Chromophobe renal cell carcinoma (chRCC) typically shows ~7 chromosome losses (1, 2, 6, 10, 13, 17, and 21) and ~31 exonic somatic mutations, yet carries ~5%–10% metastatic incidence. Since extensive chromosomal losses can generate proteotoxic stress and compromise cellular proliferation, it is intriguing how chRCC, a tumor with extensive chromosome losses and a low number of somatic mutations, can develop lethal metastases. Genomic features distinguishing metastatic from nonmetastatic chRCC are unknown. An integrated approach, including whole-genome sequencing (WGS), targeted ultradeep cancer gene sequencing, and chromosome analyses (FACETS, OncoScan, and FISH), was performed on 79 chRCC patients including 38 metastatic (M-chRCC) cases. We demonstrate that TP53 mutations (58%), PTEN mutations (24%), and imbalanced chromosome duplication (ICD, duplication of ≥ 3 chromosomes) (25%) were enriched in M-chRCC. Reconstruction of the subclonal composition of paired primary-metastatic chRCC tumors supports the role of TP53, PTEN, and ICD in metastatic evolution. Finally, the presence of these 3 genomic features in primary tumors of both The Cancer Genome Atlas kidney chromophobe (KICH) (n = 64) and M-chRCC (n = 35) cohorts was associated with worse survival. In summary, our study provides genomic insights into the metastatic progression of chRCC and identifies TP53 mutations, PTEN mutations, and ICD as high-risk features.
Chromophobe renal cell carcinoma (chRCC) typically shows ~7 chromosome losses (1, 2, 6, 10, 13, 17, and 21) and ~31 exonic somatic mutations, yet carries ~5%-10% metastatic incidence. Since extensive chromosomal losses can generate proteotoxic stress and compromise cellular proliferation, it is intriguing how chRCC, a tumor with extensive chromosome losses and a low number of somatic mutations, can develop lethal metastases. Genomic features distinguishing metastatic from nonmetastatic chRCC are unknown. An integrated approach, including whole-genome sequencing (WGS), targeted ultradeep cancer gene sequencing, and chromosome analyses (FACETS, OncoScan, and FISH), was performed on 79 chRCC patients including 38 metastatic (M-chRCC) cases. We demonstrate that TP53 mutations (58%), PTEN mutations (24%), and imbalanced chromosome duplication (ICD, duplication of ≥ 3 chromosomes) (25%) were enriched in M-chRCC. Reconstruction of the subclonal composition of paired primary-metastatic chRCC tumors supports the role of TP53, PTEN, and ICD in metastatic evolution. Finally, the presence of these 3 genomic features in primary tumors of both The Cancer Genome Atlas kidney chromophobe (KICH) (n = 64) and M-chRCC (n = 35) cohorts was associated with worse survival. In summary, our study provides genomic insights into the metastatic progression of chRCC and identifies TP53 mutations, PTEN mutations, and ICD as high-risk features.
sequencing (WES), mitochondrial DNA sequencing, DNA copy number array, mRNA sequencing, microRNA sequencing, and DNA methylation array, affirmed an extremely hypodiploid genome with prevalent loss of the chRCC-7set chromosomes, demonstrated the low (~31) exonic somatic mutation rate, and identified TP53 (20%–32%) and PTEN (6%–9%) as the only 2 significantly mutated genes (5, 21). However, the clinical significance of these mutations in chRCC has not been established.

Since extensive chromosomal losses can generate proteotoxic stress and compromise cellular proliferation (22), it is puzzling that chRCC could develop lethal metastases. To investigate the molecular underpinnings of metastatic progression in chRCC, we performed integrated genomic analyses on 49 tumor samples from 38 M-chRCC cases, and conducted ultradeep targeted cancer gene sequencing on tumors from 41 nonmetastatic chRCC cases, and evaluated the impact of identified high-risk genomic features on the overall survival of chRCC patients using 2 independent cohorts.

Results

Clinical information of cases and sample cohorts is outlined. Forty-nine tumor samples derived from 38 M-chRCC patients, including 37 primary and 12 metastatic tumors were denoted M-chRCC (Supplemental Figure 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.92688DS1). The additional 41 primary tumors from patients with nonmetastatic disease at the last documented follow-up were denoted as indeterminate (ID-chRCC) due to the heterogeneous clinical follow-up information and the unknown incidence of late recurrence among these patients (Supplemental Figure 1). For the validation of results observed from this 79-patient cohort, genomic and updated clinical data of 66 chRCC cases from the published Cancer Genome Atlas kidney chromophobe (TCGA-KICH) study (5) were analyzed, among which 12 developed metastasis (M-chRCC) and the remaining 54 nonmetastatic chRCC were also denoted as ID-chRCC. Detailed clinical data can be found in Supplemental Table 1.

WGS on 5 M-chRCC cases identifies recurrent genomic aberrations. To investigate potential genomic features that underlie chRCC metastasis, we performed WGS using DNA purified from available fresh-frozen specimens representing 5 M-chRCC cases (4 primary tumors, JHCHR3, 4, 6, and 7; and 1 metastatic tumor, JHCHR5) along with matched normal tissue or blood (Figure 1 and Supplemental Figures 2 and 3). Tumors were sequenced to a mean haploid depth coverage of ~90×, while normal samples were sequenced to ~50×. Large-scale subclonal copy number changes in WGS samples were analyzed using the Battenberg algorithm, which also supports the detection of subclonal copy number events (23). Subclonal analysis was performed using an n-dimensional Bayesian Dirichlet process to group clonal and subclonal mutations (24). The fraction of tumor cells carrying each mutation was calculated from the mutant allele fraction, taking into account the tumor purity and local copy number state (23, 24).

A wide range of genomic rearrangements (0–149, mean 61), insertions/deletions (148–776, mean 334), and nucleotide substitutions (908–6,578; mean 2,725) were identified (Figure 1 and Supplemental Figures 2 and 3). Recurrent somatic events were mutations in TP53 and PTEN, and homozygous loss of CDKN2A (Figure 1B). Surprisingly, only 1 tumor exhibited a diploid genome with typical chRCC-7set–chromosomes loss (JHCHR5), while all other tumors had ploidy estimates above 2 (2.49–3.55, mean 3.05) (Figure 1B). This observation is in stark contrast to the known prevalent ~7-chromosome loss in chRCC that results in a hypodiploid genome (5, 18). However, B allele frequency (BAF) and copy number analyses (25) suggested gains in chromosomes that resulted in polyploid genomes, yet grossly exhibited the classical pattern of chRCC-7set–chromosome loss (Figure 1 and Supplemental Figure 2). This could be due to either loss of the chRCC-7set chromosomes followed by duplication of the remaining genome (JHCHR3, 4, and 7) (Figure 1B), or duplication of multiple chromosomes excluding the chRCC-7set (JHCHR6) (Figure 1B). To our knowledge, this pattern of imbalanced sequential losses and gains of sets of multiple chromosomes has not been observed in other cancer types (19, 26). Hence, we refer hereafter to this nonrandom duplication of 3 or more chromosomes in chRCC as imbalanced chromosome duplication (ICD). Altogether, our WGS analysis of 5 M-chRCC cases revealed recurrent somatic genomic events including mutations in TP53 and PTEN, homozygous loss of CDKN2A, and ICD.

TP53 and PTEN mutations are enriched in the primary tumors of M-chRCC patients. To further evaluate these potentially high-risk genomic features of chRCC, we assembled an additional cohort consisting of 74 formalin-fixed paraffin-embedded (FFPE) primary tumors from 33 M-chRCC and 41 nonmetastatic ID-chRCC patients (Supplemental Figure 1). Clinical features of this exploratory cohort (n = 74) are summarized in Table 1. Primary tumors and matched normal tissues were analyzed by ultradeep next-generation sequencing
Figure 1. Whole-genome sequencing analysis of 5 metastatic chromophobe renal cell carcinoma (M-chRCC) cases in the exploratory cohort. (A) Diagrams summarize the somatic alterations identified by the whole-genome sequencing (WGS) analysis of primary tumors from patients JHCHR3, JHCHR4, and JHCHR6. The uppermost track depicts inter- and intrachromosomal rearrangements as arcs. The copy number (CN) track shows the number of paternal and maternal copies of each chromosome (blue and red bars, respectively). The other tracks show total CN log-ratio (logR) and B allele frequency (BAF). See supplemental methods for additional details. (B) Clonality analysis of 5 M-chRCC cases based on WGS data. Patient identification numbers are listed along with corresponding purity and ploidy estimates for the tumors generated with the Battenberg algorithm. Each branch represents a subclone and is annotated with alterations that could be assigned to the corresponding subclone. The length is proportional to the number of substitutions assigned by Dirichlet process clustering (see supplemental methods and Supplemental Figure 3 for details).
their roles in metastatic progression of chRCC. Of note, more than 1 TP53 mutation was detected in the primary tumors of 6 M-chRCC and 3 ID-chRCC cases (Figure 2A), reminiscent of the convergent mutation evolution of 3p tumor suppressors and the resulting intratumor heterogeneity in ccRCC (28).

The role of TP53 and/or PTEN mutations in metastatic evolution of chRCC is further evaluated through sequencing paired primary-metastatic tumor samples. To this end, we analyzed 11 available metastatic tumors from 8 paired primary-metastasis cases (Figure 2B, Supplemental Figure 1, and Supplemental Table 4). TP53 mutations were detected in 9 (82%) of 11 tumors from 7 (88%) of 8 patients, and PTEN mutations were in 5 (45%) of 11 tumors from 4 (50%) patients (Supplemental Table 4). In the corresponding 8 primary tumors, 5 (63%) had TP53 and 2 (25%) had PTEN mutations (Figure 2A). In summary, both TP53 and PTEN mutations were more frequent in the metastatic tumors compared with the matched primary tumors.

To examine the spatiotemporal sequence of genomic events concerning metastatic progression, we were able to reconstruct the subclonal composition of 5 of the 8 paired primary-metastasis cases that have high tumor purity in both primary and metastatic samples (Figure 2B, and Supplemental Table 4). In patient 1, the TP53 R273H mutation was detected in the most recent common ancestor (MRCA) clone and hence present in both primary and metastatic lymph node samples (29). Private somatic events, such as NTRK1 and PLCG2 mutations, were present in either primary or metastatic samples, indicating branched evolution (30, 31). In patients 2 and 3, TP53 mutation was absent in the primary tumor but present in the metastatic lymph nodes. In patient 4, the same TP53 N310Tfs mutation was present in primary and metastatic lymph node samples, whereas 5 different PTEN mutations were detected and distributed among samples, demonstrating convergent mutation evolution of PTEN (28, 31, 32). In patient 5, both TP53 and PTEN mutations were detected in the dominant clone and hence present in both primary and metastatic liver samples. ICD was detected in the primary but not the liver-metastasis tumor of patient 5. Notably, deletion of the chRCC-7set chromosomes was evident in all MRCA of these 5 phylogenies, further supporting the role of multiple chromosome loss as a very early oncogenic driver event in the pathogenesis of chRCC.

Cross-platform analysis validates ICD in chRCC. In addition to TP53 and PTEN mutations, our WGS also identified ICD as a recurrent somatic event in M-chRCC (Figure 1). To evaluate chromosomal events, we employed fraction and allele-specific copy number (ASCN) estimates from tumor sequencing (FACETS) (33), a novel statistical and computational pipeline for analyzing ASCNs from NGS data of tumor-normal sample pairs. To validate chromosome-level events beyond NGS platforms (WGS, WES, and MSK-IMPACT), an orthogonal approach on assessing genome-wide DNA copy number alterations and allelic imbalances was performed using the OncoScan FFPE Array (Affymetrix) in samples with

Table 1. Clinical and histopathologic characteristics of chRCC patients included in the exploratory cohort

<table>
<thead>
<tr>
<th></th>
<th>M-chRCC (n = 33)</th>
<th>ID-chRCC (n = 41)A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Age (min, max)</td>
<td>55 (39, 82)</td>
<td>55 (29, 86)</td>
</tr>
<tr>
<td>Median Tumor Size (min, max)a</td>
<td>12 (2.5, 29)</td>
<td>7.3 (1.2, 21.5)</td>
</tr>
<tr>
<td>Gender</td>
<td>16 males (48.5%)</td>
<td>25 males (65.8%)</td>
</tr>
<tr>
<td></td>
<td>17 females (51.5%)</td>
<td>13 females (34.2%)</td>
</tr>
<tr>
<td>AJCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical information</td>
<td>NA = 0</td>
<td>NA = 3</td>
</tr>
<tr>
<td>Histology</td>
<td>NA = 0</td>
<td>NA = 0</td>
</tr>
<tr>
<td>Eosinophilic</td>
<td>4 (12.1%)</td>
<td>11 (28.9%)</td>
</tr>
<tr>
<td>Sarcomatoid features</td>
<td>7 (21.2%)</td>
<td>1 (2.6%)</td>
</tr>
</tbody>
</table>

A Clinical information of the nonmetastatic ID-chRCC cohort demographic information was summarized based on 38 cases due to insufficient clinical information in 3 of the 41 cases. aMedian pathologic tumor size (cm). M-chRCC, metastatic chromophobe renal cell carcinoma; ID-chRCC, indeterminate chRCC; AJCC, American Joint Committee on Cancer; NA, not available.
sufficient remaining DNA (34). Furthermore, fluorescence in situ hybridization (FISH) assays utilizing chromosome-specific probes were conducted in select samples with sufficient numbers of available unstained slides to visualize chromosome gains or losses. Altogether, FACETS results were independently validated with OncoScan in 17 samples and with FISH analysis over selectively deleted or amplified chromosomal regions in 14 samples (Supplemental Table 1).

The overall estimates of total copy number and BAF were comparable between OncoScan and FACETS (Figure 3). One unique feature of FACETS is the integer copy number output, where the total copy number and the minor allele copy number were corrected for tumor purity, ploidy, and clonal heterogeneity (33). Indeed, the loss of 1 copy of chromosome 2 detected by OncoScan and FACETS (Figure 3, A and B) in the tumor was visualized with FISH assays using a 2p23/ALK probe (orange), and the retention of 2 copies of chromosomes 1q and 4 was confirmed using a 1q23/NTRK1 (red) and a chromosome 4 pericentromere (green) probe, respectively (Figure 3C). There was no ICD detected in this patient’s tumor (Figure 3, A–C).

In contrast, ICD was detected in another patient’s tumor (Figure 3, D–F). The total copy number plots from OncoScan (Figure 3D) and FACETS (Figure 3E) showed a diploid pattern, whereas the
corresponding BAF plots showed widespread loss of heterozygosity (LOH). The chRCC-7set chromosomes (1, 2, 6, 10, 13, 17, and 21) exhibited the most conspicuous allelic imbalance, whereas a range of allelic imbalance was observed in the other chromosomes (Figure 3, D and E). For example, the near-diploid copy number of chromosome 4 with a wide-split BAF indicated extreme allelic imbalance and suggested a loss of chromosome 4 followed by a duplication of the remaining chromosome 4 in most cells, whereas the increased copy number plot on chromosome 14 with a moderate-split BAF suggested a gain of 1 copy of chromosome 14 without a prior chromosomal loss (Figure 3, D and E). Indeed, FISH analysis (Figure 3F) demonstrated the presence of 2 copies of chromosome 4 (4p16/FGFR3 probe) and 3 copies of chromosome 14 (14q32/IGH probe).

The prevalence of ICD in chRCC is evaluated in our chRCC and the TCGA-KICH cohorts. Our data demonstrated that the FACETS algorithm offers a reliable resource for analyzing ASCNs based on targeted NGS panel data. Accordingly, we performed FACETS on all of our samples and the TCGA-KICH cohort to assess ICD. All FACETS results were manually inspected, and the samples were assigned as ICD (+) status if 3 or more chromosomes were duplicated. Samples with a tumor purity estimate less than 30% or large amount of noise and artifacts were excluded and assigned a not available (NA) status (Figure 2A, Table 1, and Supplemental Figure 4). Within the exploratory cohort (n = 74), ICD was present in 9 (12%),
absent in 50, and NA in 15 primary kidney tumors (Figure 2A). Among the 59 primary tumors with evaluable ICD status, ICD was present in 6 (25%) of 24 M-chRCC and in 3 (9%) of 35 nonmetastatic ID-chRCC cases. Within the TCGA-KICH cohort, there were 66 primary kidney tumors, among which ICD was present in 13 (20%) and absent in 53 cases (Supplemental Figure 4). ICD was present in 7 (58%) of 12 M-chRCC and in 6 (11%) of 54 nonmetastatic ID-chRCC cases within the TCGA-KICH cohort. Altogether, our results show a 2.6- to 5.3-fold increase in the incidence of ICD in primary tumors of M-chRCC versus ID-chRCC patients, suggesting a role for ICD in the metastatic progression of chRCC.

The presence of TP53 mutations, PTEN mutations, or ICD in the primary tumors associates with inferior survival outcome in the TCGA-KICH cohort. To evaluate the impact of TP53 and PTEN mutations and ICD on chRCC patient survival, we analyzed the TCGA-KICH cohort (n = 64) (Figure 4 and Supplemental Table 5). Of note, 2 of the originally reported 66 cases (5) were excluded from survival analysis due to insufficient clinical information (Supplemental Figure 4 and Supplemental Tables 1 and 5). When stratified based on the TP53 mutation status, overall survival (OS) favored TP53 wild-type over TP53 mutant chRCC cases (P = 0.017), and 5-year OS of TP53 wild-type versus mutant chRCC was 94% (95% CI, 87%–100%) versus 67% (48%–93%) (Figure 4A and Supplemental Table 5). When stratified based on PTEN mutation status, OS favored wild-type over mutant cases (P = 0.034), and 5-year OS of PTEN wild-type versus mutant was 89% (81%–99%) versus 57% (30%–100%) (Figure 4B and Supplemental Table 5). When stratified based on ICD, OS favored ICD (–) over (+) cases (P = 0.012), and 5-year OS of ICD (–) versus (+) was 91% (83%–100%) versus 55% (30%–100%) (Figure 4C and Supplemental Table 5). When all 3 high-risk features were combined, OS favored cases without any of these high-risk features over those with any, but does not reach statistical significance (P = 0.059), and 5-year OS of patients with no versus any feature was 97% (90%–100%) versus 74% (58%–93%) (Figure 4D and Supplemental Table 5). These data support the role of these 3 genomic features in the clinical outcomes of chRCC. Notably, although histology reports on the sarcomatoid status were missing in 40 of the 64 TCGA-KICH patients (Figure 4 and Supplemental Table 1), all 3 confirmed sarcomatoid cases have ICD, which is consistent with the reported high incidence of chromosomal gains in sarcomatoid chRCC tumors (35).

The presence of TP53 mutations, PTEN mutations, or ICD in the primary tumors associates with inferior survival in M-chRCC patients. We next evaluated the association between TP53 mutations, PTEN mutations, or ICD and survival outcome in our M-chRCC cases cohort (n = 35) (Figure 5 and Supplemental Tables 1 and 6). When stratified based on the TP53 mutation status, the OS favored TP53 wild-type over TP53 mutant, but this association did not reach statistical significance (P = 0.099) (Figure 5A and Supplemental Table 6). When stratified based on PTEN, OS was similar between wild-type and mutant cases (P =
When stratified based on ICD, OS favored ICD (−) over (+) cases ($P = 0.003$) (Figure 5C and Supplemental Table 6), and when stratified based on sarcomatoid status, OS favored sarcomatoid histology (−) over (+) cases ($P = 0.043$) (Figure 5D and Supplemental Table 6). Notably, cases with any of the 3 high-risk genomic features include all sarcomatoid cases (Figure 2A), and OS favored cases without any of these high-risk features over those with any ($P = 0.013$) (Figure 5E and Supplemental Table 6).

Transcriptomic analysis of the TCGA-KICH cohort identifies expression signature associated with the ICD status. To explore the potential molecular feature associated with ICD, we analyzed gene expression data from RNA sequencing by comparing ICD (+) to ICD (−) chRCC within the TCGA-KICH cohort. We identified 976 differentially expressed genes ($P < 0.01$, t test; 451 up, 525 down) that were associated with a number of pathways (Figure 6A and Supplemental Table 7). For example, genes involved in cell proliferation were higher in ICD (+) chRCC, whereas those involved in mitochondrial function, located at chromosome 19q13 locus, and targeted by TP53 were lower (Figure 6A). When evaluating the association between transcript levels and DNA methylation at corresponding CpG islands, coordinated changes were detected in chromosome 19q13 genes (Figure 6B and Supplemental Table 7). Notably, the downregulation of 19q23 genes was shown to associate with SETD2 mutations in clear or papillary kidney cancers (20), whereas SETD2 mutation was not detected in chRCC (Supplemental Table 3) (5, 20).
There are shared molecular features between the ICD (+) cases of chRCC in the TCGA-KICH cohort and the whole-genome duplication (+) cases of ccRCC in the TCGA-KIRC cohort. Recent pan-cancer analysis of somatic copy number alterations demonstrated that a majority of cancer is nearly diploid, whole-genome duplication (WGD) occurs in 37% of cancers with a ~3.31 mean ploidy, and recurrent whole-chromosome loss is uncommon (26). Of note, WGD within the TCGA kidney renal clear cell carcinoma (TCGA-KIRC) cohort was 20%. As ICD and WGD both manifest with gains of many chromosomes, we determined their molecular relationship in RCC and identified striking transcriptomic similarities between the ICD (+) chRCC and the WGD (+) ccRCC. Among the 451 upregulated genes in ICD (+) chRCC, 401 (89%) were up in WGD (+) ccRCC compared with WGD (–) ccRCC, and among the 525 downregulated genes in ICD (+) chRCC, 415 (79%) were down (χ² test, \( P < 1 \times 10^{-30} \)) (Figure 6A). Hence, the molecular signature conferred by ICD in chRCC is highly concordant with that by WGD in ccRCC, implying a shared pathway convergent evolution between 2 kidney cancer subtypes (28, 36).

WGD (+) ccRCC patients in the TCGA-KIRC cohort exhibit inferior survival outcome. Our study thus far identified ICD as a candidate high-risk genomic feature of chRCC (Figure 2A and Supplemental Figure 4) and recognized the marked similarity in transcriptomes between chRCC with ICD and ccRCC with WGD (Figure 6A). We next investigated whether the presence of...
WGD in the primary tumors impacts the survival outcome of ccRCC patients, which to our knowledge has not been interrogated. Kaplan-Meier survival curves of the TCGA-KIRC cohort patients stratified based on the WGD status of primary tumors showed a statistically significant inferior survival with WGD (+) ccRCC patients ($P < 0.0001$) (Figure 6C and Supplemental Table 7). Notably, WGD also negatively impacts the survival of colorectal and adrenocortical cancer patients (37, 38). Here we demonstrated the shared transcription signature and inferior survival between ICD (+) chRCC and WGD (+) ccRCC, implicating ICD as a WGD variant. How and whether ICD/WGD actively participates in RCC cancer progression and impacts clinical outcome remain to be determined.

Discussion

The principal cause of cancer death is distant metastasis, which entails diverse mechanisms and invokes distinct genes/pathways that could be either common to all cancers or relatively unique to individual cancer types. Alteration in chromosome number is a cardinal feature of human cancer and could reflect underlying genome instability (39). However, as chromosomal loss could compromise cellular proliferation (22), we had long been intrigued by how chRCC, a tumor characterized by many whole-chromosome losses, can metastasize and kill patients. Due to its rarity, contemporary genomic reports on chRCC are limited to smaller cohorts that mainly consist of nonmetastatic cases (5). Here, we investigated the genomic features underlying the metastatic progression of chRCC through comparing genomics of the primary tumors between M-chRCC patients and those who had not developed metastasis (ID-chRCC). A WGS cohort of 5 M-chRCC cases and an MSK-IMPACT cohort of an additional 74 cases (33 M-chRCC, and 41 nonmetastatic ID-chRCC) were compiled in addition to the publically available TCGA-KICH cohort of 66 cases (12 M-chRCC and 54 nonmetastatic ID-chRCC). Upon genomic analysis of the primary tumors from 144 chRCC patients (49 M-chRCC and 95 nonmetastatic ID-chRCC), increased frequencies of TP53 mutations, PTEN mutations, and ICD were noted in M-chRCC (55%, 27%, and 43%) compared with nonmetastatic ID-chRCC (25%, 7%, and 10%). Moreover, these 3 high-risk genomic events negatively impacted the survival outcome, lending clinical significance concerning these metastasis-enriched genomic features in chRCC.

Our study confirmed the nearly universal loss of chRCC-7set chromosomes in the MRCA clone of all genomically examined ccRCC, which probably denotes a very early driver event in the pathogenesis of chRCC, and determined an increasing incidence of TP53 and PTEN mutations during the metastatic progression of chRCC. Phylogenetic studies of paired primary-metastatic samples demonstrated that TP53 mutations were detected in 82% of M-chRCC samples, indicating that TP53 mutation is the second most favored genomic event. Interestingly, after the chRCC-7set–chromosomes loss and the TP53 mutation, ICD and PTEN mutation were the third most common genomic events and they appeared to occur in a mutually exclusive manner (Fisher, $P = 0.0033$) (Supplemental Figure 5), suggesting a functional convergence through PTEN mutation or ICD. As TP53 and PTEN are located at chRCC-7set chromosomes 17 and 10, heterozygous mutations in either gene would result in complete functional loss, which might explain the observed preferential mutations at these 2 tumor suppressor genes (TSGs) more than the other TSGs such as CDKN2A located at chromosome 9.

Although chRCC is relatively indolent, once metastasized the survival of M-chRCC patients is as poor as metastatic ccRCC (13). Due to the rarity (~5% of RCC) and the low (5%–10%) incidence of eventual metastases, only limited clinical parameters such as tumor sizes and sarcomatoid changes were shown to affect clinical outcome (40), and no specific follow-up guideline has been developed for chRCC patients. Currently, nonmetastatic RCC patients of all histological subtypes after surgery were followed according to contemporary guidelines that are generalized and formulated based on integrated prognostic algorithms developed for ccRCC (41, 42). Within the TCGA-KICH cohort, patients whose primary tumors exhibited none of the 3 high-risk features, i.e., TP53 mutations, PTEN mutations, or ICD, had a 5-year OS of 97%, whereas those with any had a 5-year OS of 74%. Furthermore, within our M-chRCC case cohort, patients whose primary tumors exhibited none of these features had an excellent 5-year OS of 100% and those with any had a dismal 5-year OS of 37%. Thus, TP53 mutation, PTEN mutation, and ICD might have prognostic value in chRCC. Likewise, BAP1, PBRM1, SETD2, and TP53 mutations in ccRCC (43–45), the CpG island methylator phenotype (CIMP) in pRCC (4), and the NF2 mutations in uRCC (8) are genomic features that have been shown to carry clinical significance in an RCC-subtype-specific manner. Thus far, all these genomic discoveries were made in retrospective studies utilizing archived tumor samples, which warrant
further validation. Once validated, they would be invaluable in the future construction of individualized follow-up plans and treatment algorithms for kidney cancer patients.

Methods

Tumor samples. A total of 90 chRCC tumor samples, 5 frozen and 85 FFPE tumor samples along with their matched germline blood or uninvolved tissue, representing 79 patients, were collected under respective IRB protocols to undergo NGS. Forty-nine tumor samples, including 37 primary and 12 metastatic tumors from 38 patients with advanced chRCC (M-chRCC), were collected at the Memorial Sloan Kettering Cancer Center (24 cases), Mayo Clinic (7 cases), and UCLA (7 cases). The remaining 41 primary tumors were derived from 41 nonmetastatic chRCC patients who presented with localized disease at the time of sampling and the last documented follow-up. Each slide was re-reviewed by a single genitourinary pathologist (S.K.T.) to confirm histology of chRCC and tumor content before microdissection. Equivocal lesions were excluded from further DNA extraction and preparation. Clinical features of the exploratory cohort ($n = 74$) are summarized in Table 1 and Supplemental Table 1.

WGS. WGS of fresh-frozen tumors and matched normal tissues from 5 M-chRCC cases (4 primary tumors, 1 metastasis) was performed by the New York Genome Center on an Illumina HiSeq 2500. Tumor samples were sequenced to mean haploid depth coverages of $87\times$ and $91\times$, while normal samples were sequenced to $51\times$ and $47\times$. More than 96% of the genome was sequenced to at least $10\times$ in all cases. Short-insert paired-end reads were aligned to the GRCh37 reference human genome with 1,000 genomes decoy contigs using BWA-mem (46). For further information on somatic mutation calling for substitutions, insertions, deletions, structural rearrangements, and copy number changes, please refer to the Supplemental Data section.

Custom target sequencing. Seventy-four primary tumor and 11 metastasis samples were analyzed using MSK-IMPACT, as previously described. This clinical sequencing assay is a hybridization capture-based NGS assay for targeted deep sequencing of all exons and selected introns of 341 or 410 oncogenes, TSGs, and members of pathways deemed actionable by targeted therapies (Supplemental Table 2). Unstained sections from paraffin-embedded tissue blocks were microdissected to ensure tumor purity. DNA was purified using the DNeasy Blood & Tissue Kit (Qiagen, 69504). Genomic DNA from tumor and patient-matched normal samples underwent library preparation using the KAPA HTP protocol (Kapa Biosystems) and the Biomek FX system (Beckman Coulter). Custom DNA probes targeting exons and selected introns were synthesized using the NimbleGen SeqCap EZ library custom oligo system and are listed in Supplemental Table 2. Details on massive parallel sequencing and alignment of postcapture libraries on the Illumina HiSeq 2500 and sequence data analysis are described in the Supplemental Methods section. The impact-targeted DNA sequencing data is available for public access at cBioPortal (http://cbioportal.mskcc.org/study?id=chrcc_mskcc_2015#summary) (47). Any other data that are included in the article and its supplemental material will be available from the authors upon request.

FACETS for somatic copy number analysis. FACETS is an ASCN analysis tool and open-source software with a broad application to WGS, WES, as well as targeted-panel sequencing platforms developed by V.E.S. (33). We employed FACETS for WES data from 66 primary tumor samples from the TCGA-KICH project and for 73 primary tumor and 12 metastases samples profiled by the MSK-IMPACT platform. In both sample types analyzed, FACETS enhanced the sensitivity to facilitate the identification of aneuploid tumors by joint modeling of total and allele-specific patterns. In addition, FACETS enabled the systematic identification of clonal and subclonal copy number events through a cellular fraction feature in the model. Moreover, reliable interpretation of NGS-based gene copy number calls in clinical sequencing panels was possible due to accurate, purity-, ploidy-, and clonal heterogeneity-adjusted, integer copy number calls delivered by FACETS. For detailed information please refer to the related publication (33).

In brief, the key steps for joint-segmentation and ASCN analysis were as follows. (a) Generation of total copy number log-ratio (logR): Sequence read count information was first parsed from paired tumor-normal BAM files, with a normalizing constant being calculated for each tumor-normal to correct for total library size. To reduce hypersegmentation in SNP-dense regions of the genome, subsampling within a 150- to 250-bp interval was applied. logR was then computed from the total read count in the tumor versus normal for all SNPs that had a minimum depth of coverage in the normal. (b) Generation of allelic copy number log odds-ratio (logOR): To avoid significant bias toward higher mapping rates for the reference allele compared with those for the variant allele at heterozygous loci, we used the logOR of the variant-allele count in
tumor versus normal, which is an unbiased estimate of allelic copy ratio. (c) Joint segmentation analysis: To identify regions of the genome that have constant copy number using change-point detection methods, we employed a model distinct from the conventional method in which both logR and BAF are computed for heterozygous loci only. In fact, we extended the circular binary segmentation (CBS) algorithm (48) to a joint segmentation of logR and logOR based on a bivariate Hotelling T2 statistic. After segmentation, we clustered the segments into groups of the same underlying genotype to reduce the number of latent copy number and cellular fraction states needed in subsequent modeling.

**FACETS call of integer copy number.** To model both clonal and subclonal events, we obtained integer copy number (major and minor) and the associated cellular fraction estimates for each segment cluster by modeling the expected values of logR and logOR given total (t), and each parental (m,p) copy as a function of a cellular fraction (cf) parameter \( \phi \), using a combination of parametric and nonparametric methods. This allowed us to model both clonal and subclonal events.

**Call of ICD.** Based on the outputs of FACETS, we manually analyzed the samples to assess ICD. We excluded all samples with purity estimation below 0.3 (tumor DNA content < 30%) and/or with high level of noise due to purity issues in order to avoid confounders due to hypersegmentation. Excluded samples were called NA. The logR was compared to the logOR and the copy number output. If copy number changes greater than 2 were computed for chromosomes with corresponding LOH, the estimates of cancer cell fraction (CCF) for these specific regions were evaluated to assess reliability of duplication. A finding of 3 or more chromosomes duplicated in the genome of 1 sample was called ICD.

**Affymetrix OncoScan.** The performance of our ASCN analysis method FACETS was validated with 2 independent methods on selected samples. First we analyzed genome-wide DNA copy number alterations and allelic imbalances by SNP array using the Affymetrix OncoScan FFPE Assay in 17 samples (34). This is a molecular inversion probe technology for the identification of copy number alterations and LOH. We used 80 ng of genomic DNA for each sample. Processing of samples was performed according to the manufacturer’s guidelines (Affymetrix). Assay data were analyzed by the combination of OncoScan Console software (Affymetrix) and OncoScan Nexus Express software (BioDiscovery) using the Affymetrix TuScan algorithm. OncoScan SNP-array data were analyzed by the software couple, OncoScan Console and Nexus Express, using the Affymetrix TuScan algorithm. All array data were also manually reviewed for subtle alterations not automatically called by the software. The total copy number and BAF plots generated by this platform allow overcoming the low resolution of the analysis originated using MSK-IMPACT data and confirm the copy number analysis by FACETS.

**FISH.** To further leverage our copy number findings from FACETS, FISH was performed on 14 FFPE samples using probes targeting chromosome loci that are supposed to be informative for verifying copy number changes demonstrated by FACETS. FISH probes employed in this study included a 3-color in-house probe for 1q23 (NTRK1 locus), 2p23 (ALK locus), and pericentromere of chromosome 4 (4p11 locus), and a 2-color commercial probe purchased from Abbott Molecular for 4p16 (FGFR3 locus) and 14q32 (IGH locus). Probe labeling, tissue processing, hybridization, posthybridization washing, and fluorescence detection were performed according to standard procedures. Slides were scanned using a Zeiss Axioplan 2i epifluorescence microscope equipped with a megapixel CCD camera (CV-M4+CL, JAI) controlled by Isis 5.5.9 imaging software (MetaSystems Group Inc.). The nontumor regions or normal cells (stromal cells and infiltrating lymphocytes) within tumors served as internal controls to assess quality of hybridization. The entire section was scanned under an \( \times63 \) or \( \times100 \) objective to assess copy number and representative regions imaged through the depth of the tissue. Signal enumeration was performed on the captured images and a minimum of 100 discrete nuclei analyzed. FISH analysis with these probes enabled us to conclusively confirm copy number changes detected by sequencing data in single cells from primary and metastatic tumors.

**Reconstruction of subclonal composition.** To reconstruct the subclonal composition of the analyzed tumors, we based our assumptions on previously published methods for subclonal reconstruction (49, 50). Given the high coverage of the MSK-IMPACT assay (432× for samples included in the analysis), we consider the subclonal reconstruction feasible. Segments of copy number changes of 20-markers long or less were discarded. Segments with an insufficient number of heterozygous SNPs and where FACETS results did not estimate minor copy number call were excluded. Alterations (somatic copy number and gene mutations) with CCF above 75% were considered clonal (100%), as previously described (51). Pairs of copy number alterations or mutations with CCF within 10% of each other
were considered to be in the same subclone with CCF equal to average value (there were no triplets of mutations within 20% range of each other). Copy number alterations and mutations with CCF within 5% were considered to be the same subclone.

We calculated CCF based on the formula, $CCF = \frac{V(CR + 2[1 - R])}{R}$, where $V$ is the variant allele frequency of the mutation, $R$ is the tumor purity, and $C$ is the local copy number. For mutations on the X chromosome in males, $CCF = \frac{V}{R}$. Multiple subclones with CCF less than 50% are consistent with more than 1 possible scenario: nested subclones (mutations in the same cells), shown in figure for simplicity, and parallel subclones (mutations in disjoint sets of cells). Since the copy number estimates are based on a targeted-sequencing panel, complex copy number changes could not be estimated with sufficient precision. Therefore, specific chromosomes were excluded from the reconstruction, given complex copy number patterns estimated by FACETS in one of the tumors: patient 1 (P-0002449), chromosome 11; patient 2 (JHkic60/P-0003003), chromosomes 5p, 8, 9p, 16p, 19q; patient 3 (JHkic59), chromosomes 1q, 5, 9, 10q, 12p, 15, 17q, 22, 23; patient 4 (JHnceRCC004), chromosomes 5p, 21, 23; patient 5 (P-0002825), chromosomes 2q, 3, 4q, 6q, 16, 19p.

Statistics. To address the question, whether detected mutations in chRCC are associated with metastatic chroomphobe disease, we analyzed a cohort of 99 patients, upon removal of patients with missing clinical information or indeterminate clinical status. Of these patients, the 64 TCGA patients (TCGA-KICH cohort) (52) with available clinical information served as a population-based sample of controls, with 0 samples missing ICD status. Thirty-five metastatic patients from Mayo, Memorial Sloan Kettering Cancer Center, and UCLA served as the sample of cases (M-chRCC). Of these patients, 9 are missing ICD status and were therefore excluded from any analysis of this variable. OS was calculated from date of diagnosis of renal cancer to date of death or last documented follow-up. The Kaplan-Meier method estimated OS probability and the log-rank test was used to compare OS between groups. Cox regression was used to estimated hazard ratios (HRs). A $P$ value less than 0.05 was considered statistically significant.

Study approval. For the sequencing study of all cases from Memorial Sloan Kettering Cancer Center, patients signed an informed consent (06-107 or 89-076) and the study was performed under a protocol reviewed and approved by the Memorial Sloan Kettering Cancer Center's IRB under IRB protocol WA-0395-12. Data and samples collected from the Mayo Clinic and UCLA were exempted from IRB approval as all samples were anonymous and completely de-identified.

Transcriptomic analysis of chRCC. To compare ICD (+) and ICD (–) chRCC samples within the TCGA-KICH (52) cohort, we analyzed the gene expression profiles by RNA sequencing in both groups stratified by our ICD calls based on FACETS analysis. Furthermore, we also analyzed WGD (+) and WGD (–) ccRCC samples from the TCGA-KIRC (53) cohort in the same fashion (26). FDRs corresponding to the differential genes were computed using a previously described method (54) and are listed in Supplemental Table 5. Of the ICD differential genes at a significance level of $P$ less than 0.01, true positives should be represented on the order of 80%.

Pathway enrichment testing was by 1-sided Fisher’s exact test ($P < 0.01$ significance level), using SigTerms software tool (55). Significance of overlap between ICD and WGD for upregulated genes was $P$ less than 0.0001 (169 genes intersecting), and for downregulated genes $P$ less than 0.0001.

Author contributions
JC, EHC, AAH, and JJH were responsible for the conception and design of the study. Acquisition of data was performed by JC, AR, AAH, PR, NMD, JCC, RHT, AJP, MEA, BJM, and SKT. Analysis and interpretation of data was performed by JC, NW, GG, ECZ, PIW, ED, GJN, WL, IO, VS, LW, CJC, and EP. Statistical analysis was performed by ECZ, ED, IO, and VS. Drafting of the manuscript was carried out by JC, AN, and JJH. Critical revision of the manuscript for important intellectual content was carried out by GG, ECZ, EHC, IO, EP, CJC, AAH, and JJH. Administrative, technical and material support was provided by AR, SKT, LW, GJN. JJH supervised the work.

Acknowledgments
We thank our patients for donating their tumors. This study is funded by the Tuttle Family Fund for Rare Kidney Cancer Research. JC is sponsored by the German Research Foundation (DFG) grant CA1403/1-1. BJM is supported by the Ruth L. Kirschstein National Research Service Award T32CA082088. CJC is supported by NIH grant P30 CA125123. JJH is supported by The Jill and Jeffrey Weiss Fund to the Cure...
of Kidney Cancer and J. Randall & Kathleen L. MacDonald Kidney Cancer Research Fund. The authors also wish to acknowledge assistance from the Bioinformatics Core, which is funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

Address correspondence to: James J. Hsieh, Medical Oncology Division, Department of Medicine, Campus Box 8069, Washington University St. Louis, 660 S. Euclid Avenue, St. Louis, Missouri 63110, USA. Phone: 314.273.1688; Email: jhsieh@wustl.edu.

32. Voss MH, et al. Tumor genetic analyses of patients with metastatic renal cell carcinoma and extended benefit from mTOR