SUPPLEMENTARY MATERIALS AND METHODS

STR assay

DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). DNA samples were quantified using NanoDrop ND-2000 and Qubit with DNA measurement solution (Thermo Fisher Scientific) to ensure quality. Samples were submitted to the University of Illinois Chicago Center for Genomic Research-DNA services Facility for STR profiling. Results were compared to known STR profiles for NSCLC cell lines available on the ATCC website.

	HCC4006		HCC4006 GeR		HCC4006 OR		HCC4006 shNT #2		HCC4006 shCXCR7 509 #2		HCC4006 GeR shNT #2		HCC4006 GeR shCXCR7 509 #2		HCC4006 GeSBR shNT #2		HCC4006 GeSBR shCXCR7 509 #2	
TH01	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
D21S11	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31
D5S818	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
D13S317	11	12	12	12	12	12	11	12	11	12	11	12	11	12	11	12	11	12
D7S820	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12
D16S539	11	12	12	12	12	12	11	12	12	12	12	12	11	12	11	12	11	12
CSF1PO	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
AMEL	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
vWA	16	17	16	17	16	17	16	17	16	17	16	17	16	17	16	17	16	17
TPOX	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9
Remarks																		

	NCI-H1975		NCI-H1975 shNT		NCI-H1975 shCXCR7 509		HCC827		HCC827 shNT		HCC827 shCXCR7 509		HCC827 pBABE		HCC827 pBABE-MEKDD	
TH01	7	7	7	7	7	7	6	6	6	6	6	6	6	6	6	6
D21S11	28	28	28	28	28	28	31	31	31	31	31	31	31	31	31	31
D5S818	11	12	11	12	11	12	12	12	12	12	12	12	12	12	12	12
D13S317	10	10	10	13	10	13	9	9	9	9	9	9	9	9	9	9
D7S820	8	8	8	8	8	8	11	12	11	12	11	12	11	12	11	12
D16S539	9	12	9	12	9	12	12	12	12	12	12	12	12	12	12	12
CSF1PO	12	12	12	12	12	12	11	11	11	11	11	11	11	11	11	11
AMEL	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
vWA	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
TPOX	8	11	8	11	8	11	8	8	8	8	8	8	8	8	8	8

	HCC4006 shNT (OR)		HCC4006 shCXCR7 509 (OR)		HCC4006 shCXCR7 512 (OR)		HCC4006 shNT (CXCR4)		HCC4006 shCXCR4 4054		HCC4006 GeR shNT (CXCR4)		HCC4006 GeR shCXCR4 4054	
TH01	7	7	7	7	7	7	7	7	7	7	7	7	7	7
D21S11	31	31	31	31	31	31	31	31	31	31	31	31	31	31
D5S818	12	12	12	12	12	12	12	12	12	12	12	12	12	12
D13S317	11	12	11	12	11	12	11	12	11	12	11	12	11	12
D7S820	9	12	9	12	9	12	9	12	9	12	9	12	9	12
D16S539	11	12	11	12	11	12	11	12	11	12	11	12	11	12
CSF1PO	10	10	10	10	10	10	10	10	10	10	10	10	10	10
AMEL	X	X	X	X	X	X	X	X	X	X	X	X	X	X
vWA	16	17	16	17	16	17	16	17	16	17	16	17	16	17
TPOX	8	9	8	9	8	9	8	9	8	9	8	9	8	9

List of Antibodies

Target	Application	Vendor	Catalog Number
β-Actin	WB	CST	4967
β-Arrestin 1/2 (D24H9)	WB	CST	4674
β-Tubulin (9F3)	WB	CST	2128
Akt (pan)(11E7)	WB	CST	4685
Bim (C34C5)	WB	CST	2933
CD44	WB	CST	3578
Cleaved Caspase-3 (Asp175)	WB	CST	9661
E-Cadherin (24E10)	WB	CST	3195
EGF Receptor (D38B1) XP	WB	CST	4267
HER2/ErbB2 (29D8)	WB	CST	2165
HER3/ErbB3 (1B2E)	WB	CST	4754
HSP90	WB	CST	4874
MEK1/2 (47E6)	WB	CST	9126
Met	WB	CST	4560
N-Cadherin (D4R1H) XP	WB	CST	13116
p44/42 MAPK (Erk1/2)	WB	CST	9102
PARP	WB	CST	9542
Phospho-Akt (Ser473)(D9E) XP	WB	CST	4060
Phospho-Bim (Ser69)	WB	CST	4581
Phospho-EGF Receptor (Tyr1068)(D7A5) XP	WB	CST	3777
Phospho-HER2/ErbB2 (Tyr1221/1222)(6B12)	WB	CST	2243
Phospho-HER3/ErbB3 (Tyr1289)(21D3)	WB	CST	4791

Phospho-c-Met (Tyr1230, Tyr1234, Tyr1235) Polyclonal Antibody	WB	Invitrogen	44-888G
Phospho-p44/42 MAPK (Erk1/2)(Thr202/Tyr204)	WB	CST	9101
Phospho-S6 Ribosomal Protein (Ser240-244)	WB	CST	2215
S6 Ribosomal Protein (5G10)	WB	CST	2217
Snail (C15D3)	WB	CST	3879
TCF8/ZEB1 (D80D3)	WB	CST	3396
twist (H-81)	WB	SCBT	sc-15393
TWIST1	WB	CST	46702
V5 Tag Monoclonal Antibody, HRP	WB	Invitrogen	R961-25
Vimentin (D21H3) XP	WB	CST	5741
Vinculin (E1E9V) XP	WB	CST	13901
Anti-Mouse IgG, HRP-linked	WB	CST	7076
Anti-Rabbit IgG, HRP-linked	WB	CST	7074
Anti-GPCR RDC1 (EPR9321)	WB	abcam	ab138509
Anti-ZEB2	WB	abcam	ab138222
CXCR7 (C1C2)	WB/IF	Genetex	GTX100027
Alexa Fluor 546 F (ab')2 fragment of goat anti-rabbit IgG (H+L)	IF	Invitrogen	A11071
ProLong Gold antifade reagent with DAPI	IF	Invitrogen	P36931
Human CXCR4 Alexa Fluor 700-conjugated	FC	R&D Systems	FAB170N
Human CXCR4 PE-conjugated	FC	R&D Systems	FAB170P
Human/Mouse CXCR7/RDC-1 PE-conjugated	FC	R&D Systems	FAB227P
DAPI Stain	FC	CST	4083S
RDC1-CXCR7	IHC	Abcam	ab38089
CDH2	IHC	Epitomic	2019-1
Vimentin	IHC	SCBT	SC-7557
Phospho-p44/42 MAPK (Erk1/2)(Thr202/Tyr204)	IHC	CST	4370
Phospho-Akt (Ser473)(D9E) XP	IHC	CST	4060
Beta-arrestin	IHC	SCBT	SC-74591

Abbreviations: WB: Western blot, IF: Immunofluorescence, FC: Flow Cytometry, IHC: Immunohistochemistry, CST: Cell Signaling Technology, SCBT: Santa Cruz Biotechnology.

Luminex assay

Bio-Rad 40-plex Luminex assay kit was used for the quantification of cytokines and chemokines. NSCLC cells were seeded in 60mm dishes and grown to 70% confluency on the day of assay. Cells were treated with 500nM gefitinib for indicated time and supernatant was harvested. Cell debris was removed by centrifugation and resulting supernatants were subject to Luminex assay immediately.

Several aliquots were made and snap-frozen in LN2, and kept in -80°C freezer until use. The volume of supernatant and the number of the cells or the amount of protein upon cell lysis were recorded at the time of the harvest to normalize the target quantity. The resulting values were log2 converted and the heatmap was generated using Gene-E and Morpheus (https://software.broadinstitute.org/morpheus/).

cDNA

pLX304 was a gift from David Root (Addgene plasmid # 25890 ; http://n2t.net/addgene:25890 ; RRID:Addgene_25890) (1). pBABE-puro was a gift from Hartmut Land & Jay Morgenstern & Bob Weinberg (Addgene plasmid # 1764 ; http://n2t.net/addgene:1764 ; RRID:Addgene_1764) (2) pDONR223-MEKDD was also a gift from David Root (Addgene plasmid # 31202; http://n2t.net/addgene:31202 ; RRID:Addgene_31202) (1). LacZ in pLX304 was a gift from William Hahn (Addgene plasmid # 42560 ; http://n2t.net/addgene:42560 ; RRID:Addgene_42560) (3). cDNAs for CXCR7 and SNAIL were obtained from DNASU, Arizona State University Biodesign Institute, and were cloned to pLX304 vector.

shRNA knockdown and sequence used for study

shRNA knockdown and viral transduction and siRNA transfection Lentiviral production and infection were performed as reported previously (2). Using TransIT-LT1 transfection reagent (Mirus), the 293LTV cell line (Cell Biolabs) was transfected with pLKO.1 constructs and packaging plasmids, pCMV-ΔR8.2dvpr and pCMV-VSV-G. Lentiviral supernatants were collected and quality of viral supernatant was routinely tested with Lenti-X Go Stix (Clontech). The supernatants were applied to target cells. Following puromycin selection, cell viability assays were performed 7 days post-infection.

Target	RNAi Consortium Number	Sequence	Remarks
Non-Target	N/A	5’ – GCGCGATAGCGCTAATAATTT – 3’	Sigma SHC-002
EGFR	TRCN0000010329	5’ – AGAATGTGGAATACCTAAGG – 3’	
CDH1	TRCN0000039664	5’-AATGCCATCGTTGTTCACTGG-3’	
CXCR7 #1	TRCN0000014509	5'-CCGGAAGATCATCTTCTCCTA-3'	
CXCR7 #2	TRCN0000014512	5'-CGCTCTCCTTCATTTACATTT-3'	

CXCR4	TRCN0000004054	5'-CTTTGTCATCACGCTTCCCTT-3'	

sgRNA transfection

Target sequences for CRISPR interference were designed using the sgRNA designer (https://dharmacon.horizondiscovery.com/gene-editing/crispr-cas9/crispr-design-tool/). A non-targeting sequence was used as a scramble sgRNA for control. Sequences were as follows:

	Target sequence (5'-3')
Scramble sgRNA	GGTCTTCGAGAAGACCT
CXCR7 sgRNA1	CTCTTCGACTACTCAGAGCC
CXCR7 sgRNA2	TGAAGGAGAGCGTGTAGAGC
CXCR7 sgRNA3	CTGGTCACTGGACGCCGAGA

CRISPR vectors were cloned as previously described (Ran et al., 2013). Briefly, HCC4006-GeR cells were transduced with pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmids (Addgene plasmid #62988) and selected using 5 μg/ml puromycin.

Genome engineering using the CRISPR-Cas9 system. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. *Nat Protoc. 2013 Nov;8(11):2281-308. doi: 10.1038/nprot.2013.143. Epub 2013 Oct 24.* 10.1038/nprot.2013.143 PubMed 24157548

siRNA Transfection

siRNA (pooled oligos) was transfected into the target cells using TransIT-siQUEST reagents (Mirus) according to manufacturer's instructions. The catalog numbers for the siRNAs used are as follows.

siRNA	**Catalog Number**
Control siRNA-A	Santa Cruz Biotechnology sc-37007
β-Arrestin-1 siRNA	Santa Cruz Biotechnology sc-29741
β-Arrestin-2 siRNA	Santa Cruz Biotechnology sc-29208

TaqMan RT-PCR

Probe Catalog Number	**Target**	**Brand**	**Fluorochrome**
Hs00664172_s1	ACKR3	LifeTech	FAM
Hs00607978_s1	CXCR4	LifeTech	FAM
Hs00195591_m1	SNAI1	LifeTech	FAM
Hs00361186_m1	TWIST1	LifeTech	FAM
Hs00232783_m1	ZEB1	LifeTech	FAM
Hs00207691_m1	ZEB2	LifeTech	FAM
GUSB CONTROL MIX	AB 4325789	LifeTech	VIC

Analysis of Microarray Gene Expression Data

Gene expression profiles of indicated experimental conditions were generated using the genome-wide Illumina Human-HT-12 v4 Expression BeadChip platform (Genomics Core Facility, University of Valencia). Illumina GenomeStudio v2011.1 was used for the export of summarized probe intensities. Next, the expression data were imported into Genepattern suite for quantile normalization and gene annotation. After normalization, probes representing the same genes were collapsed into a single value, and differential expression analysis using the t-test statistic and the false discovery rate (FDR) adjustment was carried out to find differentially expressed transcripts (http://www.broadinstitute.org/cancer/software/genepattern).

GSEA (http://www.broadinstitute.org/gsea) was used to determine the gene set enrichment of experimental signatures related with EMT, TGFβ, and EGFR TKI resistance obtained from the MsigDB (http://www.broadinstitute.org/gsea/msigdb). GSEA estimates whether the members of a given gene signature are found at the top or bottom of a list of genes ranked by signal-to-noise ratio, indicating they are associated with a specific phenotype (EGFR TKI resistance), rather than being distributed uniformly or randomly across the gene list. An enrichment score (ES) is calculated to quantify the degree to which a gene signature is overrepresented at the top or bottom of the entire ranked list. GSEA normalizes the ES for each gene set to account for the variation in set sizes, yielding a normalized enrichment score (NES) and a false discovery rate (FDR). The FDR gives an estimate of the probability that a set with a given NES represents a false positive finding; it is computed by comparing the tails of the observed and empirical phenotype-based permutated null distributions for the NES. The original expression data has been submitted to Gene Expression Omnibus (GEO, accession numbers GSE123031 and GSE123066).

EGFR-mutant NSCLC patients and IHC

We analyzed formalin-fixed paraffin-embedded (FFPE) tumor samples from a retrospective cohort of advanced EGFR-mutant NSCLC patients diagnosed between July 2012 and May 2017 in the Hospital Universitari i Politècnic La Fe and Hospital Clínico Universitario (Valencia, Spain). Those

samples and data from Hospital Clínico Universitario were provided by the INCLIVA BioBank (PT13/0010/0004), integrated in the Valencian Biobanking Network and the Spanish National Biobanks Network and they were processed following standard operating procedures with the appropriate approval of the Ethical and Scientific Committees. Eligibility criteria included: (I) histologically confirmed NSCLC with locally advanced or metastatic disease (stages IIIB or IV); (II) presence of activating EGFR mutations; (III) ≥18 years old; (IV) Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0–2; (V) treatment with EGFR TKIs gefitinib or erlotinib and progression by RECIST criteria; and (VI) availability of paired pre- and post-TKI FFPE biopsies. Routine molecular diagnosis was performed on FFPE tissues. Five μm thick sections were macrodissected by a pathologist to select regions containing the highest proportion of tumor cells. Genomic DNA was isolated from FFPE sections using Deparaffinization Solution and QIAamp DNA Investigator Kit (Qiagen) according to manufacturer's protocol. DNA concentration was quantified by spectrophotometer using NanoDrop 2000 (ThermoFisher Scientific). Samples were tested by real time polymerase chain reaction (PCR) in a cobas z480 analyzer using the cobas EGFR Mutation Test v2 (CE-IVD; Roche Molecular Diagnostics), which can detect 42 mutations in exons 18–21 of the EGFR gene. All patient tumor samples analyzed were obtained under institutional review board–approved protocols with informed consent obtained from each patient and conducted in accordance with the Declaration of Helsinki.

All IHC procedures were performed in the Department of Pathology at Hospital Quirónsalud (Valencia, Spain). Pre and post-TKI FFPE biopsies were sectioned on slides with 5 μm thickness. Then paraffin sections were deparaffinized in xylene and rehydrated in a graded alcohol series and boiled with citrate buffer pH 6. Sections were incubated with primary antibodies specific for CXCR7 (Abcam ab38089) and Nkx2.1 (Dako, IR05661-2). For antibody detection, Envision detection system peroxidase/DAB (Dako) was used. The procedure was performed on a Dako Autostainer Link 48. Image acquisition and analyses were performed with a Leica DMD108 microscope. The CXCR7 staining score were developed by the board-certified pathologists and designated as follows: -, negative expression; +, weak expression; ++, moderate expression; +++, strong expression; ++++, very strong expression.

Surgical resected lung adenocarcinoma patients and gene expression analysis by RT-PCR

55 patients undergoing surgical resection of a primary lung tumor from June 2011 to October 2016 were included in the study, of which 42 cases were males and 13 cases females. 47 patients underwent a lobectomy resection, 5 a segmental resection, 2 a bilobectomy and 1 a pneumonectomy. The resection margin of the surgery was at least 1 cm to ensure the removal of all the tumoral tissue. The mean age of patients was 65.6 ± 9.1 years (25 cases ≤65 years old; 30 cases >65 years old). 20 patients were smokers, 19 ex-smokers and 9 never-smokers. Postoperative pathological staging included exclusively stage I. The tumor stage was classified according to the Revisions in the International System for Staging Lung Cancer. Patients with previous adjuvant therapy or with previous or synchronous neoplasms were excluded for the study. The study protocol was approved by the ethics committee of the Hospital Universitario de la Ribera and conducted in accordance with the Declaration of Helsinki.

Total RNA was isolated from fresh lung tumor and normal adjacent tissue using QIAGEN RNA Mini Kit (QIAGEN). Retrotranscription was performed to obtain cDNA using the GeneAmp RT-PCR kit (Applied Biosystems) from 1 µg of total RNA. Gene expression analysis of CXCR7 and an internal reference gene (GAPDH), was performed using TaqMan gene expression assays: Hs00664172_s1 and Hs02758991_g1a for GAPDH. PCR amplification was carried out with 2.5 µL of cDNA and each sample in triplicate. Amplification was carried in an ABI7500 instrument (Applied Biosystems). The expression levels were defined from the threshold cycle (Ct) and relative values were calculated by the 2-ΔΔCt method after normalizing expression to GAPDH and each tumor sample value to its adjacent normal tissue.

REFERENCES

1. Yang X, Boehm JS, Yang X, Salehi-Ashtiani K, Hao T, Shen Y*, et al.* A public genome-scale lentiviral expression library of human ORFs. Nat Methods **2011**;8:659-61
2. Morgenstern JP, Land H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic acids research **1990**;18:3587-96
3. Rosenbluh J, Nijhawan D, Cox AG, Li X, Neal JT, Schafer EJ*, et al.* beta-Catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. Cell **2012**;151:1457-73