 10^{-4}$) indicate that c-Src(mt) effector genes are implicated 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 [33]. 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 comprise integrin receptor components *ITGA2*, *ITGA6* and *ITGAV*, all of which were transcriptionally downregulated (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 deregulation promotes invasiveness. For instance, *ITGAV* expression has been linked with metastatic activity in colorectal cancer [34], and *ITGA2* genetic variants may be associated with invasion of gastric cancer [35]. On the other hand, a diversity of components of the ECM itself was deregulated at the mRNA level by the action of c-Src(mt). This includes upregulation of Decorin (*DCN*), Fibronectin 1 (*FN1*), Fibrillin 1 (*FBN1*), and downregulation of E-cadherin (*CDH1*) transcripts. Various mRNAs of collagen chain genes were either up- (*COL8A1*, *COL12A1*, *COL28A1*) or downregulated (*COL17A1*). Of note, overexpression of Decorin [36], Fibronectin 1 [37] and collagen chains COL8A1 [38] and COL12A1 [39] have been recently shown to promote invasiveness of different cancer types. This indicates that c-Src(mt) significantly altered cell-matrix interactions, thereby decreasing cell adhesion and promoting cellular migration and invasiveness. Moreover, matrix metalloproteinase-2 (MMP2), an enzyme degrading type IV collagen, the major constituent of basement membranes, was transcriptionally upregulated 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 deregulated by c-Src(wt) at the mRNA level (Fig. 2C), further verifying that the activating mutation of c-Src(mt) promoted 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 upregulated by c-Src(mt) (Fig. 2C). Adducins are components important for the cortical cytoskeleton network and are subject to Caspase-3-mediated cleavage during cisplatin-mediated apoptosis [40]. 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 TNF receptor-mediated apoptosis [41]. Upregulation of cytoskeletal components such as Adducin 1 and Plectin by c-Src(mt)

might thus interfere with caspase-mediated cytoskeleton instability, thereby suppressing apoptosis. While *ADD1* transcripts were not deregulated by c-Src(wt), *PLEC* mRNA was upregulated, however, less markedly compared with 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 found a surprisingly large number of pathways involved in translation and protein synthesis with an overrepresentation of c-Src(mt) effector genes (Fig. 2B). These include various steps of protein synthesis, such as mRNA transport and degradation, 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 upregulated by c-Src(mt) (Fig. 2C). mRNAs of small and large subunit ribosomal proteins, *RPL29*, *RPL31* and *RPS24*, were also upregulated. Amino acid transport across the cell membrane may have been influenced by c-Src(mt), since a number of mRNAs of genes encoding amino acid-transporting solute carrier proteins (SLCs), such as *SLC1A5*, *SLC3A2*, *SLC7A2*, *SLC7A5* and *SLC38A1* were deregulated. In addition, mRNAs of amino acid biosynthesis genes were regulated by c-Src(mt), including downregulation of transcripts of serine hydroxymethyltransferases 1 and 2 (*SHMT1/2*), enzymes that catalyze the conversion of serine to glycine, and downregulation of the glycine dehydrogenase (decarboxylating) gene (*GLDC*) mRNA, 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 downregulated by c-Src(mt), while the polyadenylate-binding protein 1 (*PABPC1*) gene, essential for translation initiation, was upregulated at the mRNA level. Moreover, the overall higher proportion of upregulated

compared to downregulated transcripts induced by the expression of c-Src(mt) indicated an elevated pool of available mRNAs, further increasing protein expression, including that of pro-invasion and anti-apoptosis genes (Fig. 2A-C). Of note, only few of the above mentioned mRNAs involved in protein synthesis were deregulated by c-Src(wt) (Fig. 2C). We conclude that the activation of protein synthesis-related transcripts required the constitutively active form of c-Src, c-Src(mt).

In summary, overexpression of c-Src(mt) was sufficient to modulate transcription of genes involved in cell migration/adhesion, apoptosis and protein synthesis. In particular, deregulation of mRNAs whose gene products are involved in integrin signaling, ECM remodeling, composition of the cytoskeleton and of the translation machinery were 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 the scientific literature in 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 found in PubMed abstracts. Of these, 344 (82.9%) were significantly linked to migration, which might reflect that previous studies investigating Src activity were biased towards its migration-promoting activity. On the other hand, 71 genes (17.1%) were not linked to migration (highlighted green in Table S1). These 71 genes are of particular interest as they might contain genes that have previously not been associated with Src activity. To get an insight about the role of these 71 genes, they were subjected to Gene Ontology (GO) overrepresentation analysis. Interestingly, these genes appeared to be mainly involved in the splicing process of pre-mRNAs (GO:000569 “U12-type spliceosomal complex” and

GO:0071013 “catalytic step 2 spliceosome”, both q = 3.62 x 10^{-2}) (Table S4). De-regulated splicing is a known feature of transformed cells [42]. Our data so far suggests that the constitutively active c-Src(mt) mutant, in addition to promoting metastasis, may have an impact on protein synthesis and splicing. These functions, to our knowledge, have not been attributed to Src activity before.

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

Increased Src activity is a frequent feature of different human cancers [1]. Therefore, we wondered 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 [43] for comparison with c-Src(mt) effector genes. The gene sets in ONCOMINE consist of genes whose transcripts were found differentially expressed when comparing either tumor vs. normal tissue or metastatic vs. primary tumors and are provided as gene sets within the database. 38 of 70 (54%) of the ONCOMINE gene sets of normal tissue vs. cancer pairs showed significant ($q \leq 0.05$) overlap (Fig. 3A, left). Of note, this included most of the breast cancer (4/6), leukemia (5/6), ovarian (5/7), prostate (7/13) and renal cancers (3/5), as well as all liver cancer samples (4/4). Increased Src activity has been identified in breast, ovarian, liver cancer and leukemia [44], as well as prostate [45] and renal cancer [46]. Analysis of ONCOMINE datasets of cancer progression (metastatic vs. localized, n=42) revealed that 18 of 42 (40%) of these gene sets had significant overlap with c-Src(mt) effector genes (Fig. 3A, right), including most of the lung (4/5) and the prostate (9/16) cancer samples. Increased Src activity has been shown to be associated with increased metastasis, amongst others, in lung [44] and prostate cancer [47]. Therefore, c-Src(mt)-expressing cells proved to be a valid,

albeit simplified, model for some human cancers and their progression to metastasis. This suggests that the single amino acid mutation of c-Src(mt) can indeed affect similar transcripts that are found to be dysregulated in diverse tumor or metastatic cells.

We next sought to investigate whether the non-migration-associated c-Src(mt) effector genes (Table S1) are relevant for human cancer. Therefore, we searched the ONCOMINE database for any evidence of their deregulation in normal tissue vs. cancer pairs. We identified 18 genes that showed either significant ($q \leq 0.05$) transcriptional up- (eight genes) or downregulation (ten genes) in at least 50% of the respective cancer types (Fig. 3B) similar to our cellular model, in which mock cells represent normal tissue and c-Src(mt)-expressing cells correspond to cancer cells. Of note, most of these genes were deregulated in leukemia samples. Thus, the non-migration-associated c-Src(mt) effector genes are likely most relevant for this cancer type. This might reflect the fact that the absence of the PDZ binding motif at the C-terminus of c-Src(mt) is reminiscent of other SFK members, such as Fgr or Hck (Fig. 1A) expressed in non-adherent cells of the hematopoietic system.

Discussion

Here we have analyzed a cellular model of C-terminally mutant and constitutively active c-Src(mt) by its expression at physiological levels in non-transformed human cells. The c-Src(mt)-specific transcriptional changes inferred by RNA sequencing were in high 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. In spite of these findings, mRNA expression changes may not necessarily be reflected at the protein level for every identified gene. Recent studies, however, demonstrated good correlations between mRNA and protein levels in differential gene expression analyses, suggesting that quantifying mRNA could be more useful than

previously acknowledged [48,49]. It has to 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 deregulated 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 [44]. The change in mRNA levels of migration/adhesion-associated genes in the c-Src(mt)-expressing cells (Fig. 2B-C) suggested that these cells might also bear increased metastatic potential. To test whether c-Src(mt)-expressing cells exhibit increased capacity to metastasis *in vivo*, we employed an established mouse model of metastatic lung tissue invasion [50]. 8-weeks old SCID mice were intravenously injected with 9×10^5 naïve MCF-10A, c-Src(wt), c-Src(mt), or mock-expressing cells. To induce transgene expression in the engineered cell lines, 1 μg/mL Tet was added to the drinking water of all animals. Mice were sacrificed after 42 days, and lung tissues analyzed phenotypically (Fig. 4A). Macroscopic analysis revealed visible lung metastases exclusively in those mice that have 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 i.v. with c-Src(mt)-expressing cells, while they were absent in the lungs of mice that received c-Src(wt)-expressing cells (0/4) or mock and naïve MCF-10A cells (both 0/5) (Fig. 4B).

Representative histological H&E-stained sections are shown in Fig. 4C. The difference of metastasis formation in c-Src(mt)-expressing compared with mock-expressing cells showed a trend without statistical significance ($p = 0.1$). Therefore, 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, overexpression of the wildtype c-Src protein alone has been shown to cause only minimal changes in the migratory activity [19,29,53]. Accordingly, we observed no metastases in mice that have received c-Src(wt)-expressing cells (Fig. 4B,C).

The metastases of c-Src(mt)-expressing cells were used to qualitatively verify expression of c-Src(mt) effector genes at the protein level by immunohistological stainings. We chose 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), as well as 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 (p-S6 RP), a marker of elevated protein synthesis [54]. The stainings confirmed that Adducin 1, matrix metalloproteinase-2, Plectin and p-S6 RP were expressed within cells of the metastases. Expression of Fibronectin 1 appeared to be restricted to the border of the metastasis.

Literature mining revealed that more than 80% of c-Src(mt) effector genes have previously been linked to cellular migration. The remaining 17% of genes showed an enrichment of genes involved in mRNA splicing (Table S4), which is relevant for gene expression at the level of translation. Interestingly, various cellular pathways relevant for protein synthesis showed an enrichment of genes that were transcriptionally deregulated by c-Src(mt) (Fig. 2B). Moreover, immunohistochemistry revealed the expression of phosphorylated S6 ribosomal protein, a marker of increased translation [54], in c-Src(mt)-expressing cells (Fig.

5). To our knowledge, the deregulation 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, c-Myc [55]. The importance of protein synthesis in cancer development has been neglected historically [55]. However, recent evidence underlines the importance of deregulated translation in different cellular processes, including cancer progression. Using ribosomal profiling, Hsieh et al. [56] 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 [55]. In line with these findings, we found evidence that Src activity caused the deregulation of different steps of the translational control. Proteins whose mRNAs were upregulated 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 activate phosphoinositide 3-kinase (PI3K) by phosphorylation, whose downstream targets include mTOR [55]. mTOR stimulates protein synthesis by phosphorylating EIF4EBP1 and ribosomal S6 kinase p70S6K1/2. Consequently, mTOR and downstream targets p70S6K1/2, EIF4EBP1 and ribosomal protein 6 are phosphorylated/activated by Src [55]. 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 [57,58]. FAK is directly phosphorylated by Src kinase, leading to the activation of multiple intracellular signaling pathways that are relevant, amongst others, for apoptosis and cell migration [58]. FAK and Src kinase synergize in cell migration and invasion [59,60]. 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 of this study revealed that the mRNA expression level of FAK was not significantly influenced by c-Src(mt), as FAK (gene symbol: *PTK2* for 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 while transcription levels remain unaltered. Of note, both upregulation and increased phosphorylation of FAK have been detected in numerous human metastatic tumors compared to benign, non-metastatic tumors or normal tissues [58]. The overexpression of FAK in different human neoplastic diseases, such as invasive colorectal carcinoma or oesophageal squamous cell carcinoma [58] 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 [61,62].

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). Of note, we detected significant overlap with mRNA profiles of some colon cancers (3/6 tumor vs. normal tissue and 1/3 metastatic vs. localized), for which activating C-terminal mutations of the c-Src gene have been reported before [24]. Thus, c-Src(mt)-expressing cells not only exerted increased

migratory potential, but also showed similarities to the transcription signature of some primary cancer diseases, which may become metastatic later. It has to be noted that tumorigenesis and progression to metastasis *in vivo* is a multi-step process that involves the inactivation of tumor suppressor genes and the acquisition of oncogenic mutations [63,64]. This study aimed to characterize the function and effectors of activated Src kinase with loss of binding to PDZ domains as a model for metastasis.

Previous genome-wide analyses of the effect of Src activity on gene transcription include microarray analyses of v-Src-transformed mouse cells [3,65] or RSV-transformed chicken cells [66,67]. 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 ER [7]. In contrast, the cellular model described herein reduces the metastatic mechanism to a single point mutation in the c-Src protein and escape from tumor suppressors, leading to increased kinase activity. This might explain why there is only little concordance with the c-Src(mt) effector genes identified in this study and the v-Src-regulated genes identified by others [3,7,65-67]. Other more complicated multi-component mechanisms certainly exist but may be difficult to analyze. 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 intact C-terminal GENL peptide. This could 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 showed transcriptional deregulation in human primary cancer samples as well, especially in leukemia (Fig. 3B). SFK members lacking the

ability to bind to PDZ proteins, such as Fgr or Hck, are typical for non-adherent hematopoietic cells. Thus, non-adherent hematopoietic cells resemble metastatic alterations of adherent cells.

One way through which c-Src(mt) might 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 [68]. A number of these motifs showed overlap with known binding sites of 17 different human TFs (Fig. 6). 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 [84]. Similarly, STAT5A may be activated by Src kinase [85]. Moreover, Src signaling has been shown to regulate the activity of the TF p300 that is endowed with histone acetyltransferase (HAT) activity, thereby contributing to epigenetic gene regulation [86]. Src kinase positively regulates the HAT activity of p300, and thereby, its ability to increase the transcription of target genes [86]. Of note, neither the three 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 likely altered through post-translational modifications involving c-Src(mt), or increased mRNA turnover.

Overall, this study defined the regulatory landscape of a specific C-terminally mutated form of the 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*, as metastasizing cells. Among known genes, we describe a number of genes not yet associated

with this metastatic behavior. They might deserve further investigation as putative biomarkers of invasive cells or 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 Vetenary Authority of Zurich in accordance to the guidelines of the Swiss Animal Protection Law, which strictly follows the principles declared in the Basel Declaration, namely conducting research in animals according to the International Council for Laboratory Animal Science (ICLAS). All expression data of human tumors and tissues analyzed in this study was retrieved from the ONCOMINE public domain database [43].

Cells

MCF-10A cells (ATCC-CRL-10317) were obtained from the American Type Culture Collection and grown in complete growth medium [DMEM/F12 (Gibco, Grand Island, NY, USA) with 2 mM L-glutamine supplemented with 20 ng/mL epidermal growth factor (Sigma-Aldrich, St. Louis, MO, USA), 100 ng/mL cholera toxin (Sigma), 10 ng/mL insulin (Sigma), 500 ng/mL hydrocortisone (Sigma) and 5% horse serum (Gibco)]. Cells were maintained at 37°C with 5% CO_2.

Lentiviral constructs

The HA-tagged Src constructs are described elsewhere [17-19]. Lentiviral particles for TetR-inducible expression of c-Src(wt) (LVP-Src-wt) and c-Src(mt) (LVP-Src-mt), as well as a short nonsense construct (LVP-mock) conferred resistance to Blasticidin S. LVP017-Neo allows constitutive expression of TetR and confers resistance to G418 (Fig. 1C). Lentiviral

particles were prepared by AMS Biotechnology (Abingdon, United Kingdom). 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 at a multiplicity of infection (MOI) of ten and lentiviral expression particles (MOI = 3) for 72h. Then, medium was replaced with fresh complete growth medium containing 60 μg/mL G418 (Gibco) and 3.5 μg/mL Blasticidin S (Invitrogen, Carlsbad, CA, USA) for 14 days to select double-positive cells. Untransfected control cells were killed within 14 days when exposed to either 60 μg/mL G418 or 3.5 μg/mL Bsd. Double-positive cells were maintained in complete growth medium with 30 μg/mL G418 and 1.75 μg/mL 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 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 Blot. For the subsequent gene expression analyses we selected cell populations that showed similar levels of c-Src(wt) and c-Src(mt) expression at physiological levels (Fig. 1D).

Preparation of total RNA

Cells were grown to ~70% confluence in complete medium supplemented with 30 μg/mL G418 and 1.75 μg/mL Bsd. Tetracycline (Sigma-Aldrich) was added to the culture medium and cells were incubated for 24h. Then, total RNA was extracted with Trizol (Invitrogen) according to the manufacturer's recommendations. The RNA was precipitated at -80°C and pelleted by centrifugation at 12,000 x g for 10 minutes. The pellets were washed with 75% ethanol, air-dried and solubilized in nuclease-free water. Remaining DNA was removed with the RNase-free DNase set from Qiagen according to the manufacturer's protocol. Then, the

solution was subjected to phenol-chloroform extraction. RNA integrity was assessed by gel electrophoresis using 300 ng of total RNA per lane in a 1% (w/v) agarose in 1x TAE gel. The gel was run at 130 V for 45 min. and stained with ethidium bromide. As size marker the GeneRuler 1kb DNA Ladder (Fermentas, Waltham, MA, USA) was used.

Primer design and synthesis

Primers for quantitative Real-Time PCR (qRT-PCR) were designed such that PCR products were smaller than 300 bp and intron-spanning, using the primer3 program (http://primer3.wi.mit.edu) [87] and cDNA sequences retrieved from the UCSC Genome Browser and the current release of the human genome [88]. The GAPDH primer pair has been described previously [89]. The forward primer for mu SRC_HA is situated within the HA tag of c-Src constructs, c-Src(wt) and c-Src(mt), to avoid amplification of endogenous c-Src expressed by MCF-10A cells. Primers were synthesized by Metabion (Martinsried, Germany). For a complete list of primer sequences used in this study refer to Table 1.

Quantitative real-time PCR

Total RNA preparations were reverse-transcribed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosciences, Foster City, CA, USA) according to the manufacturer's recommendations. RT-products were quantified with the 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 Blots

Cells were resuspended in buffer A (10 mM HEPES, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1 mM DTT, protease inhibitor cocktail) and layzed mechanically. After centrifugation for 5 min at 2500 x g the pellet then was subjected to DNase I digestion (10 mM HEPES, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.2U DNase I, 1 mM DTT, protease inhibitor cocktail) and

centrifuged again as described before. The resulting supernatant was combined with the first one and the total protein concentration determined using the Bradford method (Sigma-Aldrich) following the manufacturer's instructions. 30 µg of lysate were subjected to SDS-polyacrylamide gel electrophoresis (PAGE) and immunoblotting. Primary antibodies used in the experiment were: anti-Src clone GD11 (Merck Millipore, Billerica, MA, USA), and anti-GAPDH (Cat.# AM4300, Ambion, Carlsbad, CA, USA). The secondary antibody was Peroxidase-AffiniPure Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch, West Grove, PA, USA).

RNA sequencing

1 µg of total RNA preparations of the four cell lines were subjected to cDNA library preparation with 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 with a 100 bp paired-end multiplex run, yielding 99,100,500 (naïve MCF-10A), 46,988,060 (mock), 38,071,856 (c-Src(wt)) and 56,106,530 (c-Src(mt)) quality control-passed reads.

Sequence read alignment

The Illumina paired-end sequence reads were mapped to the UCSC hg19 human reference genome using the Bowtie read aligner (0.12.8 release, default parameters) [21].

Gene expression and determination of differential expression

Via the Illumina RNA-Seq technology, gene expression in four different cell lines was measured:

(1) non-transfected cells (naïve MCF-10A)

(2) cells transfected with the mock lentiviral vector

(3) cells transfected with the c-Src(wt) lentiviral vector

(4) cells transfected with the c-Src(mt) lentiviral vector

RNA expression levels were determined with 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 was then fed into DESeq that 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 (DE) of transcripts comparing 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 with 'nbinomTest(cds, "cell line A", "cell line B"). Corresponding genes with an FDR-adjusted q value ≤ 0.05 in their fold change were considered as differentially expressed. For any given gene of the comparison cell line A vs. cell line B, if transcripts were detectable in A, but non-detectable in B, the reported FC value is "Inf". *Vice versa*, the reported FC value is "-Inf".

Gene Ontology overrepresentation

The ConsensusPathDB [33] 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 done using the FDR (false discovery rate) method that resulted in q values.

Pathway overrepresentation

Overrepresented pathways involving the candidate genes were accumulated *via* the interface of the ConsensusPathDB (Release 25) [33], a pathway database that integrates interaction data from currently 31 public resources. p and q values were calculated analogously to 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 x 10^5 MCF-10A cells or c-Src(wt), c-Src(mt) or mock cells. Lung tissues were evaluated after 42 days. 1 µg/mL of Tet was added to the drinking water of all mice to induce transgene expression in engineered cell lines.

Immunohistochemistry

Lung tissues were fixed with 4% paraformaldehyde and embedded in paraffin. Paraffin sections (2 µm) were either stained with haematoxylin / eosin (H&E) or automated immunohistochemistry staining with rabbit anti-MMP2 (Abcam, Cambridge, UK), rabbit anti-Fibronectin (Abcam), mouse anti-alpha Adducin (Abcam), rabbit anti-Plectin (Abcam), rabbit anti-Phospho-S6 Ribosomal Protein (Ser235/236) (Cell Signaling Technology, Danvers, MA, USA), rabbit-anti-rat HRP, anti-rabbit-poly HRP and anti-mouse HRP (Leica Bond Polymer Refine Detection DS9800). Image acquisition was performed on DotSlide BX51 (Olympus, Tokyo, Japan), SCN400 (Leica, Wetzlar, Germany), Axio Z1 (Zeiss, Oberkochen, Germany) or BX53 (Olympus) microscopes.

Literature mining

We aimed to find genes that can be associated to cellular migration characteristics but have so far not been dealt with in that 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 the genes' co-occurrence 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.

Here, P_{ij} is the frequency of co-occurrence of the lexicon term j 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. It is thus of particular interest to further analyze those genes that have a low or negative evidence score. 71 genes showed no evidence.

Cancer gene expression profiles

Gene set enrichment analysis was performed on the basis of publicly available cancer gene expression datasets from the ONCOMINE database [43]. We looked for cancer datasets that showed similar expression profiles to our c-Src(mt)-expressing cells vs. mock cells. Focus was laid on studies that compared different stages of cancer tissue, i.e. sets marked with NP for ‘normal tissue vs. cancer’ and CP for ‘cancer progression vs. primary tumor’. At the state of writing there were 112 of such data sets to be found, all of which were considered. Ideally, those sets containing data for normal cancer tissue as well as metastatic tissue show a

significant overlap of differentially expressed genes when compared to our set of c-Src(mt) effector genes. To calculate the significance of overlap of our effector genes with the gene sets from the ONCOMINE database, we performed Fisher's exact tests and corrected for multiple testing using the FDR 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 with the Amadeus platform [68]. Standard parameters were used with motif lengths of eight to twelve basepairs and the hg19 release of the human genome. The output from Amadeus is shown in Fig. S1. Human transcription factor binding sites from TRANSFAC [90] that showed similarity to the identified motifs were chosen for further investigation. The respective position-specific weight matrices (PSSMs) were retrieved from the Cistrome database on cistrome.dfci.harvard.edu [91]. Sequence logos were computed from PSSMs with WebLogo 3 on weblogo.threeplusone.com [92].

Acknowledgements

We like to thank Dr. Jochen Heinrich (University of Zurich) for help with generating the c-Src expression constructs, Nada Kumer (Max Planck Institute for molecular Genetics) for excellent technical assistance, Dr. Algirdas Ziogas (University Hospital Zurich) for help with some of the experiments, and Profs. Peter H. Seeberger (Max Planck Institute of Colloids and Interfaces) 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 BMBF (grants 0315428A and 0316190A to RH) and the Max Planck Society for financial support. This work was supported by private funding (KM). MH was supported by the Helmholtz Alliance PCCC, an ERC Starting grant

(LiverCancerMechanism), the Helmholtz Foundation, the SFBTR36 and the Hofschneider Foundation. MRS was supported by the Volkswagenstiftung (Lichtenberg program). The authors declare no conflicts of interest.

Author contributions

FB performed the *in vitro* experiments, generated stable cell lines and conducted qRT-PCR measurements. CH, RH and MK conducted 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 conducted 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-kappaB, 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, Chang I, Deng G & Dahiya R (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. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM & Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29.
31. Dranoff G (2004) Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer* **4**, 11-22.
32. Lai SL, Chien AJ & Moon RT (2009) Wnt/Fz signaling and the cytoskeleton: potential roles in tumorigenesis. *Cell Res* **19**, 532-545.
33. Kamburov A, Stelzl U, Lehrach H & Herwig R (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res* **41(Database issue)**, D793-800.

34. 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.
35. 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.
36. El Behi M, Krumeich S, Lodillinsky C, Kamoun A, Tibaldi L, Sugano G, De Reynies A, Chapeaublanc E, Laplanche A, Lebret T, Allory Y, Radvanyi F, Lantz O, Eiján AM, Bernard-Pierrot I & Théry C (2013) An essential role for decorin in bladder cancer invasiveness. *EMBO Mol Med* **5**, 1835-1851.
37. 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.
38. 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.
39. 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.
40. 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 alpha-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.
41. 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.

42. David CJ & Manley JL (2010) Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* **24**, 2343-2364.
43. 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.
44. Irby RB & Yeatman TJ (2000) Role of Src expression and activation in human cancer. *Oncogene* **19**, 5636-5642.
45. Edwards J (2010) Src kinase inhibitors: an emerging therapeutic treatment option for prostate cancer. *Expert Opin Investig Drugs* **19**, 605-614.
46. Qayyum T, McArdle PA, Lamb GW, Jordan F, Orange C, Seywright M, Horgan PG, Jones RJ, Oades G, Aitchison MA & Edwards J (2012) Expression and prognostic significance of Src family members in renal clear cell carcinoma. *Br J Cancer* **107**, 856-863.
47. Aleshin A & Finn RS (2010) SRC: a century of science brought to the clinic. *Neoplasia* 2010, **12**, 599-607.
48. 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.
49. 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.

50. Yang S, Zhang JJ & Huang XY (2012) Mouse models for tumor metastasis. *Methods Mol Biol* **928**, 221-228.
51. Barnard GA (1945) A new test for 2×2 tables. *Nature* **156**, 177.
52. Barnard GA (1947) Significance tests for 2×2 tables. *Biometrika* **34**, 123-138.
53. Malek RL, Irby RB, Guo QM, Lee K, Wong S, He M, Tsai J, Frank B, Liu ET, Quackenbush J, Jove R, Yeatman TJ & Lee NH (2002) Identification of Src transformation fingerprint in human colon cancer. *Oncogene* **21**, 7256-7265.
54. Dufner A & Thomas G (1999) Ribosomal S6 kinase signaling and the control of translation. *Exp Cell Res* **253**, 100-109.
55. Ruggero D (2013) Translational control in cancer etiology. *Cold Spring Harb Perspect Biol* **5**, pii:a012336.
56. Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ, Wang S, Ren P, Martin M, Jessen K, Feldman ME, Weissman JS, Shokat KM, Rommel C & Ruggero D (2012) The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55-61.
57. Schlaepfer DD, Mitra SK & Ilic D (2004) Control of motile and invasive cell phenotypes by focal adhesion kinase. *Biochim Biophys Acta* **1692**, 77-102.
58. Zhao J & Guan JL (2009) Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev* **28**, 35-49.
59. 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.
60. Hsia DA, Mitra SK, Hauck CR, Streblow DN, Nelson JA, Ilic D, Huang S, Li E, Nemerow GR, Leng J, Spencer KS, Cheresh DA & Schlaepfer DD (2003) Differential regulation of cell motility and invasion by FAK. *J Cell Biol* **160**, 753-767.

61. 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.
62. 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.
63. Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646-674.
64. Vogelstein B & Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* **10**, 789-799.
65. 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.
66. 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.
67. 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.
68. 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.
69. 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.

70. Benbrook DM & Jones NC (1994) Different binding specificities and transactivation of variant CRE's by CREB complexes. *Nucleic Acids Res* **22**, 1463-1469.
71. Cai RL (1998) Human CART1, a paired-class homeodomain protein, activates transcription through palindromic binding sites. *Biochem Biophys Res Commun* **250**, 305-311.
72. Grandori C & Eisenman RN (1997) Myc target genes. *Trends Biochem Sci* **22**, 177-181.
73. 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.
74. Ko LJ & Engel JD (1993) DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol* **13**, 4011-4022.
75. 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.
76. Tronche F & Yaniv M (1992) HNF1, a homeoprotein member of the hepatic transcription regulatory network. *Bioessays* **14**, 579-587.
77. Morris JF, Hromas R & Rauscher FJ 3rd (1994) Characterization oft he 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.
78. 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.
79. 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.

80. Schoenherr CJ, Paquette AJ & Anderson DJ (1996) Identification of potential target genes for the neuron-restrictive silencer factor. *Proc Natl Acad Sci U S A* **93**, 9881-9886.
81. Rikitake Y & Moran E (1992) DNA-binding properties of the E1A-associated 300-kilodalton protein. *Mol Cell Biol* **12**, 2826-2836.
82. 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.
83. 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.
84. 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.
85. 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.
86. 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.
87. Rozen S & Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386.
88. 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.

89. 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.
90. 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.
91. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y, Pape UJ, Poidinger M, Chen Y, Yeung K, Brown M, Turpaz Y & Liu XS (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* **12**, R83.
92. Crooks GE, Hon G, Chandonia JM & Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* **14**, 1188-1190.

Table 1

Primers used in this study

Gene	Forward primer (5‘ to 3‘)	Reverse primer (5‘ to 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

Supporting information

The following supplementary material is available:

Table S1. List of 435 DE genes comparing c-Src(mt) vs. mock cells by RNA-Seq (c-Src(mt) effector genes).

Table S2. Gene Ontology overrepresentation analysis of c-Src(mt) effector genes.

Table S3. Pathway overrepresentation analysis of c-Src(mt) effector genes.

Table S4. Gene Ontology overrepresentation analysis of non-metastasis-associated c-Src(mt) effector genes.

Table S5. ONCOMINE analysis.

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

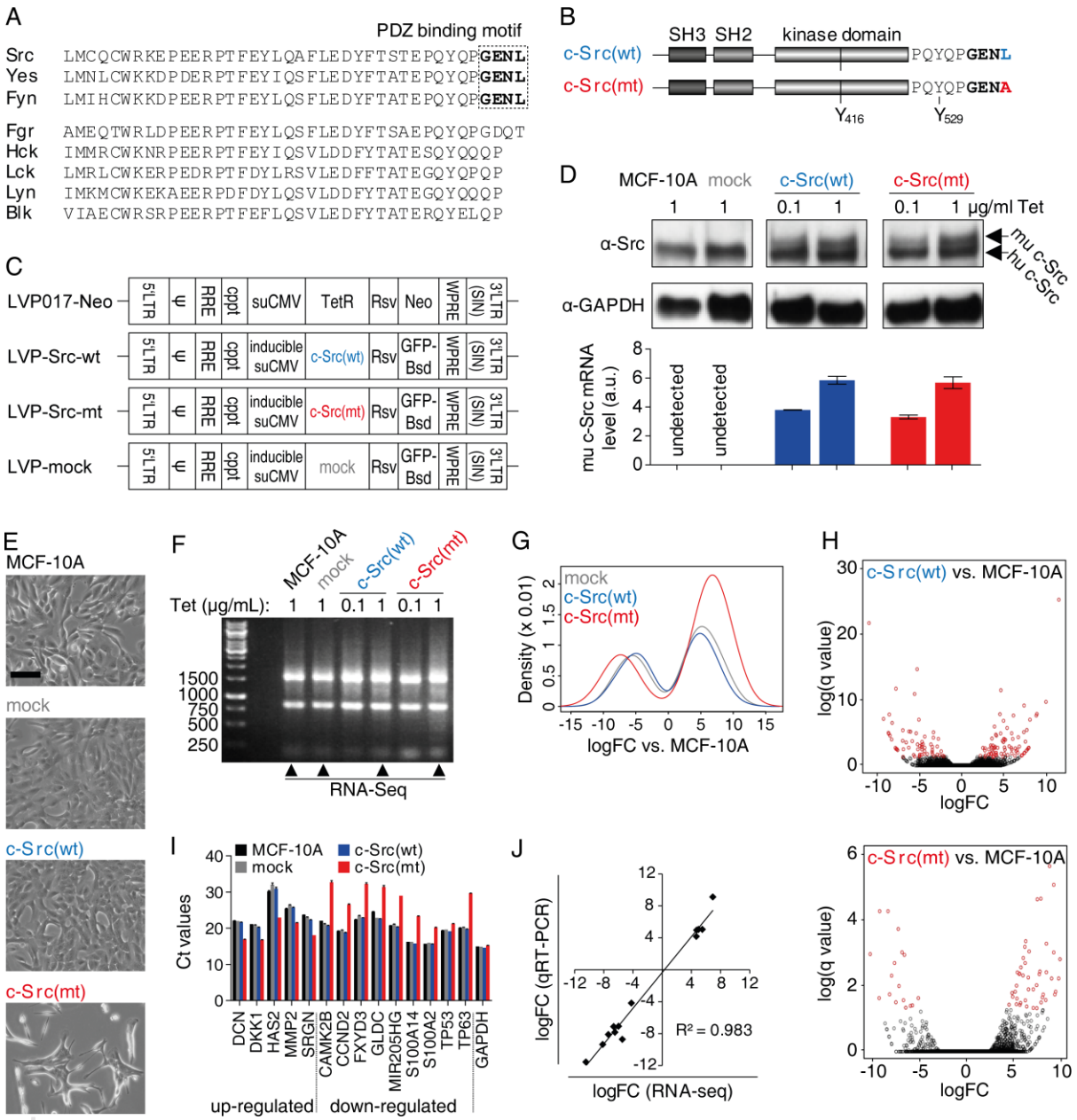


Fig. 1. Establishment of a cellular model of mutant c-Src. **(A)** Lentiviral vectors used in this study. The Neo gene confers resistance to G418. The Bsd gene, fused to the green fluorescent protein (GFP) gene, confers resistance to Blasticidin S. LTR, long terminal repeat; Ψ, psi packaging sequence; RRE, Rev-responsive element; cppt, central polypurine tract; suCMV, 'super' cytomegalovirus promoter; TetR, tetracycline regulatory gene; Rsv, Rous Sarcoma virus promoter; WPRE, Woodchuck Hepatitis posttranscriptional regulatory element; SIN, self-inactivating. The suCMV promoter is Tetracycline-inducible. 'mock' is a randomized

sequence of about 100 nucleotides. **(B)** Alignment of C-terminal moieties of human Src family kinases (SFKs). A conserved binding motif for PDZ domain-containing tumor suppressor proteins is highlighted. (**C**) Domain organization of c-Src constructs used in this study. (**D**) Detection of transgene expression in engineered MCF-10A cells by Western Blot (upper part) and qRT-PCR (lower part). The α-Src antibody binds to both human (hu) and murine (mu) c-Src proteins. The mu c-Src constructs (61.3 kD) are bigger than the endogenous hu c-Src (59.8 kDa), giving rise to a second band above that of hu c-Src in the c-Src(wt) and c-Src(mt)-expressing cells. GAPDH **is** 36 kDa. qRT-PCR data shows mean ± SD of two independent measurements. (**E**) Inverted light microscopic images of the four cell lines after exposure to 1 μg/ml Tet for 24h. The scale bar is approximately 100 μm. (**F)** Agarose gel of RNA preparations used for RNA-Seq and qRT-PCR. Sizes of the marker in basepairs are indicated on the left. (**G**) Density plot showing logarithmic fold change (logFC) values of statistically significant ($q \leq 0.05$) differentially expressed (DE) genes of mock, c-Src(wt) and c-Src(mt) compared with naïve 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. The bars show mean + SD of two independent measurements. (**J**) Correlation between RNA sequencing data and qRT-PCR measurements of 13 statistically significantly DE genes of the comparison c-Src(mt) vs. mock cells.

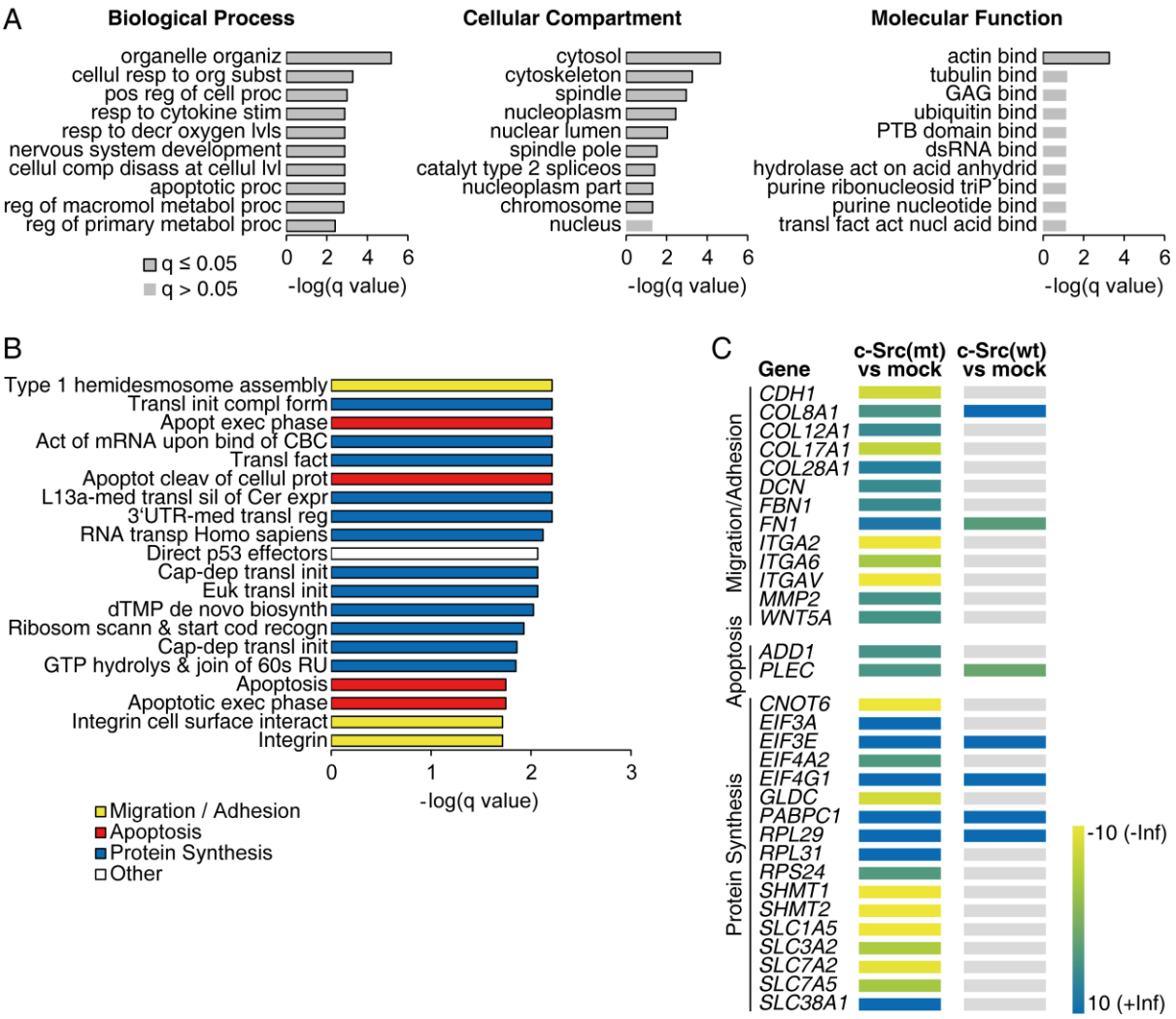
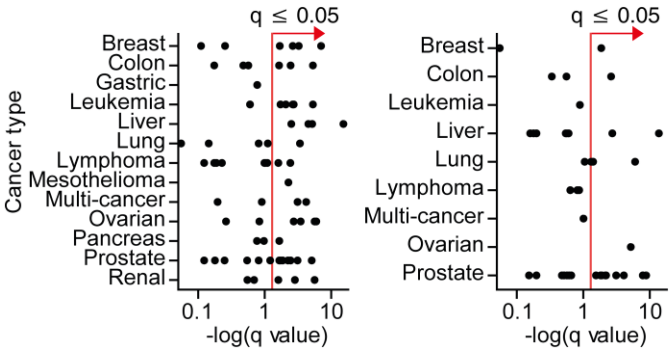


Fig. 2. Functional analysis of c-Src(mt) effector genes. (A) Gene Ontology (GO) enrichment analysis. Displayed are the ten level 4 GO terms in the indicated categories with the lowest q values (see Table S2 for a complete list and corresponding gene lists). (B) Pathway analysis. Displayed are the twenty pathways with the lowest q values (all are $q \leq 0.05$). Redundant pathways have been removed. Color code indicates pathways associated with different cellular processes. The complete list of pathways can be found in Table S3. (C) Selected genes and their relative expression levels in c-Src(mt)- and c-Src(wt)-expressing cells with respect to mock cells. logFC values ($q \leq 0.05$) are indicated by color code. Non-significant ($q > 0.05$) gene expression is shown in grey.

A



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) The q values of gene expression overlaps with cancer vs. normal tissue pairs (left) and with metastatic vs. localized cancer pairs (right) are shown. A complete list of primary cancer samples and gene lists can be found in Table **S5**. (B) Non-migration-associated c-Src(mt) effector genes with significant deregulation in cancer vs. normal tissue pairs are shown.

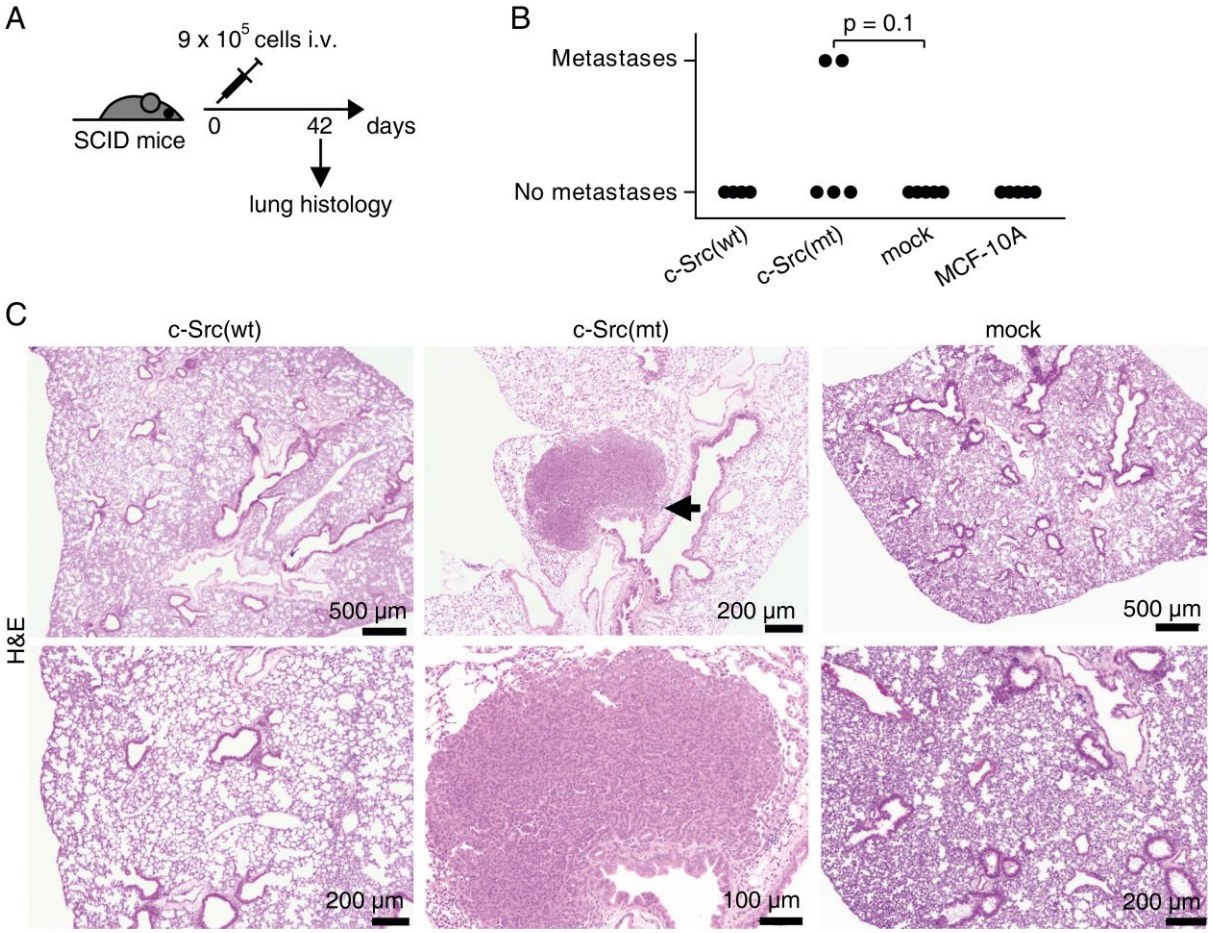


Fig. 4. c-Src(mt)-expressing cells exhibit elevated invasive potential *in vivo*. (A) SCID mice were injected i.v. with one of the four different cell lines and sacrificed 42 days later for histological analyses of their lungs. (B) Determination of lung metastases in mice injected with c-Src(mt)-expressing cells as visualized by H&E (Hematoxylin & Eosin) staining shown in C. p value was determined by Barnard's test [51,52]. (C) Macroscopic analysis of lungs. The arrow indicates a representative metastasis found in a mouse injected with c-Src(mt)-expressing cells.

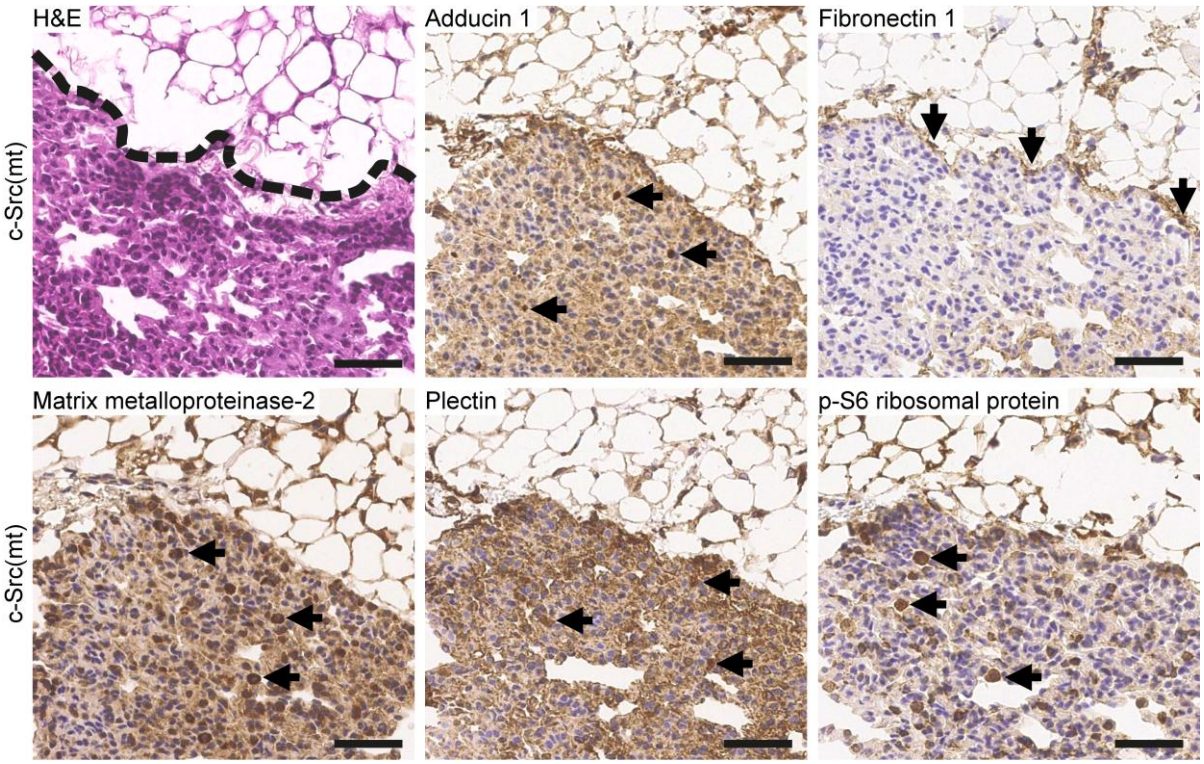


Fig. 5. Protein expression of selected effector genes in metastases of c-Src(mt)-expressing cells detected by immunohistochemistry. The border between the metastasis and normal lung tissue is indicated in the H&E-stained panel by a dashed line. Arrows indicate representative cells stained positive for the indicated proteins. Scale bars, 50 μm.

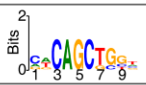
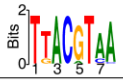
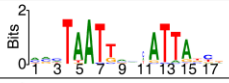
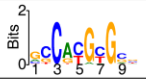
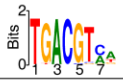
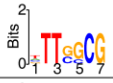
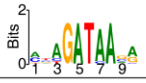
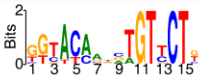
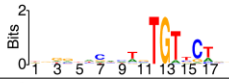
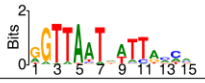
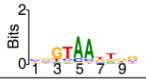
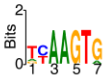
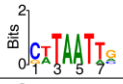
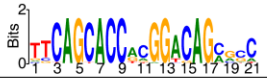
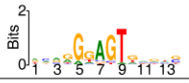
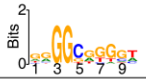
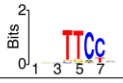
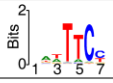
Transcription factor	Gene	Matrix ID	Sequence logo	Reference
AP-4 (Activator protein protein 4)	*TFAP*	M00176-V$AP4_Q6		[69]
ATF2 (Activating transcription factor 2)	*ATF2*	M00040-V$CREBP1_01		[70]
CART1 (Cartilage homeoprotein 1)	*ALX1*	M00416-V$CART1_01		[71]
c-Myc (Myelocystomatosis viral oncogene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

Fig. 6. Human transcription factor binding sites enriched in the promoter regions of c-Src(mt) effector genes. For details, refer to Fig. S1.