Epigenetic expansion of VHL-HIF signal output drives multi-organ metastasis in renal cancer

Sakari Vanharanta[1], **Weiping Shu**[1], **Fabienne Brenet**[2], **A. Ari Hakimi**[3,4], **Adriana Heguy**[5], **Agnes Viale**[6], **Victor E. Reuter**[7], **James J.-D. Hsieh**[3], **Joseph M. Scandura**[2], and **Joan Massagué**[1,8]

[1]Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA

[2]Department of Medicine, Weill Cornell Medical College, New York, New York 10065, USA

[3]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA

[4]Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA

[5]Geoffrey Beene Translational Oncology Core Facility, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA

[6]Genomics Core Facility, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA

[7]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA

[8]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA

Abstract

Inactivation of the *von Hippel-Lindau tumor suppressor* (*VHL*) is an archetypical tumor-initiating event in clear cell renal carcinoma (ccRCC), leading to the activation of hypoxia-inducible transcription factors (HIFs). However, *VHL* mutation status in ccRCC is not correlated with clinical outcome. Here we show that during ccRCC progression, cancer cells exploit diverse epigenetic alterations to empower a branch of the VHL-HIF pathway for metastasis, and the

Correspondence: Joan Massagué, Box 116, Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065 USA, Phone: 646-888-2044, massaguj@mskcc.org.

strength of this activation is associated with poor clinical outcome. By analyzing metastatic subpopulations of VHL-deficient ccRCC cells, we discovered an epigenetically altered VHL-HIF response specific to metastatic ccRCC. Focusing on the two most prominent pro-metastatic VHL-HIF target genes, we show that liberation from PRC2-dependent repressive histone methylation (H3K27me3) activates HIF-driven CXCR4 expression in support of chemotactic cell invasion, whereas loss of DNA methylation enables HIF-driven CYTIP expression to protect cancer cells from death cytokine signals. Thus, metastasis in ccRCC is based on an epigenetically expanded output of the tumor-initiating pathway.

The drivers of metastasis in certain cancers include genes and pathways that are mechanistically independent of the oncogenic mutations driving tumor initiation[1–4]. In other cancers, however, the pathways driving carcinoma formation additionally drive metastasis. One example of this alternative paradigm is ccRCC, a tumor type in which the VHL-HIF pathway drives both tumor initiation and metastasis[5,6].

VHL is a classical gatekeeper inhibiting renal tumor initiation[7–13]. The main tumor suppressive function of VHL is its role in mediating the degradation of the hypoxia inducible factor 2 alpha (HIF2α, also know as endothelial PAS domain protein 1, EPAS1), which drives the expression of multiple target genes with tumorigenic functions[5,14–16]. Additionally, at least one HIF target gene, *chemokine (C-X-C motif) receptor 4* (*CXCR4*), is a direct mediator of metastatic colonization[1,6,17,18]. Therefore, it was previously suggested that *VHL* loss might directly lead to metastatic tumor phenotypes through HIF activation[6]. This model, however, is challenged by the striking clinical facts that most VHL-negative ccRCCs never metastasize[19], and that *VHL* mutation status does not correlate with poor disease outcome[20,21], even though CXCR4 expression does[6,22].

We used the combined power of new experimental model systems and large clinical data sets to test the hypothesis that the activation of CXCR4 and other metastatic genes downstream of VHL-HIF is enabled by epigenetic events in metastatic subpopulations of renal cancer cells.

RESULTS

ccRCC metastasis model with clinically relevant correlates

We isolated metastatic subpopulations of the VHL-deficient ccRCC cell line 786-O, originally derived from a primary tumor of an individual with widely metastatic disease (Supplementary Fig. 1a)[23]. The parental cell line contained rare clones that upon intravenous inoculation into mice were capable of forming rapidly growing metastases in the lungs, the most frequent site of ccRCC metastasis[19] (Supplementary Fig. 1b). Isolation of cells from these lesions yielded variants (786-M1A and M1B) that were ~100-fold enriched for lung colonizing activity (Fig. 1a and Supplementary Fig. 1c,d). Second generation derivatives (786-M2A and M2B) recapitulated the behavior of the 786-M1 cells (Fig. 1a and Supplementary Fig. 1c,d). The enhanced metastatic phenotype was not associated with changes in cell proliferation *in vitro* (Supplementary Fig. 1e). The 786-M1A derivatives were more metastatic to the lung also from the orthotopic site (Fig. 1b). Orthotopic tumors

formed by 786-O cells displayed a high-grade ccRCC histology with prominent epithelioid features, whereas the metastatic 786-M1A cells showed areas of epithelioid ccRCC and also areas of sarcomatoid features (Fig. 1c and Supplementary Fig. 1f). Lung metastatic nodules presented only sarcomatoid histology (Fig. 1c and Supplementary Fig. 1f). All the cell lines tested were also more aggressive in forming osteolytic bone metastasis (Fig. 1d,e and Supplementary Fig. 2a,b) and more slowly progressing invasive lesions in the brain (Fig. 1f,g and Supplementary Fig. 2c), organs commonly affected by ccRCC[19,24].

We used genome wide transcriptional profiling to identify 155 genes associated with this metastatic phenotype (Supplementary Table 1). In an unsupervised hierarchical clustering analysis of three independent human data sets comprising 758 samples (TCGA consortium, GSE2109 and GSE3538)[25], this 155 gene set was able to identify tumor subgroups that had an expression profile resembling that of the metastatic cells (Supplementary Fig. 3a–c). These data sets were derived from surgically removed untreated primary ccRCCs, with the exception of three tumors in the GSE2109 and seven tumors in the TCGA data sets that had received prior treatment. The unfiltered 155 gene set showed clinically significant clustering of human primary tumors, even though it was likely to contain gene expression events that were specific to our cell lines. Therefore, we used the GSE2109 data set to filter out genes that behaved discordantly between our experimental model and clinical samples. This step reduced the number of genes to 50 yielding a classifier (RMS50) that provided a tight correlation with clinical outcome in both the TCGA and GSE3538 cohorts (Fig. 1h and Supplementary Fig. 3d–g).

Altered VHL-HIF response in metastatic ccRCC cells

One of the most highly overexpressed genes in the metastatic cells was *CXCR4* (52-fold; Supplementary Table 1), which was of interesting for two reasons. Firstly, CXCR4 expression correlates with poor prognosis and metastasis in ccRCC[6,22,26]. Secondly, *CXCR4* is induced upon VHL loss[6,27]. We confirmed the higher expression of *CXCR4* by quantitative RT-PCR and also showed that this induction was strongly dependent on VHL loss (Fig. 2a). Furthermore, *CXCR4* mRNA levels were associated with disease progression in a cohort of mostly early stage ccRCC samples collected in our institution (Fig. 2b). The expression of *vascular endothelial growth factor A* (*VEGFA*) and *adrenomedullin* (*ADM*), both known VHL-HIF target genes[28], remained unchanged (Fig. 2c). Also, there were no differences in *VHL* mutation status (data not shown) or HIF2α protein expression (Supplementary Fig. 4a). These findings suggested that clonal derivatives of a single *VHL*-mutant cell could display significant variation in the VHL-HIF pathway transcriptional output and that this was associated with different tumor phenotypes.

To determine whether this alteration in the VHL-HIF signal output included genes other than *CXCR4*, we performed gene expression profiling. Of the 137 genes that were significantly regulated by VHL in the 786-M1A cells (Supplementary Table 2), 37 were also present in the 155 gene set (Fig. 2d). This overlap was more than what would be expected by chance alone and included both up- and down-regulated genes. Also, these 37 overlapping genes were significantly enriched in two (of the four possible) categories: 1) induced by

VHL loss and upregulated in metastatic cells or 2) repressed by VHL loss and downregulated in metastatic cells (Supplementary Fig. 4b).

As HIFs are primarily transcriptional activators[29], we focused on the 12 genes that were both upregulated upon VHL loss and had higher expression in the metastatic cells. To avoid possible confounding effects caused by tumors in which the HIF-pathway was not activated, we analyzed the expression of these VHL-regulated genes in a set of *VHL*-mutant primary ccRCC specimens[30]. *VHL* mutation status was not associated with RMS50-status (Supplementary Fig. 4c), but a subset of metastasis-associated VHL-HIF target genes, namely *CXCR4*, *cytohesin 1 interacting protein* (*CYTIP*), *SLAM family member 8* (*SLAMF8*), *chemokine (C-X-C motif) receptor 7* (*CXCR7*), and *latent transforming growth factor beta binding protein 1* (*LTBP1*) had significantly higher expression in RMS50-positive VHL-negative tumors (Supplementary Fig. 5a). Other genes, such as *VEGFA* and *ADM*, showed no difference (Supplementary Fig. 5b). We confirmed the expression of these genes by qRT-PCR in our cell line model (Fig. 2e). The expression of these genes was dependent on HIF2α (Fig. 2f). Protein analysis confirmed CXCR4 and CYTIP overexpression in the metastatic cells (Supplementary Fig. 5c,d).

To determine the generality of these findings, we utilized two additional cell line systems. One of them (RFX-631) spontaneously formed robust lung metastasis with a mixed epithelioid and sarcomatoid histology (Supplementary Fig. 6a). The other cell line (OS-RC-2) was more indolent and gave rise to metastatic OS-LM1 variants after *in vivo* selection (Supplementary Fig. 6b–d). The OS-LM1 metastatic nodules were of the classic clear cell morphology with some regions containing cells with eosinophilic cytoplasm (Supplementary Fig. 6e). The expression of all five genes was dependent on VHL loss at least in one additional model system, and *CXCR4*, *CYTIP* and *LTBP1* scored in both RFX-631 and OS-LM1 cells (Fig. 2g,h). *CXCR4*, *CYTIP*, *LTBP1* and *SLAMF8* all showed varying degrees of overexpression in OS-LM1 cells when compared to the parental OS-RC-2 cells (Supplementary Fig. 6f), confirming that their expression was more generally selected for in cells with metastatic potential.

In conclusion, the VHL-HIF pathway not only drives ccRCC initiation but also the expression of several genes that are associated with ccRCC metastasis (Fig. 2i).

Expanded HIF signal output activates mediators of metastasis

VHL reintroduction strongly inhibited 786-M1A lung colonizing activity (Fig. 3a). Among the VHL-HIF-dependent genes associated with metastasis in our model systems, CXCR4 and CXCR7 are G protein-coupled receptors for the chemokine CXCL12[1,17,18]. CYTIP is a ~40kDa intracellular protein that can enhance NFAT/AP1 transcriptional responses[31], modulate cell adhesion and migration[32–34], and support stem cell fitness[32]. LTBP1 is a large extracellular protein that regulates TGF-β activation[35]. SLAMF8 is a cell surface receptor normally expressed in activated macrophages. To prioritize our functional analysis, we focused on *CXCR4*, *CYTIP* and *LTBP1*, genes consistently associated with metastasis across all cell line systems (Figs. 2a, 2e and 2g,h). LTBP1 was not essential for lung metastasis in the 786 model (Supplementary Fig. 7a,b), possibly due to compensation by LTBP4 (Supplementary Table 1) or other LTBP family members, whereas knockdown of CYTIP or

CXCR4 inhibited by 10-fold the lung colonizing activity of 786-M1A cells (Fig. 3b and Supplementary Fig. 8a,b). Combinatorial knockdown of both genes further reduced the lung metastatic burden (Fig. 3b). Both *CYTIP* and *CXCR4* cDNA constructs also enhanced the metastatic potential of 786-O cells (Fig. 3c and Supplementary Fig. 8c). Combinatorial knockdown of CYTIP and CXCR4 strongly inhibited OS-LM1 lung colonization as well (Fig. 3d and Supplementary Fig. 8d).

To determine the cellular mechanisms through which CXCR4 and CYTIP contributed to metastasis, we tested their effects on phenotypes relevant to metastasis *in vitro*. As reported[6], CXCR4 mediated CXCL12 chemotactic migration in metastatic 786-M1A cells, with associated changes in the Erk phosphorylation status (Fig. 3e and Supplementary Fig. 9a,b). The CXCR4 inhibitor AMD3100 blocked the chemotactic phenotype *in vitro* (Fig. 3e) and reduced metastasis formation *in vivo* (Fig. 3f).

Metastasis assays suggested a role for CYTIP in the early steps of lung colonization (Supplementary Fig. 9c). However, CYTIP knockdown also inhibited bone metastasis (Fig. 3g). As extravasation through sinusoid capillaries in the bone marrow is not thought to be rate limiting for bone metastasis, we reasoned that CYTIP might support the survival of disseminating cancer cells. Indeed, CYTIP knockdown significantly reduced cancer cell viability in the presence of TRAIL cell death cytokine (Fig. 3h).

DNA demethylation allows *CYTIP* expression in metastatic ccRCC

We used the two functionally validated target genes, *CXCR4* and *CYTIP,* to investigate the basis of the altered VHL-HIF-response. We found little evidence for genomic amplification, altered mRNA stability, or increased promoter activity as being the explanation for the increased expression of these genes (Supplementary Fig. 10a–c). We therefore asked whether epigenetic changes in transcriptional accessibility of these loci[36] could explain our observations. Indeed, the promoter regions of *CXCR4* and *CYTIP*, but not that of *VEGFA*, were more sensitive to DNase I in the metastatic cells (Supplementary Fig. 11a,b). This difference remained after enforced expression of VHL (Supplementary Fig. 11b), showing that DNase sensitivity was not simply associated with transcriptional activity.

Mutations in epigenetic regulators make these genes attractive candidates as mediators of epigenetic remodeling in ccRCC[30,37]. In our analysis, RMS50 positivity was not associated exclusively with *PBRM1*, *SETD2* or *JARID1C* mutant tumors (Supplementary Fig. 12), and these genes showed no mutational differences between our parental and metastatic cells. Also, inhibition of *PBRM1*, *JARID1C* or *SETD2* did not result in consistent changes in *CXCR4* expression (Supplementary Fig. 13).

Genome-wide assessment of DNA methylation patterns revealed that, in general, genes that were overexpressed in the metastatic cells had higher levels of DNA methylation near the transcription start site (TSS) when compared to non-changed genes or random genomic loci (Supplementary Fig. 14a). This methylation was reduced in the metastatic cells to the level of non-changed genes. The downregulated genes showed methylation levels comparable to non-changed genes both in the parental and metastatic cells (Supplementary Fig. 14b). Among genomic regions with the greatest DNA methylation difference between the

metastatic and non-metastatic cells, 756 out of 772 regions (98%, $P < 1.0 \times 10^{-10}$) had lower methylation in the metastatic cells (Fig. 4a). Of the 756 regions with reduced DNA methylation, 18 were associated with upregulated genes in the 155 gene set ($P = 5.6 \times 10^{-7}$). The vast majority was not, however, associated with transcriptional changes. *CYTIP* showed reduced methylation near the TSS, whereas *CXCR4* showed low levels of DNA methylation across the whole gene locus and no change (Fig. 4b and Supplementary Fig. 15a). We confirmed these results with targeted analysis of the most highly induced RMS50 genes (Supplementary Fig. 15b).

To evaluate the role of DNA methylation in *CYTIP* expression, we treated 786-O cells with the DNA methyltransferase inhibitor 5′-aza-deoxycytidine (5DC). This led to DNA demethylation (Supplementary Fig. 15c) and robust expression of *CYTIP*, but not *CXCR4*, *ADM* or *VEGFA* (Fig. 4c). *CYTIP* induction was dependent on VHL inactivation (Fig. 4c), and it remained high two weeks after 5DC treatment (Supplementary Fig. 15d). Concordantly, analysis of DNA methylation in 78 untreated human primary ccRCC specimens detected little methylation at the *CXCR4* promoter (Fig. 4d), whereas *CYTIP* promoter contained abundant DNA methylation (Fig. 4e). We observed a significant inverse correlation between tumor grade and stage with *CYTIP* methylation (Fig. 4e and Supplementary Fig. 16). Low levels of *CYTIP* methylation correlated with poor disease outcome (Fig. 4f). Thus, changes in *CYTIP* DNA methylation were associated with aggressive tumor phenotypes both in our experimental model and clinically.

Loss of H3K27me3 repressive mark at the *CXCR4* locus

The involvement of DNA methylation in *CYTIP* but not in *CXCR4* regulation implied that multiple mechanisms of VHL-HIF output amplification are coselected during ccRCC progression. The *CXCR4* locus in stem cells is marked by a bivalent histone H3 modification with trimethylated lysines 4 and 27 (H3K4me3 and H3K27me3, respectively)[38]. H3K4me3 is generally a mark of transcriptionally active genes, and H3K27me3 of repressed genes[39]. ChIP-seq experiments revealed that, in general, 786-M1A cells contained more H3K4me3 near upregulated genes in the 155 gene set when compared to the parental cells, whereas the opposite was true for the downregulated genes (Supplementary Fig. 17a). H3K27me3 followed an inverse pattern, where genes with higher expression in the metastatic cells showed lower levels of H3K27me3 and *vice versa* (Supplementary Fig. 17b), relative to the control experiment with IgG (Supplementary Fig. 17c). In individual genes, H3K4me3 followed in most cases the pattern expected of transcriptionally active genes, as shown for *CXCR4* and *CYTIP* (Fig. 5a). For H3K27me3, on the other hand, only few genes, including *CXCR4*, drove the observed global difference between the cell lines (Fig. 5b). The majority of genes, including *CYTIP* (Fig. 5b), contained H3K27me3 levels comparable to the negative IgG control (Supplementary Fig. 17d).

Next we identified genomic loci that showed the greatest difference in H3K4me3 and H3K27me3 enrichment between the parental and metastatic cells, irrespective of their expression pattern. For H3K4me3 this revealed a uniform distribution of both higher and lower enrichment in the metastatic cells (Supplementary Fig. 18a). In contrast, for H3K27me3 the distribution was strongly skewed, with the most differentially enriched

regions having a lower level of H3K27me3 in the metastatic cells (Supplementary Fig. 18b). Among the 155 metastasis-associated genes, we observed a strong correlation between changes in H3K4me3 and gene expression, especially for the genes that were activated in the metastatic cells (Supplementary Fig 18c,d). For H3K27me3, however, no significant overlap was seen (Supplementary Fig. 18c), the only shared gene between the 74 most differentially changed regions and the genes upregulated in metastatic cells being *CXCR4*. Hence, the majority of H3K27me3 demethylation events do not automatically lead to transcriptional activation.

We confirmed these ChIP-seq results by ChIP-qPCR and tested the effects of VHL-reintroduction on the histone modification status of *CXCR4* and *CYTIP*. Whereas the high H3K4me3 levels were strongly dependent on VHL-HIF pathway activity (Fig. 5c), H3K27me3 remained unchanged (Fig. 5d).

Loss of PRC2-mediated repression of *CXCR4* in metastatic ccRCC

H3K27me3 is a hallmark of gene repression by Polycomb Repressive Complex 2 (PRC2)[40]. Among the genes encoding the core PRC2 subunits[40] (EED, EZH1, EZH2, SUZ12, RBBP4 and RBBP7)*, SUZ12* mRNA levels were downregulated in the metastatic cells ($P = 0.009$), although the fold change in the 786-O system did not qualify for the 155 gene set. The metastatic cells expressed lower levels of SUZ12 protein, whereas both the parental and metastatic cells expressed similar levels of EZH2 (Fig. 6a). ChIP-qPCR experiments reveled that SUZ12 binding was reduced at the *CXCR4* promoter of 786-M1A cells (Fig 6b). Low levels of *SUZ12* expression correlated also with ccRCC progression (Fig. 6c). The RMS50-positive tumors in the TCGA cohort had both higher expression of *CXCR4* ($P = 5.2 \times 10^{-7}$) and lower expression of *SUZ12* ($P = 8.3 \times 10^{-9}$).

Knockdown of SUZ12 in the 786-O cells led to a reduction in H3K27me3 at the *CXCR4* promoter (Fig. 6d and Supplementary Fig. 19a) and a proportional increase in *CXCR4* expression (Fig. 6e). The SUZ12 knockdown cells displayed an enhanced metastatic phenotype *in vivo* (Fig. 6f). *CXCR4* expression was also increased with a reduction in H3K27me3 when EZH1 and EZH2 were knocked down together but not individually (Fig. 6g and Supplementary Fig. 19b,d), in line with previous reports showing functional redundancy of EZH1 and EZH2[41–43].

We then analyzed the transcriptome-wide effects of 5DC and SUZ12 knockdown, respectively, in the 786-O system (Supplementary Fig. 20a,b, Supplementary Tables 3 and 4). This revealed a significant overlap between the 155 gene set and genes regulated by DNA methylation or by SUZ12. The majority of gene expression changes caused by 5DC or SUZ12 knockdown was not part of the 155 gene set, however, indicating the involvement of additional factors imposing selectivity.

DISCUSSION

Here we show that repressive chromatin modifications and DNA methylation restrict the expression of metastasis-associated VHL-HIF target genes, and that liberation from these constrains mediates ccRCC metastasis. The transcriptional output of the VHL-HIF pathway

thus evolves during ccRCC progression to further enhance the metastatic potential of the pathway (Figure 6h). These findings provide insights into the interplay between genetic and epigenetic mechanisms in the selection of metastatic traits.

The role of the VHL-HIF pathway in ccRCC initiation is well-established[5], and the VHL-HIF target *CXCR4* has been proposed to drive ccRCC metastasis as well[6]. Clinical data suggests, however, that VHL-HIF activation does not automatically lead to the activation of metastasis genes[21]. Our data shows that a metastatic subprogram of the VHL-HIF pathway is activated only in a subpopulation of cancer cells in renal tumors. We find that CXCR4 and CYTIP are important mediators of the metastatic phenotype driven by the altered VHL-HIF response. CXCR4 has been previously implicated in the metastatic progression of other cancers[1,17,18], and CXCL12, the ligand for CXCR4 and CXCR7, can attract ccRCC cells as well[6], results confirmed in our model system. We identified additional genes that follow similar expression patterns, one of which is the intracellular signal modulator CYTIP. Our work establishes a novel role for *CYTIP* in supporting survival under stressed conditions, and points at additional HIF target genes for future studies.

The pathways that drive metastasis are often considered separate from tumor-initiating functions[4]. In the case of ccRCC, the major tumor-initiating pathway is also exploited for the acquisition of metastatic traits. The central role of the VHL-HIF pathway as a gatekeeper for renal tumorigenesis makes ccRCC an ideal model system for studying the connection between metastatic phenotypes and tumor-initiating mechanism. It is likely that the phenotypes of other tumor-initiating pathways may also expand to cover metastatic functions. The general concept emerging from this work could therefore have implications for our understanding of the evolution of metastatic traits in other cancers as well.

ONLINE METHODS

Cell lines

We obtained 786-O from ATCC, RFX-631 cells from National Cancer Institute, and OS-RC-2 cells from RIKEN Cell Bank (Japan), and cultured in RPMI1640. For retrovirus and lentivirus production, GPG29 and 293T cells, respectively, were cultured in DMEM. RPMI1640 and DMEM contained 10% FBS, L-Glutamine (2 mM), penicillin (100 IU ml^{-1}), streptomycin (100 μg ml^{-1}) and amphotericin B (1 μg ml^{-1}). In addition, the GPG29 medium contained G418 (0.3 mg ml^{-1}), doxycycline (20 ng ml^{-1}) and puromycin (2 μg ml^{-1}).

Overexpression and RNAi

For *CXCR4* and *CYTIP* cDNA overexpression, we used the pBABE-puro retroviral expression vector. The pBABE-HA-VHL plasmid was generated by the Kaelin lab[44] and obtained from Addgene (plasmid 19234). We used empty pBABE vector as control. Virus production in the GPG29 packaging cells followed published protocols[45]. The virus-containing supernatant was then concentrated by ultracentrifugation, resuspended in RPMI medium and used for infection in the presence of polybrene (8 μg ml^{-1}). Puromycin selection started the following day (4 μg ml^{-1}). Similarly, we used a triple modality

retroviral reporter plasmid to stably label cells for bioluminescence imaging[46], followed by selection of GFP-positive by FACS.

For RNAi-mediated gene silencing, we obtained clones and corresponding empty vectors from Open Biosystems. For EZH1, we used DNA oligos for the cloning of miR30-based shRNA constructs into the pGIPZ and pHAGE-puro vectors[47]. The clone numbers and oligo sequences are listed in Supplementary Table 5.

For HIF2α knockdown, we produced retrovirus as described above. For lentivirus production, we infected 293T cells with the shRNA-containing plasmid together with the packaging plasmids psPAX2 and pMD2.G using Lipofectamine 2000 (Invitrogen). We harvested virus 48 hours after transfection, then filtered and used it to infect target cells in the presence of polybrene (8 μg ml^{-1}). Puromycin selection started the following day (4 μg ml^{-1}). We used the corresponding empty vector as a control (pLKO.1 for *CYTIP*, pGIPZ for *CXCR4, LTBP1* and *SUZ12*). For combination knockdown, we added two different viral supernatants into the same transduction mix. Correspondingly, the control for combination knockdown experiments contained both pLKO.1 and pGIPZ empty vectors.

Animal studies and isolation of metastatic cells

We performed all animal experiments in accordance with a protocol approved by MSKCC Institutional Animal Care and Use Committee. For *in vivo* selection, we inoculated cells (400,000) resuspended in 100μl of PBS into the lateral tail vein of 5–7 week old male NOD/SCID mice. We followed tumor growth by bioluminescence imaging of anesthetized mice (100 mg kg^{-1} ketamine and 10 mg kg^{-1} xylazine) by injecting retro-orbitally d-luciferin (150 mg kg^{-1}) and imaging with the IVIS Spectrum Xenogen machine (Caliper Life Sciences). When growing signal was observed for several weeks, we harvested lungs, dissected tumor nodules from the lungs, minced and placed them in RPMI supplemented with 0.125% collagenase III and 0.1% hyaluronidase. Samples were then gently rocked in 37°C for 3–4 hours, briefly centrifuged and resuspended in 0.25% trypsin for ten minutes with vortexing every 2 minutes. We then centrifuged cells and plated them in RPMI medium. After reaching confluency, GFP-positive cells were sorted by FACS, and reinjected into mice. For the comparison of parental and metastatic cells in the lung colonization assays, we injected 150,000 cells (except for 786-M1B cells, for which three mice were injected with 150,000 cells and four mice with 400,000 cells). In all subsequent lung colonization assays, we used 150,000 cells. For bone and brain metastasis assays, we injected cells (100,000) into the left vetricle of 5–7 week old anesthetized athymic male nude mice. We monitored metastatic colonization by bioluminescence and X-ray imaging. For orthotopic renal subcapsular injections, we mixed cells with Matrigel. Incisions were made into the left flank of anesthetized mice and the left kidneys were exposed. Ten microliters of Matrigel containing 15,000 cells were carefully injected under the renal capsule and the incision was sutured. We confirmed the injections were confirmed by bioluminescent imaging. We confirmed brain metastases as well as lung metastasis after orthotopic injections by *ex vivo* imaging. For drug experiments, we purchased AMD3100 from Sigma and injected the drug intraperitoneally twice daily (2.5 mg kg^{-1}).

Histological analysis and immunostaining

After being sacrificed, mice were perfused with 4% paraformaldehyde. Target organs were collected and fixed over night in 4% paraformaldehyde, washed, embedded in paraffin and sectioned. Haematoxylin-eosin staining was performed by standard methods. The MSKCC Molecular Cytology Core Facility performed immunostaining for GFP (Abcam) according to standard methods.

In vitro assays

For proliferation assays, we plated cells (500 per well) on 96-well plates and measured growth using the Resazurin reagent (R&D Systems) according to the manufacturer's instructions. We read the signal using a plate reader, each time point in eight independent wells. For cell migration assays, we starved cells starved over night (0.2% FBS containing media). The next day, we plated 50,000 cells on Fluoroblock cell culture inserts (BD Falcon) with 8μm pores placed in control, recombinant human CXCL12 (R&D) or CXCL12 + AMD3100 (Sigma) containing media. Fourteen hours later the cells were fixed and stained with DAPI. Migration to the basolateral side of the membrane (nuclei stained with DAPI) was quantified using ImageJ software (three random fields for replicate wells) from images obtained by AMG Evos FL microscope. We quantified cell survival by the CellTiter-Glo cell viability assay (Promega). We plated cells (100,000) on 6-well plates with no serum, adding recombinant human TRAIL (PeproTech) the next day, and measuring cell survival two days later.

DNA Sequencing

The MSKCC DNA sequencing core facility sequenced *VHL* by capillary sequencing using previously published primers sequences were[48]. *SETD2*, *JARID1C*, *UTX* and *PBRM1* capillary sequencing was performed in the MSKCC Beene Translational Oncology Core Facility.

mRNA and protein detection

We extracted total RNA using PrepEase RNA spin kit (USB). RNA (1 μg) was used to generate cDNA with Transcriptor First Strand cDNA Synthesis Kit (Roche). For quantitative PCR, we used predesigned Taqman gene expression assays (Applied Biosystems) and the ABI 7900HT Fast Real-Time PCR System. *TBP* served as a housekeeping control gene in the analysis peformed by the SDS2.2.2 software (Applied Biosystems). For immunoblotting, we washed cells with PBS and lysed them using RIPA buffer. Proteins were separated by SDS-PAGE, transferred to PVDF membranes and blotted with antibodies recognizing HIF2α (NB100-122, Novus Biologicals), EZH2 (Cat. 612666, BD Biosciences), SUZ12 (ab12073, Abcam), Erk (9102S, Cell Signaling), pERK (4370S, Cell Signaling), CYTIP (SAB4200043-25UL, Sigma), α-tubulin (11H10, Cell Signaling). Secondary antibodies were HRP (Pierce) or fluorescence (LiCor) conjugated. Immunofluorescence of CXCR4 was performed with the MAB173 antibody (R&D).

Clinical ccRCC samples

We collected a set of 91 surgically removed primary ccRCC specimens at our institution according to a protocol approved by the MSKCC Institutional Review Board and Human Biospecimen Utilization Committee. Samples were snap-frozen and DNA and RNA were extracted by standard methods.

mRNA expression analysis of clinical ccRCC samples

Direct assessment of *CXCR4* mRNA expression was performed by the nCounter Analysis System (NanoString Technologies) according to manufacturer's recommendations. The method is based on a solution-phase hybridization protocol that allows direct detection of target mRNA molecules in small amounts of input RNA. After data preprocessing and background correction, each sample was normalized for RNA loading using the ribosomal genes *RPL13A*, *RPL24*, *RPL27*, and *RPS20*. All 91 RNA samples passed the initial data quality control criteria.

DNA methylation analysis

The effects of DNA methylation on gene expression was studied by treating cells with 5′-aza-deoxycytidine (5DC) followed by quantitative RT-PCR. A concentration of 100 nM was used in all experiments. MSKCC Beene Translational Oncology Core Facility carried out analysis of CpG methylation using the EpiTyper system (Sequenom, San Diego, CA), which performs quantitative analysis of DNA methylation using base-specific cleavage of bisulfite-treated DNA and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. One μg of DNA was subjected to bisulfite treatment using the EZ-96 DNA methylation Kit following the manufacturer's instructions (Zymo Research, Orange, CA). Specific PCR primers with T7-promoter tags were designed using the EpiDesigner software (www.epidesigner.com). After polymerase chain reaction (PCR) and shrimp alkaline phosphatase (SAP)-treatment, in vitro transcription and T cleavage were carried out using the MassCleave kit (Sequenom). Epityper reaction product was then loaded onto a SpectroCHIP II array (Sequenom) and analyzed using a Bruker Biflex III MALDI-TOF mass spectrometer (SpectroREADER, Sequenom). Results were analyzed using the Epityper Analyzer software, and manually inspected for spectra quality and peak quantification. Of the clinical samples, 78 passed the quality control criteria and were used for further analysis.

DNase 1 sensitivity assays

We assessed chromatin accessibility as a function of DNase 1 sensitivity, following previously published protocols with modifications[49,50]. Cells were trypsinized and counted, washed once with PBS and resuspended in 500μl of ice-cold RSB buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$). Cells were lysed by slowly adding 8 ml of cold lysis buffer (RSB supplemented with 0.1% IGEPAL CA-630) and incubating on ice for ten minutes. Nuclei were pelleted by centrifuging five minutes at 500 g, washed once with 5 ml of cold RSB, resuspended in DNase 1 reaction buffer (Roche) at a $\sim 7 \times 10^6$ ml^{-1} concentration and aliquoted into multiple tubes. Treatment with recombinant DNase 1 was carried out at 37°C for 20 minutes and the reaction was stopped by adding an equal volume of stop buffer (0.1 M NaCl, 0.1% (w vol^{-1}) SDS, 50 mM Tris-HCl (pH 8.0), 100 nM

EDTA). Samples were then treated with proteinase K over night and DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). The DNase 1 digestion was first optimized by testing several different concentrations (10–200U per 180 µl reaction). For final experiments, we performed 3–4 DNase treatments with the same concentration. We quantified DNA by PCR using the SYBR Green Mix (Thermo Scientific) and ABI 7900HT Fast Real-Time PCR System absolute quantification method. DNase insensitive control loci (Chr15at22Mb and Chr17at17Mb) that were randomly chosen based on DNase data from multiple cell lines available at the UCSC genome browser (http://genome.ucsc.edu/) and empirically validated as regions showing minimal sensitivity to DNase when compared to non-treated controls served as normalization controls. PCR primers listed in Supplementary Table 6.

Chromatin Immunoprecipitation (ChIP)

For ChIP, we used the ChIP Assay Kit (Millipore) according to the manufacturer's instructions. Cells were sonicated in 20 second pulses for a total of 3 min 20 s. The following antibodies were used: H3K4me3 (04-745, Millipore), H3K27me3 (17-622, Millipore), SUZ12 (ab12073, Abcam). Control experiments were done without antibody. DNA was extracted using the QIAquick PCR Purification Kit (Qiagen) followed by quantitative PCR essentially as described above for DNase sensitivity assays.

Gene expression profiling

We extracted total RNA using the PrepEase RNA spin kit (USB). MSKCC Genomics Core Facility prepared the samples according to standard protocols and hybridized them on Affymetrix HG-U133A or HG-U133 Plus 2.0 microarrays. We processed the data using GCRMA together with updated probe set definitions (hs133phsentrezg)[51]. Differentially expressed genes between groups were identified with fold-change (<0.5 or >2) and t-test (P < 0.05) cutoffs. From the initial comparison of the parental 786-O cells to the metastatic derivatives, we identified 155 genes (the 155 gene set).

Bioinformatic analysis

We conducted all bioinformatic analyses using R. The following data sets were used: GSE2109 (192 ccRCC samples), GSE3538 (Ref. 25), GSE17895 (Ref. 30), TCGA[52]. GSE2109 was processed as described above for the cell line data. For GSE3538 and GSE17895, preprocessed data was downloaded from the Stanford Microarray Database (http://smd.stanford.edu, normalized log2 ratio, >70% good data), and from GEO, respectively. For GSE3538, missing values were imputed with the k-nearest neighbors method, using the 10 neighboring genes. For the TCGA data set, normalized mRNA z-scores were downloaded from the TCGA cBio portal[52]. Differentially expressed genes between groups were identified by fold-change and t-test without the equal variance assumption (P < 0.05).

We performed unsupervised hierarchical clustering using gene Z-scores and the function *heatmap.2* in the R-package *gplots*. To find the most suitable method for clustering, we compared three different algorithms (Pearson's correlation coefficient, Spearman's correlation coefficient and Kendall's correlation coefficient) to calculate correlation between

samples using the 155 gene set in the GSE2109 data set. As a measure of cluster reproducibility, we calculated R-indices as described[53]. Pearson's correlation coefficient yielded the highest R-indices for multiple cutoff levels (number of clusters), and was thus selected for all subsequent analysis.

155 gene set-positive tumors were identified by unsupervised hierarchical clustering. Visual inspection revealed a clear cluster in all three data sets: TCGA, GSE2109 and GSE3538. RMS50, a derivative of the 155 gene set, was generated by using the GSE2109 data set and including genes that fulfilled the following criteria: upregulated in the metastatic derivatives and higher expression in the 155 gene set-positive cluster, or downregulated in the metastatic derivatives and lower expression in 155 gene set-positive cluster.

For categorical data (RMS50 status), differences in survival between groups were evaluated using the log-rank test. For continuous variables (*CYTIP* methylation data, *CXCR4* mRNA expression, *SUZ12* mRNA expression), a Cox proportional hazards model was utilized as implemented in the *coxph* function in the R-package *survival*.

ChIP-seq library preparation

Subconfluent cells were cross-linked (1% formaldehyde, 10 min), quenched (200 mM Glycine, 5 min), washed with ice-cold PBS and harvested by scraping. ChIP was performed essentially as described[54]. ChIP library was prepared for sequencing by the SOLiD 4 (H3K4me3) or HiSeq2000 (H3K27me3) instruments according to manufacturer recommended procedures.

STAMP assay library preparation

Genomic DNA was randomly fragmented to a modal length of 120 bp and then methylated DNA was enriched and processed for massively parallel sequencing using the sequence tag analysis of DNA methylation patterns (STAMP) method[55]. MBD-enriched fragment libraries were sequenced using SOLiD 4.

ChIP-seq and STAMP Data Analysis

Sequenced fragments were mapped to Hg18 and genomic coverage maps were generated by extending single-end reads to the modal fragment length or based upon mapping of paired-ends as described[55]. Coverage maps were normalized to the total library size and were then used to evaluate relative enrichment near each Refseq transcription start site ($\pm$ 5,000 bp). The significance of differential coverage was estimated by counting tags within 100 bp windows tiled across the regions of interest and then estimating the significance of differential tag counts using empirical Bayes estimation and exact tests based on the negative binomial distribution[56].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Kang Y, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell. 2003; 3:537–549. [PubMed: 12842083]

2. Nguyen DX, et al. WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. Cell. 2009; 138:51–62. [PubMed: 19576624]

3. Sonoshita M, et al. Suppression of colon cancer metastasis by Aes through inhibition of Notch signaling. Cancer Cell. 2011; 19:125–137. [PubMed: 21251616]

4. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009; 9:274–284. [PubMed: 19308067]

5. Kaelin WG Jr. The von Hippel-Lindau tumour suppressor protein: O2 sensing and cancer. Nat Rev Cancer. 2008; 8:865–873. [PubMed: 18923434]

6. Staller P, et al. Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL. Nature. 2003; 425:307–311. [PubMed: 13679920]

7. Kaelin WG. Von Hippel-Lindau disease. Annu Rev Pathol. 2007; 2:145–173. [PubMed: 18039096]

8. Chauveau D, et al. Renal involvement in von Hippel-Lindau disease. Kidney Int. 1996; 50:944–951. [PubMed: 8872970]

9. Mandriota SJ, et al. HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. Cancer Cell. 2002; 1:459–468. [PubMed: 12124175]

10. Gnarra JR, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nat Genet. 1994; 7:85–90. [PubMed: 7915601]

11. Herman JG, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A. 1994; 91:9700–9704. [PubMed: 7937876]

12. Montani M, et al. VHL-gene deletion in single renal tubular epithelial cells and renal tubular cysts: further evidence for a cyst-dependent progression pathway of clear cell renal carcinoma in von Hippel-Lindau disease. Am J Surg Pathol. 2010; 34:806–815. [PubMed: 20431476]

13. Kapitsinou PP, Haase VH. The VHL tumor suppressor and HIF: insights from genetic studies in mice. Cell Death Differ. 2008; 15:650–659. [PubMed: 18219317]

14. Kondo K, Klco J, Nakamura E, Lechpammer M, Kaelin WG Jr. Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. Cancer Cell. 2002; 1:237–246. [PubMed: 12086860]

15. Kondo K, Kim WY, Lechpammer M, Kaelin WG Jr. Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth. PLoS Biol. 2003; 1:E83. [PubMed: 14691554]

16. Shen C, et al. Genetic and functional studies implicate HIF1alpha as a 14q kidney cancer suppressor gene. Cancer Discov. 2011; 1:222–235. [PubMed: 22037472]

17. Muller A, et al. Involvement of chemokine receptors in breast cancer metastasis. Nature. 2001; 410:50–56. [PubMed: 11242036]

18. Zhang XH, et al. Latent bone metastasis in breast cancer tied to Src-dependent survival signals. Cancer Cell. 2009; 16:67–78. [PubMed: 19573813]

19. Schlesinger-Raab A, Treiber U, Zaak D, Holzel D, Engel J. Metastatic renal cell carcinoma: results of a population-based study with 25 years follow-up. Eur J Cancer. 2008; 44:2485–2495. [PubMed: 18783939]

20. Yao M, et al. VHL tumor suppressor gene alterations associated with good prognosis in sporadic clear-cell renal carcinoma. J Natl Cancer Inst. 2002; 94:1569–1575. [PubMed: 12381710]

21. Young AC, et al. Analysis of VHL gene alterations and their relationship to clinical parameters in sporadic conventional renal cell carcinoma. Clin Cancer Res. 2009; 15:7582–7592. [PubMed: 19996202]

22. D'Alterio C, et al. Differential role of CD133 and CXCR4 in renal cell carcinoma. Cell Cycle. 2010; 9:4492–4500. [PubMed: 21127401]

23. Williams RD, Elliott AY, Stein N, Fraley EE. In vitro cultivation of human renal cell cancer. II Characterization of cell lines. In Vitro. 1978; 14:779–786. [PubMed: 721102]

24. Hess KR, et al. Metastatic patterns in adenocarcinoma. Cancer. 2006; 106:1624–1633. [PubMed: 16518827]

25. Zhao H, et al. Gene expression profiling predicts survival in conventional renal cell carcinoma. PLoS Med. 2006; 3:e13. [PubMed: 16318415]

26. Wang L, et al. Strong expression of chemokine receptor CXCR4 by renal cell carcinoma cells correlates with metastasis. Clin Exp Metastasis. 2009; 26:1049–1054. [PubMed: 19859817]

27. Zagzag D, et al. Stromal cell-derived factor-1alpha and CXCR4 expression in hemangioblastoma and clear cell-renal cell carcinoma: von Hippel-Lindau loss-of-function induces expression of a ligand and its receptor. Cancer Res. 2005; 65:6178–6188. [PubMed: 16024619]

28. Semenza GL. Targeting HIF-1 for cancer therapy. Nat Rev Cancer. 2003; 3:721–732. [PubMed: 13130303]

29. Mole DR, et al. Genome-wide association of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha DNA binding with expression profiling of hypoxia-inducible transcripts. J Biol Chem. 2009; 284:16767–16775. [PubMed: 19386601]

30. Dalgliesh GL, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010; 463:360–363. [PubMed: 20054297]

31. Chen Q, Coffey A, Bourgoin SG, Gadina M. Cytohesin binder and regulator augments T cell receptor-induced nuclear factor of activated T Cells. AP-1 activation through regulation of the JNK pathway. J Biol Chem. 2006; 281:19985–19994. [PubMed: 16702224]

32. Watford WT, et al. Cytohesin binder and regulator (cybr) is not essential for T- and dendritic-cell activation and differentiation. Mol Cell Biol. 2006; 26:6623–6632. [PubMed: 16914744]

33. Boehm T, et al. Attenuation of cell adhesion in lymphocytes is regulated by CYTIP, a protein which mediates signal complex sequestration. EMBO J. 2003; 22:1014–1024. [PubMed: 12606567]

34. Hofer S, et al. Dendritic cells regulate T-cell deattachment through the integrin-interacting protein CYTIP. Blood. 2006; 107:1003–1009. [PubMed: 16204317]

35. Saharinen J, Hyytiainen M, Taipale J, Keski-Oja J. Latent transforming growth factor-beta binding proteins (LTBPs)--structural extracellular matrix proteins for targeting TGF-beta action. Cytokine Growth Factor Rev. 1999; 10:99–117. [PubMed: 10743502]

36. Bell O, Tiwari VK, Thoma NH, Schubeler D. Determinants and dynamics of genome accessibility. Nat Rev Genet. 2011; 12:554–564. [PubMed: 21747402]

37. Varela I, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011; 469:539–542. [PubMed: 21248752]

38. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]

39. Kouzarides T. Chromatin modifications and their function. Cell. 2007; 128:693–705. [PubMed: 17320507]

40. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature. 2011; 469:343–349. [PubMed: 21248841]

41. Margueron R, et al. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell. 2008; 32:503–518. [PubMed: 19026781]

42. Shen X, et al. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell. 2008; 32:491–502. [PubMed: 19026780]

43. Shi J, et al. The Polycomb complex PRC2 supports aberrant self-renewal in a mouse model of MLL-AF9;Nras(G12D) acute myeloid leukemia. Oncogene. 201210.1038/onc.2012.110

44. Li L, et al. Hypoxia-inducible factor linked to differential kidney cancer risk seen with type 2A and type 2B VHL mutations. Mol Cell Biol. 2007; 27:5381–5392. [PubMed: 17526729]

45. Minn AJ, et al. Genes that mediate breast cancer metastasis to lung. Nature. 2005; 436:518–524. [PubMed: 16049480]

46. Ponomarev V, et al. A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. Eur J Nucl Med Mol Imaging. 2004; 31:740–751. [PubMed: 15014901]

47. Dow LE, et al. A pipeline for the generation of shRNA transgenic mice. Nat Protoc. 2012; 7:374–393. [PubMed: 22301776]

48. Banks RE, et al. Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. Cancer Res. 2006; 66:2000–2011. [PubMed: 16488999]

49. McArthur M, Gerum S, Stamatoyannopoulos G. Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. J Mol Biol. 2001; 313:27–34. [PubMed: 11601844]

50. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010; 2010:pdb prot5384. [PubMed: 20150147]

51. Dai M, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. 2005; 33:e175. [PubMed: 16284200]

52. Cerami E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2:401–404. [PubMed: 22588877]

53. McShane LM, et al. Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data. Bioinformatics. 2002; 18:1462–1469. [PubMed: 12424117]

54. Kashyap V, et al. Epigenomic reorganization of the clustered Hox genes in embryonic stem cells induced by retinoic acid. J Biol Chem. 2011; 286:3250–3260. [PubMed: 21087926]

55. Brenet F, et al. DNA methylation of the first exon is tightly linked to transcriptional silencing. PLoS One. 2011; 6:e14524. [PubMed: 21267076]

56. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]
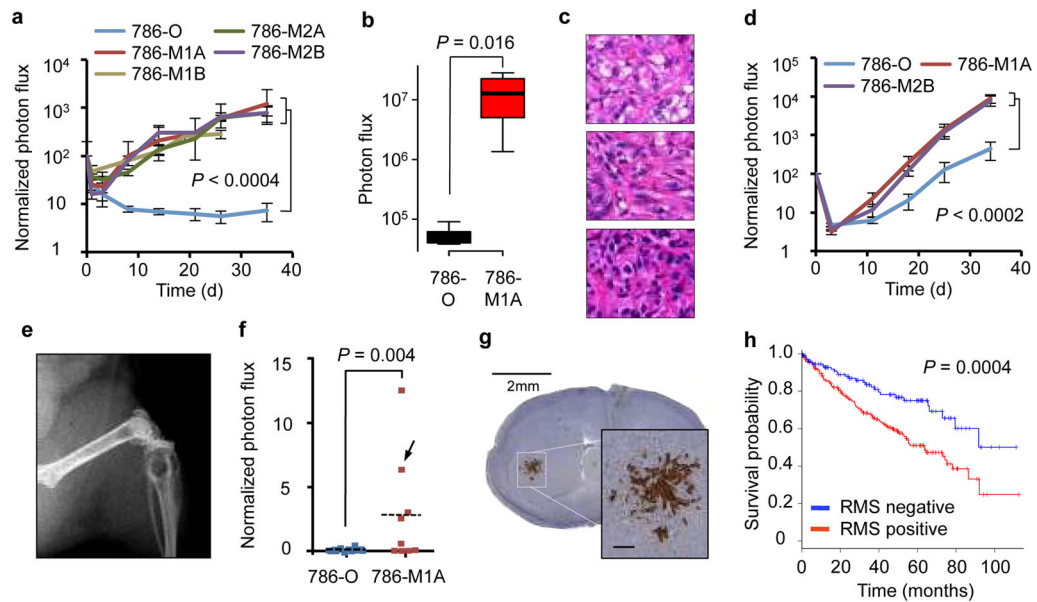
Figure 1. Experimental Model System and Gene Expression Signature for ccRCC Metastasis

a. Quantification of lung photon flux (photons per second) normalized to day 0 after tail vein inoculation of cancer cells. $n = 6$–9 for all groups. P-value calculated by the two-tailed Wilcoxon rank sum test. Error bars represent standard error of the mean.

b. Lung photon flux *ex vivo* 66 days after renal subcapsular inoculation of cells. $n = 5$ for 786-O cells, $n = 4$ for 786-M1A cells. P-value calculated by the two-tailed Wilcoxon rank sum test.

c. H&E staining of histological sections. Orthotopic tumor formed by 786-O cells (top). Orthotopic tumor formed by 786-M1A cells (middle). Lung metastasis formed by 786-M1A cells (bottom).

d. Normalized photon flux from hind limbs after intracardiac inoculation of cells. $n = 9$–10 for all groups. P-value calculated by the two-tailed Wilcoxon rank sum test. Error bars represent standard error of the mean.

e. X-ray image of a bone metastasis formed by the 786-M1A cells.

f. Normalized photon flux from whole brains imaged *ex vivo* at day 64 after intracardiac inoculation of cells. The arrow indicates the sample from which the histological section shown in (g) was derived.

g. Whole brain section and a close-up showing immunohistochemical staining of a brain metastasis using an anti-GFP antibody. Scale bar in the insert, 0.2 mm.

h. Disease-specific survival for RMS50-positive and RMS50-negative tumors in the TCGA data set. $n = 389$. P-value calculated by the log-rank test.
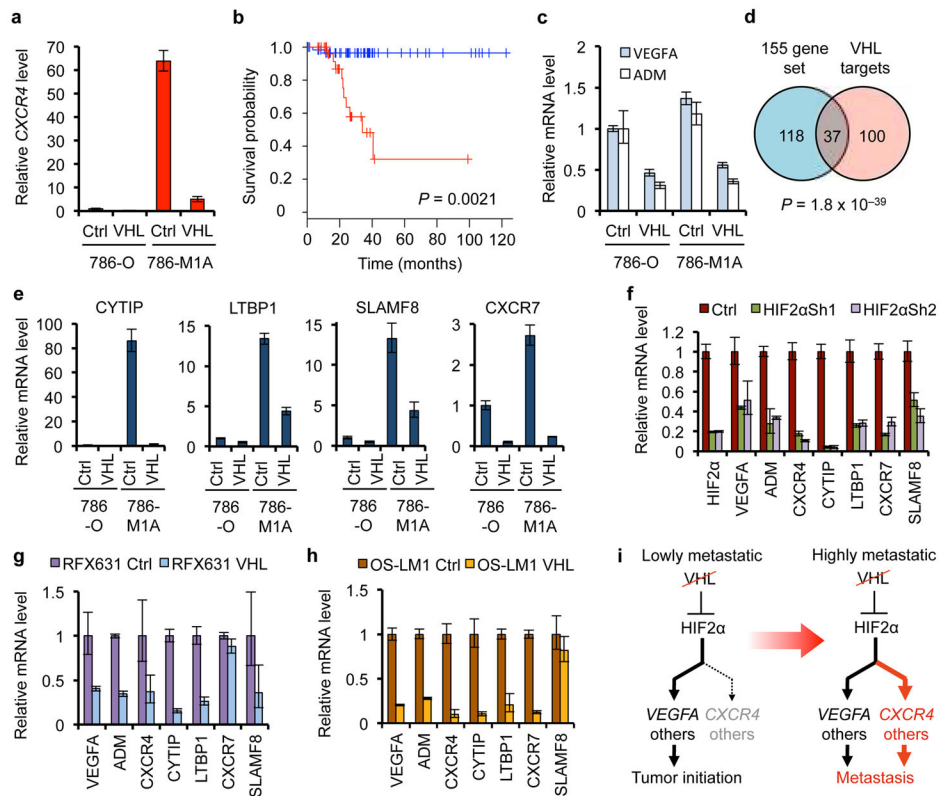
Figure 2. VHL-HIF Signal Output Modulation Associated with ccRCC Progression

a. Effects of VHL reintroduction on *CXCR4* expression measured by quantitative RT-PCR analysis. Expression normalized to 786-O cells transduced with the control vector.

b. Kaplan-Meier analysis of metastasis-free survival in patients with ccRCC. Classification based on *CXCR4* mRNA expression in primary ccRCCs, top 1/3 shown in red, bottom 2/3 in blue. P-value from a Cox proportional hazards model with *CXCR4* expression treated as a continuous variable. $n = 91$.

c. Effects of VHL reintroduction on *VEGFA* and *ADM* expression measured by quantitative RT-PCR analysis. Expression normalized to 786-O cells transduced with the control vector.

d. Venn-diagram showing the overlap between the 155 metastasis-associated genes and genes changed after *VHL* reintroduction into the 786-M1A cells. Statistical significance of overlap tested by Fisher's exact test.

e. Effects of VHL reintroduction on gene expression measured by quantitative RT-PCR analysis. Expression normalized to 786-O cells transduced with the control vector.

f. Effects of HIF2α knockdown on gene expression measured by quantitative RT-PCR analysis in 786-M1A cells. Expression normalized to the cells transduced with the hairpin control vector.

g. Effects of VHL reintroduction on gene expression measured by quantitative RT-PCR analysis. Expression normalized to RFX-631 cells transduced with the control vector.

h. Effects of VHL reintroduction on gene expression measured by quantitative RT-PCR analysis. Expression normalized to OS-LM1 cells transduced with the control vector.

i. A schematic summarizing the hypothesis that when *VHL* loss stabilizes HIF, early ccRCC genes such as *VEGF* and *ADM* are highly expressed whereas the activation of pro-metastatic
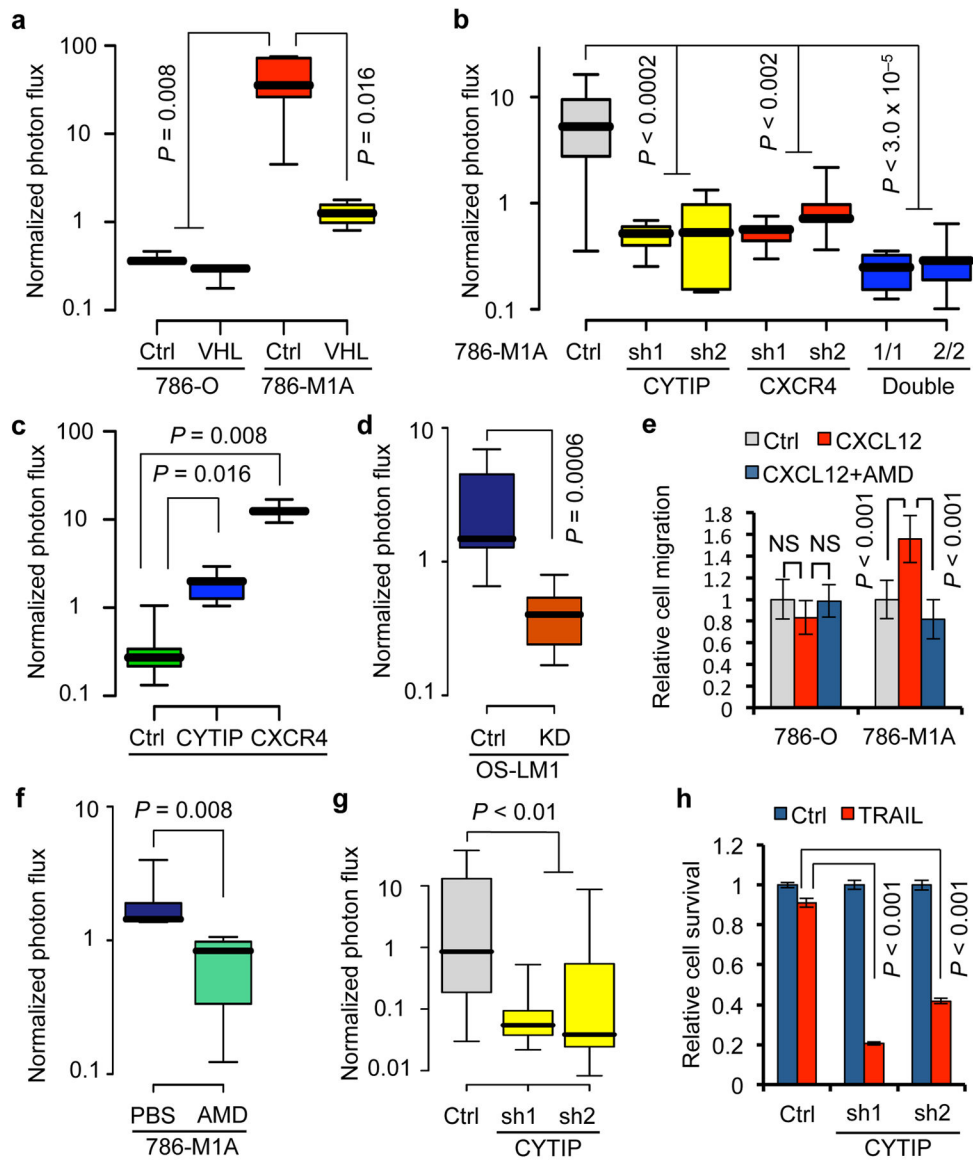
HIF target genes such as *CXCR4* and others requires additional amplifying mechanisms. Error bars represent the 95% confidence interval based on multiple PCR reactions in all panels.

Figure 3. VHL-HIF Pathway Modulation Activates Functional Mediators of ccRCC Metastasis

a. Normalized bioluminescence photon flux of the lungs with and without reintroduced VHL at day 28 after tail vein inoculation of cancer cells. For 786-M1A-VHL cells $n = 4$, for all other groups $n = 5$.

b. Normalized bioluminescence photon flux of the lungs at day 28 after tail vein inoculation of 786-M1A cells with *CYTIP* and *CXCR4* knockdown individually or in combination. Control ($n = 25$) is a pool of the following groups: pLKO.1 $n = 5$, pGIPZ $n = 10$, pLKO.1/ pGIPZ $n = 10$. For CYTIPsh1/CXCR4sh1 $n = 10$, for all other groups $n = 5$.

c. Normalized bioluminescence photon flux of the lungs at day 64 after tail vein inoculation of 786-O cells with transduced *CYTIP* or *CXCR4* cDNA. $n = 5$ per group.

d. Normalized bioluminescence photon flux of the lungs at day 49 after tail vein inoculation of OS-LM1 cells with CYTIP and CXCR4 double knockdown. Control (pLKO.1/pGIPZ) $n = 8$, CYTIPsh1/CXCR4sh1 $n = 7$.

e. Effects of CXCL12 on cancer cell migration measured by the Boyden chamber assay. AMD, the CXCR4 inhibitor AMD3100. Error bars represent standard deviation of the mean.

f. Normalized bioluminescence photon flux of the lungs at day 28 after tail vein inoculation of 786-M1A cells. Mice treated with AMD3100 or PBS. $n = 5$ for both groups.

g. Normalized photon flux from hind limbs after intracardiac inoculation of 786-M1A cells with CYTIP knockdown. $n = 9$–10 for all groups.

h. Effects of CYTIP inhibition on 786-M1A cell survival under TRAIL treatment in low confluency and low serum conditions. Error bars represent standard deviation of the mean. P-value calculated by the two-tailed Wilcoxon rank sum test (**a–d**, **f**, **g**) or the two-tailed Student's t-test (**e**, **h**).
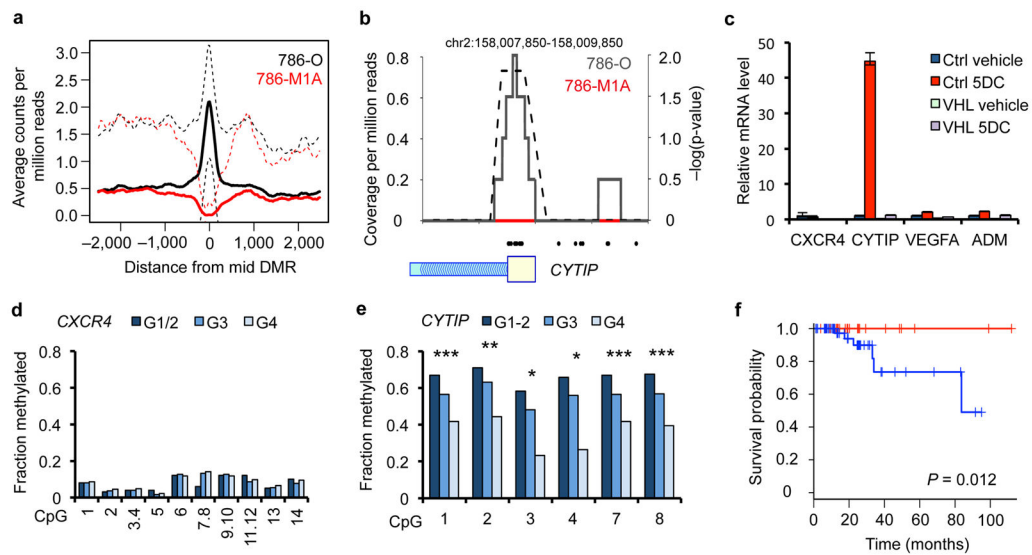
Figure 4. DNA Demethylation Activates *CYTIP* Expression in Metastatic ccRCC

a. Average enrichment of DNA methylation near the 772 genomic loci that showed the greatest difference in DNA methylation between the 786-O and 786-M1A cells plotted for both 786-O cells (black solid line) and 786-M1A cells (red solid line). The black and red dashed lines show the average enrichment plus 1 SD for 786-O and 786-M1A cells, respectively. DMR, differentially methylated region.

b. Enrichment of methylated DNA around (± 2,000 bp) *CYTIP* TSS as determined by STAMP analysis. Y-axis shows the relative counts of high-throughput sequencing reads recovered after enrichment for methylated DNA. The dashed black line indicates an estimate of significance of difference in enrichment between 786-O and 786-M1A cells as determined by a negative binomial error model. The black circles below the x-axis are CpG positions. The first *CYTIP* exon is in tan, intron in light blue with transcription direction indicated by hatching. The metastatic cells show a value of zero throughout the region.

c. Gene expression measured by quantitative RT-PCR analysis of the 786-O cells with or without VHL reintroduction. Cells were treated with either 5DC or vehicle (dimethyl sulfoxide). Error bars represent the 95% confidence interval based on multiple PCR reactions.

d. DNA methylation at *CXCR4* promoter in clinical ccRCC samples ($n = 78$) grouped by tumor grade. The CpG dinucleotides analyzed correspond to the ones highlighted in red in Supplementary Figure 15b.

e. *CYTIP* DNA methylation in clinical ccRCC samples ($n = 78$) grouped by tumor grade. (***) $P < 0.001$, (**) $P < 0.01$. (*) $P < 0.05$, linear regression analysis. The CpG dinucleotides analyzed correspond to the ones highlighted in blue in panel Supplementary Figure 15b.

f. Kaplan-Meier analysis of overall survival in patients with ccRCC. Classification based on average *CYTIP* methylation, top 1/3 shown in red, bottom 2/3 in blue. P-value from a Cox proportional hazards model with average *CYTIP* methylation treated as a continuous variable. $n = 78$.
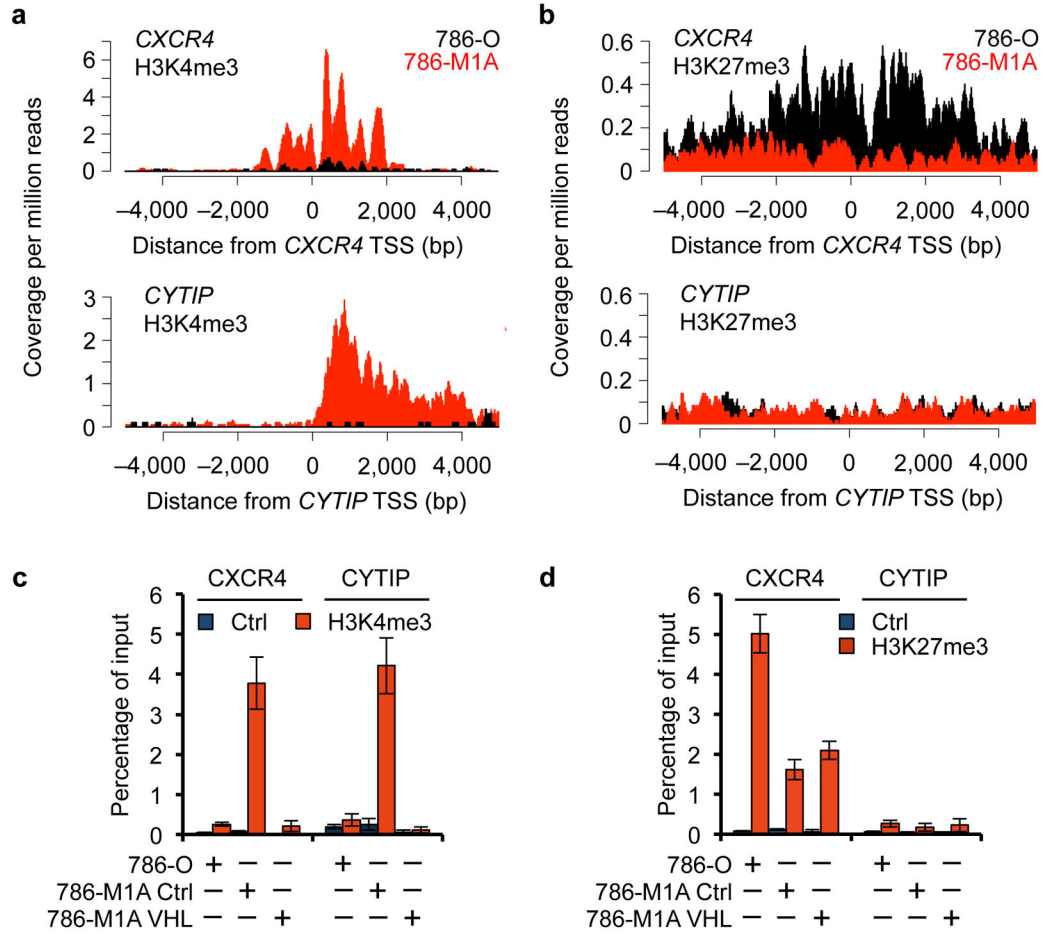
Figure 5. Histone Modification Patterns Linked to ccRCC Progression

a. Normalized H3K4me3 ChIP-seq signal relative to *CXCR4* and *CYTIP* TSS.

b. Normalized H3K27me3 ChIP-seq signal relative to *CXCR4* and *CYTIP* TSS.

c. H3K4me3 chromatin immunoprecipitation followed by quantitative PCR analysis at the *CXCR4* and *CYTIP* loci. Error bars denote standard deviation.

d. H3K27me3 chromatin immunoprecipitation followed by quantitative PCR analysis at the *CXCR4* and *CYTIP* loci. Error bars denote standard deviation.
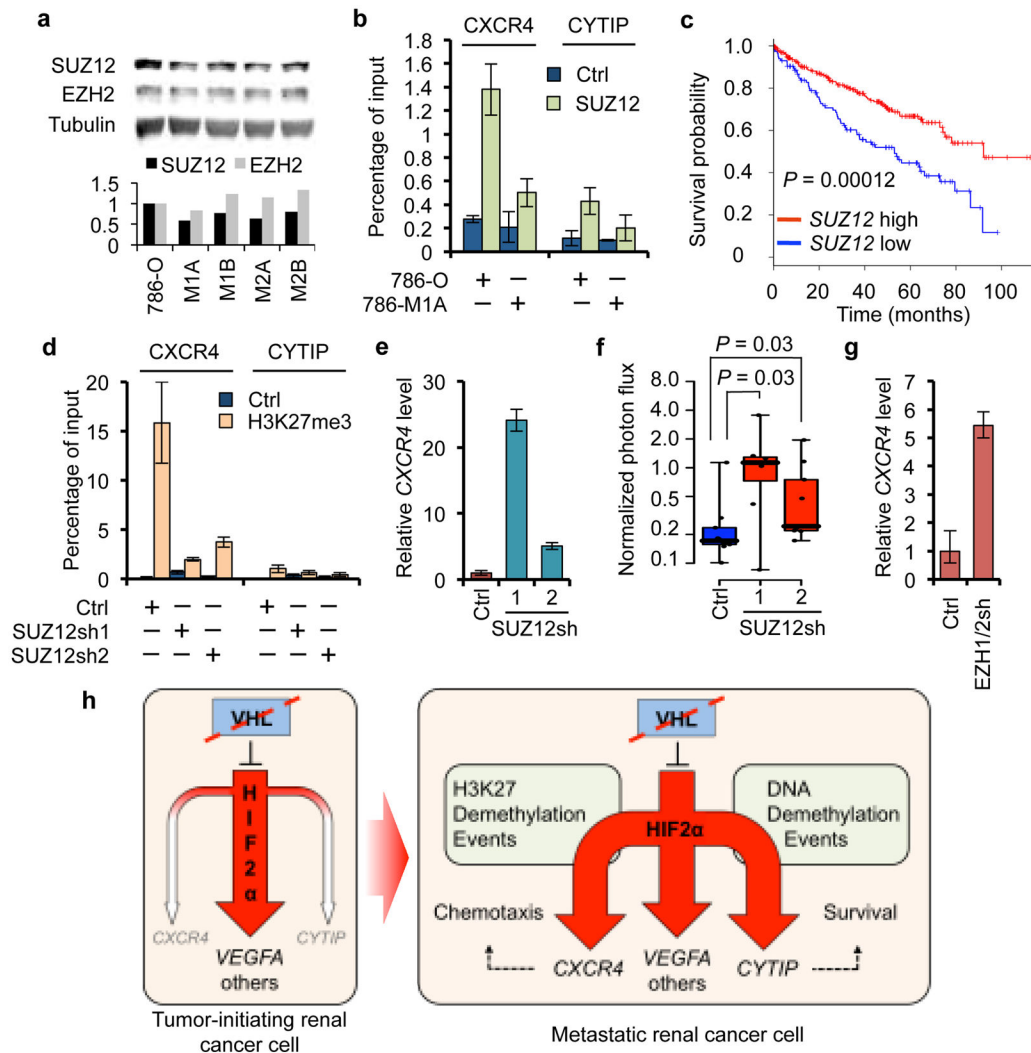
Figure 6. Loss of PRC2-Dependent Repression Activates *CXCR4* Expression in Metastatic ccRCC

a. Top, SUZ12 and EZH2 protein levels measured by immunoblotting. Bottom, quantification of signal normalized to the loading control tubulin.

b. SUZ12 chromatin immunoprecipitation followed by quantitative PCR analysis at the *CXCR4* and *CYTIP* loci. Error bars denote standard deviation.

c. Kaplan-Meier analysis of overall survival in patients with ccRCC from the TCGA data set. Classification based on SUZ12 expression, top 70% shown in red, bottom 30% in blue. P-value from a Cox proportional hazards model with *SUZ12* expression treated as a continuous variable. *n* = 389.

d. H3K27me3 chromatin immunoprecipitation followed by quantitative PCR analysis at the *CXCR4* and *CYTIP* loci. Error bars denote standard deviation.

e. *CXCR4* expression measured by quantitative RT-PCR analysis after *SUZ12* knockdown. Error bars represent 95% confidence interval based on multiple PCR reactions.

f. Lung colonization of 786-O cells after SUZ12 knockdown. Normalized bioluminescence photon flux of the lungs at day 68. n = 7–10. P-value calculated using the two-tailed Wilcoxon rank-sum test.

g. *CXCR4* expression measured by quantitative RT-PCR analysis after combinatorial knockdown of both *EZH1* and *EZH2*. Error bars represent 95% confidence interval based on multiple PCR reactions.

h. Model. *VHL* loss initiates renal tumorigenesis by stabilizing HIF2α, which leads to the activation of several target genes, such as *VEGFA*. At this stage, many potential VHL-HIF target genes are only weakly expressed due to various inhibitory mechanisms (e.g. DNA methylation and PRC2-dependent repression). Changes in the epigenetic landscape can allow hyperinduction of VHL-HIF target genes. If these genes are beneficial to the tumor (e.g. *CXCR4* and *CYTIP*), they will drive ccRCC progression and get selected for.