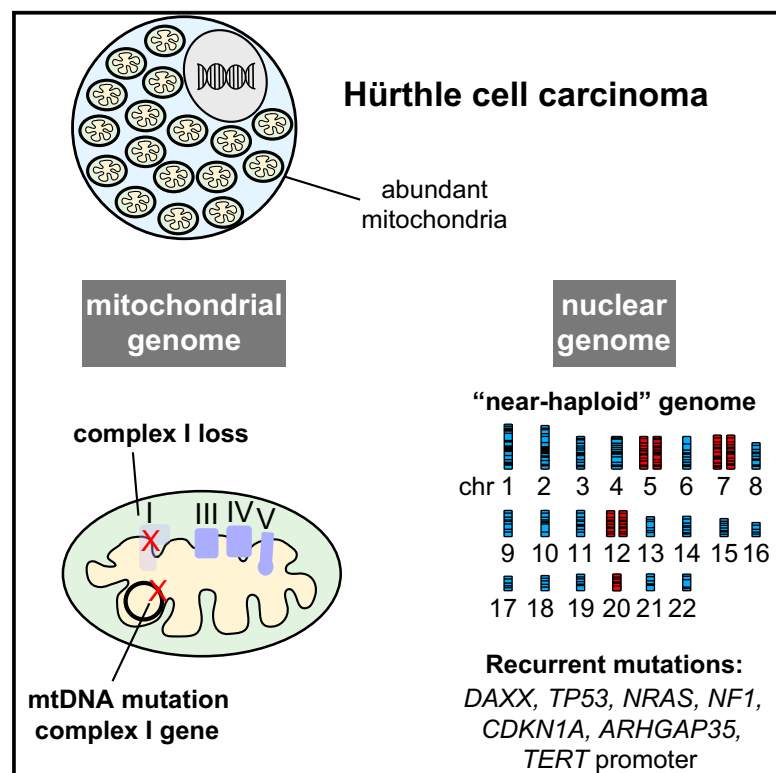# Widespread Chromosomal Losses and Mitochondrial DNA Alterations as Genetic Drivers in Hürthle Cell Carcinoma

Graphical Abstract

Authors

Raj K. Gopal, Kirsten Kübler,
Sarah E. Calvo, ...,
Dora Dias-Santagata, Gad Getz,
David G. McFadden

Correspondence

gadgetz@broadinstitute.org (G.G.),
david.mcfadden@
utsouthwestern.edu (D.G.M.)

In Brief

Gopal et al. identify recurrent alterations in *DAXX*, *TP53*, *NRAS*, *NF1*, *CDKN1A*, *ARHGAP35*, and the *TERT* promoter, as well as in mtDNA-encoding complex I of the electron transport chain, in Hürthle cell carcinomas (HCC). Many HCCs harbor widespread chromosomal loss culminating in a near-haploid state.

Highlights

- HCC is driven by unique alterations of the nuclear and mitochondrial genomes

- Early widespread loss of chromosomes leads to a stable near-haploid state in HCC

- mtDNA mutations in complex I of the electron transport chain are enriched in HCC

- *DAXX*, *TERT*, *TP53*, *NRAS*, *NF1*, *CDKN1A*, and *ARHGAP35* are recurrently altered in HCC

Widespread Chromosomal Losses and Mitochondrial DNA Alterations as Genetic Drivers in Hürthle Cell Carcinoma

Raj K. Gopal,[1,2,6,8,9,11,19] Kirsten Kübler,[2,8,11,19] Sarah E. Calvo,[6,8,9] Paz Polak,[2,4,8,11,16] Dimitri Livitz,[8] Daniel Rosebrock,[8] Peter M. Sadow,[2,4,11] Braidie Campbell,[1,2] Samuel E. Donovan,[1,2] Salma Amin,[2,5] Benjamin J. Gigliotti,[1] Zenon Grabarek,[6,8,9] Julian M. Hess,[8] Chip Stewart,[8] Lior Z. Braunstein,[8,17] Peter F. Arndt,[8,18] Scott Mordecai,[4] Angela R. Shih,[4,11] Frances Chaves,[4] Tiannan Zhan,[7] Carrie C. Lubitz,[2,5,7,11] Jiwoong Kim,[14] A. John Iafrate,[4,11] Lori Wirth,[1,2,11] Sareh Parangi,[2,5,11] Ignaty Leshchiner,[8] Gilbert H. Daniels,[1,2,3,11] Vamsi K. Mootha,[1,6,8,9,10,20] Dora Dias-Santagata,[4,11,20] Gad Getz,[2,4,8,11,20,*] and David G. McFadden[1,3,12,13,15,20,21,*]

[1]Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[2]Cancer Center, Massachusetts General Hospital, Boston, MA 02114, USA
[3]Thyroid Unit, Massachusetts General Hospital, Boston, MA 02114, USA
[4]Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA
[5]Department of Surgery, Massachusetts General Hospital, Boston, MA 02114, USA
[6]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[7]Institute for Technology Assessment, Massachusetts General Hospital, Boston, MA 02114, USA
[8]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[9]Howard Hughes Medical Institute, Chevy Chase, MD, USA
[10]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[11]Harvard Medical School, Boston, MA 02115, USA
[12]Department of Internal Medicine, Division of Endocrinology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[13]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[14]Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[15]Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[16]Present address: Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[17]Present address: Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[18]Present address: Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany
[19]These authors contributed equally
[20]Senior author
[21]Lead Contact
*Correspondence: gadgetz@broadinstitute.org (G.G.), david.mcfadden@utsouthwestern.edu (D.G.M.)
https://doi.org/10.1016/j.ccell.2018.06.013

SUMMARY

Hürthle cell carcinoma of the thyroid (HCC) is a form of thyroid cancer recalcitrant to radioiodine therapy that exhibits an accumulation of mitochondria. We performed whole-exome sequencing on a cohort of primary, recurrent, and metastatic tumors, and identified recurrent mutations in *DAXX*, *TP53*, *NRAS*, *NF1*, *CDKN1A*, *ARHGAP35*, and the *TERT* promoter. Parallel analysis of mtDNA revealed recurrent homoplasmic mutations in subunits of complex I of the electron transport chain. Analysis of DNA copy-number alterations uncovered widespread loss of chromosomes culminating in near-haploid chromosomal content in a large fraction of HCC, which was maintained during metastatic spread. This work uncovers a distinct molecular origin of HCC compared with other thyroid malignancies.

Significance

Hürthle cell carcinoma of the thyroid (HCC) is a form of thyroid cancer with a marked increase in mitochondria. HCC exhibits both a failure to concentrate radioactive iodine and avidity for fluorodeoxyglucose, an imaging signature suggestive of metabolic reprogramming. We performed a parallel analysis of the nuclear and mitochondrial genomes of HCC using exome sequencing. We identified distinct genomic features of HCC, including alterations in mtDNA complex I subunits, and widespread chromosomal losses leading to a near-haploid state. A driver role for these events was further suggested by the maintenance, and convergent evolution, of these genetic features in related primary tumors and distant metastases.

INTRODUCTION

Hürthle cell carcinoma of the thyroid (HCC) is a type of oncocytic neoplasm responsible for approximately 3% of all thyroid cancers, and has traditionally been classified as a variant of follicular thyroid cancer (Wada et al., 2002). Hürthle cells are large, polygonal cells with variable nuclear atypia, and a prominent, granular cytoplasm as a result of mitochondrial abundance (Maximo and Sobrinho-Simoes, 2000). Oncocytic change or metaplasia (referred to as Hürthle cell change in the thyroid gland), reflecting mitochondrial accumulation, occurs in non-neoplastic conditions, such as chronic lymphocytic (Hashimoto's) thyroiditis, as well as benign and malignant neoplasms originating in several organs, including kidney, adrenal gland, salivary gland, thyroid, and others (Gasparre et al., 2010).

Although a diagnosis of minimally invasive HCC (exhibiting invasion only through the tumor capsule or at most a few foci of vascular invasion) portends a low risk of recurrence or metastasis, widely invasive HCC (with extensive vascular invasion) exhibits more frequent recurrences and distant metastases (Stojadinovic et al., 2002). Compared with relatively indolent forms of thyroid cancer, including papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC), HCC exhibits a greater incidence of distant metastases and reduced progression-free survival (Sabra et al., 2016; Shaha et al., 1996). Further complicating the clinical management, HCC infrequently concentrates radioactive iodine, used for both imaging and treatment of well-differentiated thyroid cancer (Haugen et al., 2016; Pryma et al., 2006). Hürthle cell tumors conversely show intense uptake of fluorodeoxyglucose (^{18}F-FDG) by PET imaging (Pathak et al., 2016; Pryma et al., 2006). The increased ^{18}F-FDG uptake and mitochondrial accumulation point to metabolic reprogramming in HCC and suggest unique biological features compared with other forms of thyroid cancer. This notion is supported by the description of a distinct gene expression signature in HCC compared with PTC and FTC (Ganly et al., 2013).

Studies of the somatic genetic alterations in HCC are limited. Prior targeted analyses suggested that HCC harbors distinct alterations compared with other forms of thyroid cancer with infrequent mutations in the mitogen-activated protein kinase pathway, *TP53*, *PTEN*, and *MEN1*, along with infrequent *PAX8/PPARγ* rearrangements (Evangelisti et al., 2015; Kasaian et al., 2015; Nikiforova et al., 2013; Wei et al., 2015). Two genomic features of HCCs, not frequently observed in PTC and FTC, have been reported from studies of small numbers of HCCs. First, mitochondrial-encoded mutations were reported to be a marker of benign and malignant oncocytic thyroid tumors. However, whether these mutations act as passengers or drivers has not been established (Gasparre et al., 2007). In addition, loss of heterozygosity (LOH) across many chromosomes was reported in oncocytic thyroid tumors (Corver et al., 2012, 2014). Whether the mitochondrial or LOH events occur in the context of other genomic alterations, and whether these events are prevalent in recurrent and metastatic disease has not been addressed. In this study, we utilized whole-exome sequencing (WES) to characterize the genomes of surgically resected HCC primary tumors, regional recurrences, and distant metastases. We performed a parallel analysis of the nuclear and mitochondrial genomes of primary and metastatic tumors to uncover the genetic drivers of HCC.

RESULTS

Recurrent Nuclear-Encoded Mutations

Exome sequencing was performed on 70 HCC specimens obtained from 41 patients, as well as matched DNA isolated from blood or adjacent normal thyroid (Table S1). The cohort included 32 (47%) primary tumors (P), 13 (19%) distant metastases (DM), and 25 (34%) locoregional recurrences (LR), thereby representing various disease states (Figure S1A). We first compared overall frequencies of non-silent exonic mutations between clinically defined subsets of HCC, including minimally invasive primary tumors (MIPs), widely invasive primary tumors (WIPs), and DM/LRs, as well as to The Cancer Genome Atlas study of PTC (consisting of primary tumors) (Cancer Genome Atlas Research Network, 2014). The non-silent mutation frequency in DM/LRs was higher than that observed in MIPs (DM/LR median, 0.7/Mb; range, 0.1–6.97 versus MIPs median, 0.4/Mb; range, 0.03–0.9, Wilcoxon test, p = 0.0006; Figure S1B; Table S2). Compared with PTC, WIPs, but not MIPs, exhibited statistically higher non-silent mutation frequencies (WIPs median, 0.6/Mb; range, 0.2–16.13 versus PTC median, 0.4/Mb; range, 0.03–2.17; Wilcoxon test, p = 0.025).

To identify significantly mutated genes, we integrated mutation data, including single nucleotide variations (SNVs) and short insertions and deletions (indels), from all 70 samples. To avoid double-counting mutations observed in different samples from the same patient, we created one "aggregated tumor" for each patient, which represented the union of mutations observed in all samples obtained from each patient. Multiple occurrences of the same event were counted once in order to catalog only independently arising mutations (the pre-aggregated mutation data are shown in Figure S1C). We found *DAXX*, *TP53*, *NRAS*, and *NF1* (Figure 1A, upper) to be significantly mutated using MutSig2CV (Lawrence et al., 2013). We also performed a restricted recurrence analysis using 326 previously published cancer genes (Forbes et al., 2017; Lawrence et al., 2014), and identified two additional recurrently mutated genes, *CDKN1A* and *ARHGAP35* (gene names depicted in green). Finally, we manually reviewed the data and identified additional mutations in known oncogenes and tumor suppressors (TS), including those previously identified in thyroid cancer (Figure 1A, middle) (Cancer Genome Atlas Research Network, 2014). Mutations in *TSHR*, *BRAF*, *EIF1AX*, *MEN1*, *PTEN*, *MTOR*, *ATM*, *MSH6*, and *NFE2L2* were also identified. Based on the cancer cell fraction (CCF) of each mutation, calculated using ABSOLUTE, most putative driver mutations were predicted to be clonal events (CCF $\geq$ 0.9), i.e., existing in all tumor cells in the sample, supporting an early origin during tumor outgrowth (median CCF in individual tumors of 0.99; Figure S1C) (Carter et al., 2012; Landau et al., 2013).

We also used targeted sequencing approaches to identify somatic alterations not detected by exome sequencing, including *TERT* promoter variants and fusion transcripts relevant to thyroid cancer (Landa et al., 2013; McFadden et al., 2014a). Hotspot *TERT* promoter mutations (C228T and C250T) were identified in 18/49 tumors from 12/38 (32%) patients from which sufficient nucleic acid was obtained (Figure 1A, lower; Table S3). Focal amplifications of the *TERT* gene were also observed in three tumors from two patients (Figures 1A, lower and S2, see below).
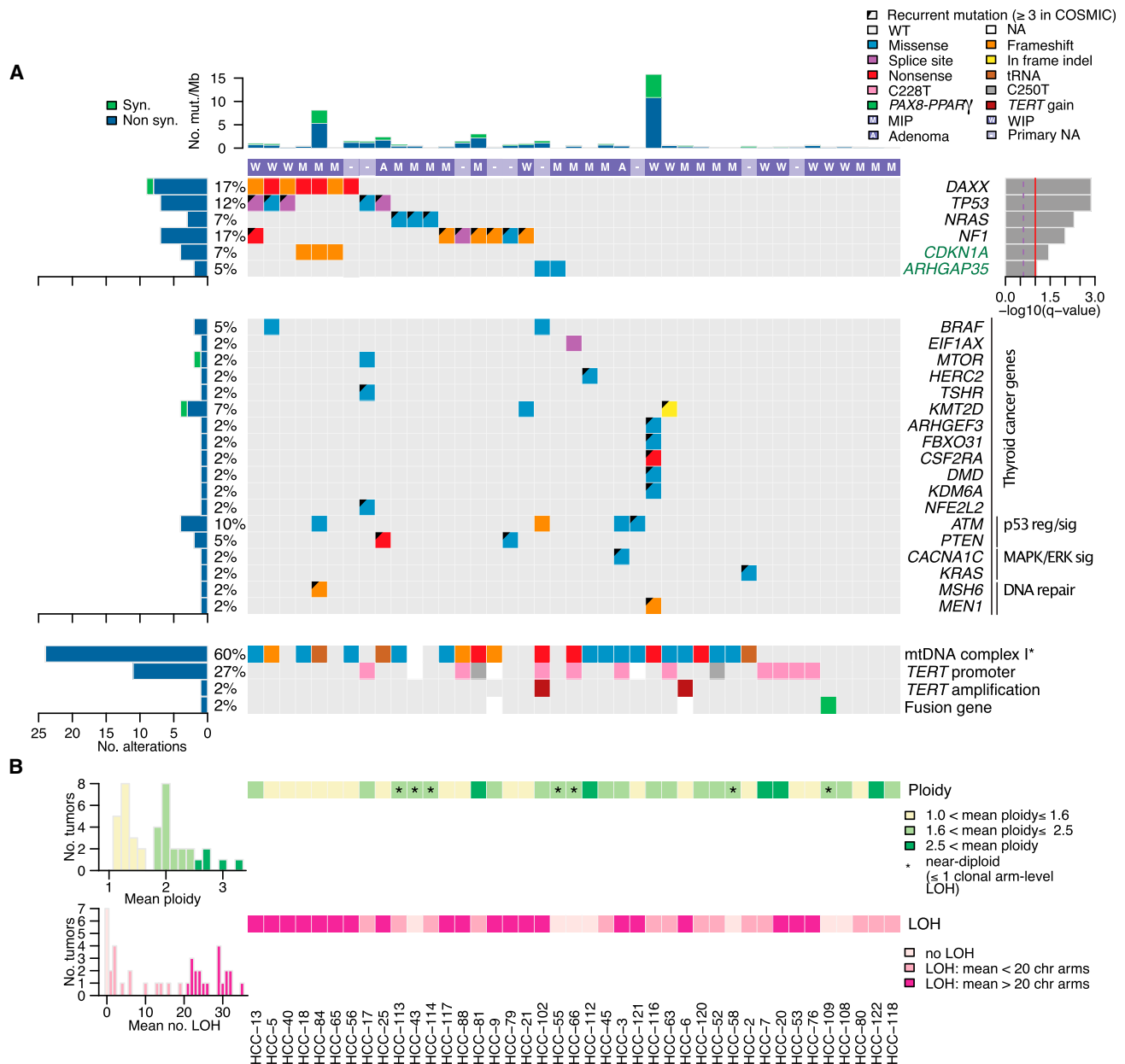
Figure 1. The Genomic Landscape of HCC

(A) Top: mutation status for significantly mutated nuclear genes using unrestricted hypothesis testing across the cohort (black font) and restricted hypothesis testing on known cancer genes (green font) are shown. Middle: genes previously associated with thyroid cancer, in which at least one non-silent mutation that appeared ≥ 3 times in the COSMIC database ("recurrent mutation"), are depicted. Bottom: mtDNA alterations, TERT promoter mutations, TERT amplifications, and fusion genes are identified. Mitochondrial alterations shown are those likely disrupting complex I, including missense mutations. The * highlights the inclusion of three leucyl-tRNA variants associated with the disease mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (brown boxes). The detection of a single fusion gene (PAX8-PPARγ) is depicted. The white boxes denote patients with insufficient DNA coverage for analysis.

(B) Top: mean tumor ploidy is shown, based on ABSOLUTE-generated ploidy values across all tumors from each patient. Histogram at left shows the distribution of mean ploidy values colored according to thresholds shown at right. Bottom: mean number of chromosomal clonal arm-level LOH events is shown (based on ABSOLUTE and averaged across all samples from each patient). Histogram at left shows distribution of mean LOH values, colored according to thresholds shown at the right. Each column represents one aggregated tumor (the union of mutations identified in all tumors from each patient, see text). In each subpanel colored boxes indicate the presence of a genomic alteration. Percentages indicate the fraction of aggregated tumors with an alteration in a gene; histograms indicate the number of alterations in a gene across the cohort of aggregated tumors.

See also Figure S1 and Tables S1, S2, and S3.

TERT alterations, including promoter mutations and amplifications, were mutually exclusive with *DAXX* mutations (Fisher's exact test on patient mutational level; p = 0.018), suggesting multiple mechanisms of stabilizing telomere length acted in HCC (Heaphy et al., 2011; Maciejowski and de Lange, 2017).

One *PAX8-PPARγ* fusion was identified in a tumor lacking alterations in other well-established driver genes (Figure 1A, lower). Out of 29 independent nuclear mutations identified by exome sequencing and selected for validation, 24 (83%) were confirmed by VariantPlex sequencing (Table S3, STAR Methods). Of 5 mutations failing this validation experiment, 4 were not adequately covered by validation sequencing reads; therefore, 24/25 (96%) adequately covered mutations were validated, suggesting high-fidelity mutation calling by our WES pipeline.

Early Widespread Chromosomal Loss Leads to a Near-Haploid State in HCC

Analysis of somatic DNA copy-number alterations (SCNAs), including determination of absolute allelic copy numbers (using ABSOLUTE), revealed widespread chromosomal losses leading to LOH across most of the genome and a near-haploid chromosomal content in the majority of tumors (38/70, 54%; Figures 1B and 2). Retained disomies of chr 5, chr 7, chr 12, and chr 20 suggested selection against the loss of these chromosomes. In contrast, few focal SCNAs were observed (Figure S2). We reviewed focally amplified and deleted regions in individual tumors for known oncogenes and tumor suppressor genes previously reported to be significantly altered in a pan-cancer analysis, and identified infrequent focal SCNAs in these genes (Table S4) (Zack et al., 2013)

We analyzed the distribution of ploidies observed in HCC, and identified three patterns in aggregated tumors, following a classification previously described for adrenocortical carcinoma: (1) "near-haploid" with a mean ploidy ≤1.6 (17/41 patients, 41.46%); (2) "quiet" with a mean ploidy of >1.6 to ≤2.5 (19/41, 46.34%), including 7 near-diploid cases; (3) "complex" with a mean ploidy >2.5 (5/41, 12.2%) (Figure 1B, top), in some cases due to whole-genome duplication (WGD) (Carter et al., 2012; Zheng et al., 2016). Of note, four aggregated tumors (HCC-3, HCC-13, HCC-102, and HCC-9) exhibited a mean ploidy >1.6 to ≤2.5 due to the occurrence of both widespread LOH (>20 arms) and WGD. WGD occurred in 9/41 (22%) patients in our cohort with 5 of these patients showing evidence of widespread LOH (>20 arms; Figures 1B, bottom and 2).

We experimentally confirmed ploidy predictions from ABSOLUTE using two complementary methods (Figures 3, S3A, and S3B). First, we performed fluorescence *in situ* hybridization (FISH) in 23 tumors, using probes for chromosomes 4, 7, and 8. FISH analysis was largely concordant with ABSOLUTE and AllelicCapSeg, but also revealed intratumoral heterogeneity in tumor populations exhibiting complex ploidy (Figures 3, middle column and S3A). Second, we analyzed the nuclear DNA (nDNA) content of individual tumor nuclei isolated from paraffin-embedded sections using imaging flow cytometry (IFC) of propidium iodide-stained nuclei. IFC enabled image collection of each event in both bright-field and fluorescent channels in order to confirm that modal peaks consisted of largely whole, intact nuclei. These studies confirmed approxi-

mately 2N nDNA content in diploid tumors, whereas near-haploid samples exhibited reduced (∼1N) nDNA content and complex samples showed increased (>2N) nDNA content (Figures 3, right column and S3B).

Genome wide loss of one copy of most chromosomes in HCC might endow a bias toward tumor suppressor mutations by enabling complete gene inactivation as the result of a single mutational event. Consistent with this possibility, mutations in the tumor suppressor genes *DAXX*, *TP53*, *NF1*, *CDKN1A*, *ATM*, *PTEN*, and *MSH6* were enriched in tumors with widespread LOH (>20 arms), compared with those without such chromosomal loss events (<20 arms; Fisher's exact test; p = 0.0001), resulting in biallelic inactivation of these genes (Figure S1C).

Complex I mtDNA Mutations as Driver Events in HCC

We next identified somatic mtDNA-encoded alterations from the exome sequencing data (STAR Methods). Coverage of the mtDNA was uniform and of high depth in control (83X) and tumor (314X) samples, with only one patient (HCC-43) excluded due to insufficient coverage (Figure S4A). Large deletions, including the common 4,977 bp deletion, were not detected (Schon et al., 1989). The overall mutational spectrum of the HCC mtDNA was similar to a recent International Cancer Genome Consortium (ICGC) pan-cancer analysis of mtDNA from 1,664 tumors across 31 cancers, including a strong strand bias for C > T and T > C transitions on the heavy and light strand, respectively (Figure S4B) (Ju et al., 2014).

Analysis of the mitochondrial genome revealed 29 protein-coding and 17 non-coding mutations with variant allele fraction (VAF) ≥ 0.3 (Table S5). Recurrent somatic loss-of-function (LOF) and missense mtDNA mutations were identified in genes encoding subunits of complex I of the electron transport chain in 21/40 patients (Figures 1A, lower and 4A). A leucyl-tRNA somatic variant MT-TL1:m.G3244A, which was previously reported to be pathogenic in an inherited case of mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes, a syndrome associated with complex I deficiency, was recurrently mutated in 3/40 patients (Kirino et al., 2005; Mariotti et al., 1995). This mutation was absent from the 1,664 tumors in a pan-cancer analysis (Ju et al., 2014). In total, 24/40 (60%) aggregated tumors harbored alterations predicted to impact complex I function. Complex I mtDNA variants were validated in 10/10 samples selected for PCR and Sanger sequencing (Table S5). Given that complex I is encoded by mitochondrial and nuclear genomes, we also searched for somatic alterations in the nuclear genome that were likely to disrupt complex I but did not identify recurrent mutations with a high likelihood of being pathogenic.

Several lines of evidence suggested a driver function of mitochondrial-encoded complex I mutations. First, complex I mutations were more prevalent in HCC compared with the ICGC pan-cancer analysis of 1,664 tumors from 31 cancer types (Ju et al., 2014). Specifically, while HCC exhibited a higher overall rate of mtDNA somatic mutations (Figure 4B), the increase was entirely driven by a 3-fold enrichment in protein-modifying mutations within complex I genes (p = 2 × 10^{-7} Fisher's exact test; Figure 4C). Conversely, no enrichment was observed for silent and non-coding variants, or alterations in complex III, IV, and V in HCC. Moreover, considering all protein-modifying variants, we noted a significant enrichment for mutations with
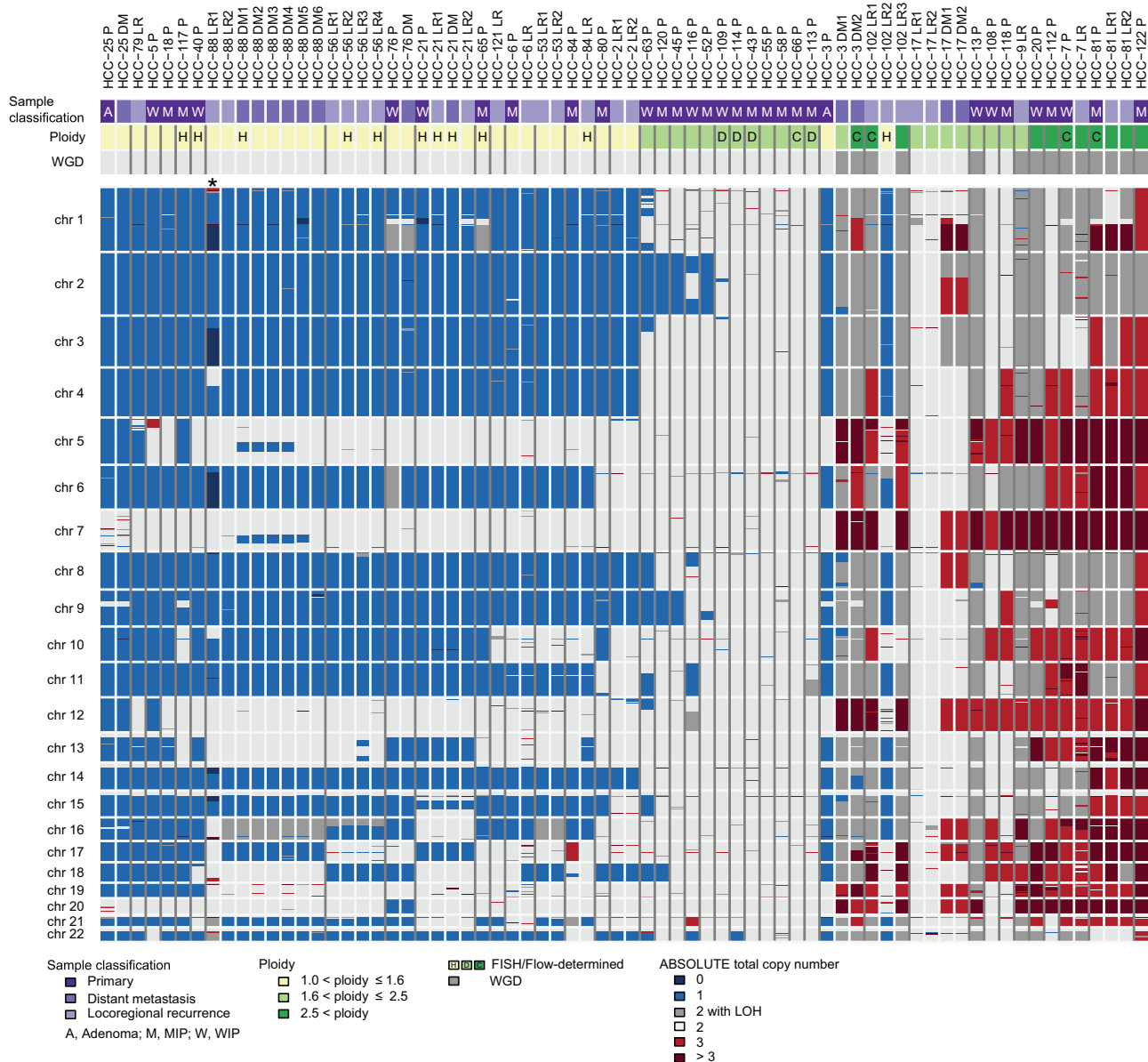
Figure 2. Widespread Loss of Chromosomes in HCC

ABSOLUTE total copy numbers of individual segments are delineated by their genomic position along the 22 chromosomes (top to bottom). Sample identifiers and classifications are shown at the top. Dark gray vertical lines separate different patients, and white vertical lines separate individual tumors from the same individual. Patients are ordered according to mean ploidy of aggregated tumors, and individual tumor samples from the same patient are ordered based on date of resection. The presence of WGD is indicated. Ploidy results of FISH and imaging flow cytometry (Flow) are indicated by letters for near-haploid (H), diploid (D), and complex (C) tumors. Colors indicate the loss, copy neutral LOH or gain at genomic loci. *HCC-88 LR1 analysis was confounded by low tumor purity. See also Figure S2 and Table S4.

VAF $\geq$ 0.3 in complex I genes in HCC compared with the expected background mutation frequency observed in the pan-cancer analysis (88% versus 60%, p = 0.007 based on two-sided test of proportions, Figure S4C). Second, in HCC but not in the pan-cancer analysis, protein-altering complex I, but not complex III, IV, or V, mutations exhibited higher VAF than silent/non-coding mutations, consistent with positive selection for complex I loss (Figures 4D and S4D). Specifically, in HCC complex I LOF and missense variants showed significantly

higher VAFs versus silent/non-coding variants (median VAF 0.98 and 0.78 versus 0.17, p = 10^{-3} and p = 9 × 10^{-3}, respectively, two-sided Wilcoxon rank-sum test), whereas the pan-cancer analysis showed significantly lower VAF for LOF variants (Figure 4D). Third, in HCC, complex I missense mutations occurred preferentially at highly conserved residues (Figure 4E), and mapped to regions within the highly charged central axis believed to be important for redox coupling and proton translocation based on the high-resolution crystal structure of complex
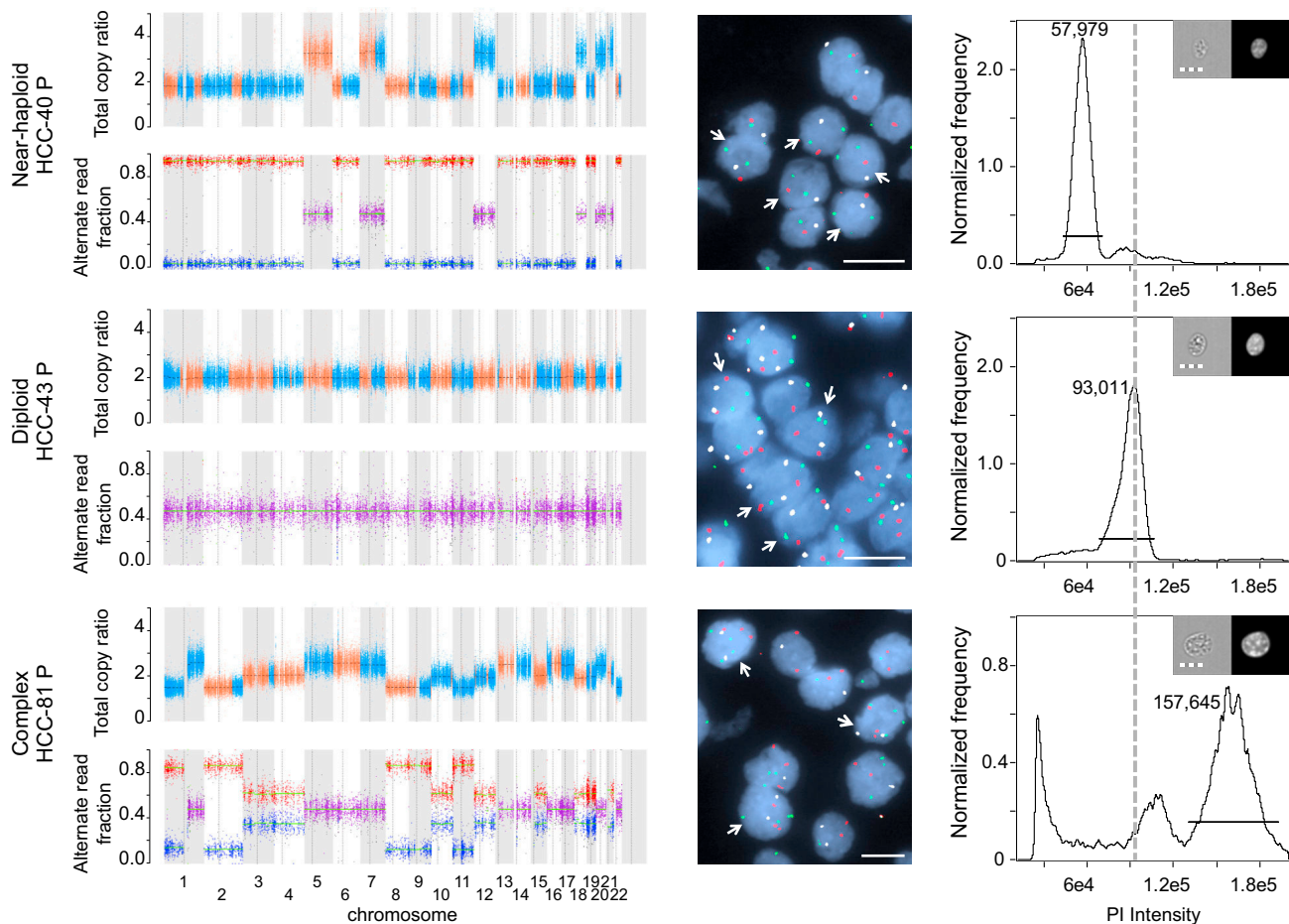
Figure 3. Experimental Validation of the Near-Haploid State

Left column: genomic plots show coverage-derived somatic copy-number ratios (upper panel, normalized to two by convention), and the allelic ratio of single-nucleotide variants across each chromosome (lower panel) for three representative tumor samples with near-haploid, diploid, and complex chromosomal content. Purple color represents heterozygous SNVs with an allelic ratio of ~0.5, whereas red represents higher allelic ratio and blue represents lower allelic ratio. Middle column: FISH images for chromosomes 4 (*PDGFRA*; red), 7 (*EGFR*; green), and 8 (centromeric; white) are shown. White arrows point to the probe signals for the modal cell population. Scale bars, 10 μm. Right column: frequency distribution histograms from imaging flow cytometry with propidium iodide (PI) labeling of nDNA are shown. Horizontal lines demarcate the gating for the modal cell population and the modal PI intensity of the parent population is shown within each plot. The vertical dashed gray line corresponds to the modal PI intensity in the diploid tumor. Bright-field and fluorescent PI images confirmed measurement of single cells and representative gated cells are shown in the top right. Scale bars, 10 μm. The same tumor is represented across each row. See also Figure S3.

I available from *Thermus thermophilus* (Figure 4F; Table S5) (Baradaran et al., 2013). Together, these findings suggest that mtDNA-encoded complex I alterations act as drivers in HCC.

Chromosomal Content Is Stable during Metastatic Progression of HCC

We next examined the evolution of the nuclear and mitochondrial genomes during disease progression. A near-haploid genome has been described as an unstable, unfavorable state, which is rescued by subsequent WGD (Corver et al., 2012; Taylor-Weiner et al., 2016; Zheng et al., 2016). This leads to tumors with a "pseudo-diploid," or uniparental disomy, chromosomal state (Safavi and Paulsson, 2017). In contrast to these studies, we observed cases in which related families of primary tumors, local recurrences, and DM harbored near-haploid nDNA content (9/14, 64%; Figure 5, top). WGD was observed in a minority of

HCC tumor families (4/14, 29%; Figure 5, bottom). The observation that the near-haploid state was maintained during metastatic seeding and recurrence of HCC suggested that subsequent genome doubling was not under positive selection during metastatic colonization or emergence, or that genome doubling is mechanically infrequent in these tumors.

Mutational Evolution during Metastatic Progression of HCC

Evolutionary relationships across multiple tumors resected from individual patients were reconstructed using PhylogicNDT with *TERT* promoter and mtDNA complex I mutations manually mapped onto phylogenetic trees. All 14 patients with recurrent and/or metastatic disease in our cohort showed a branched pattern of tumor evolution suggestive of a common ancestor in all cases (Figures 6A and S5A). We first identified shared mutations across
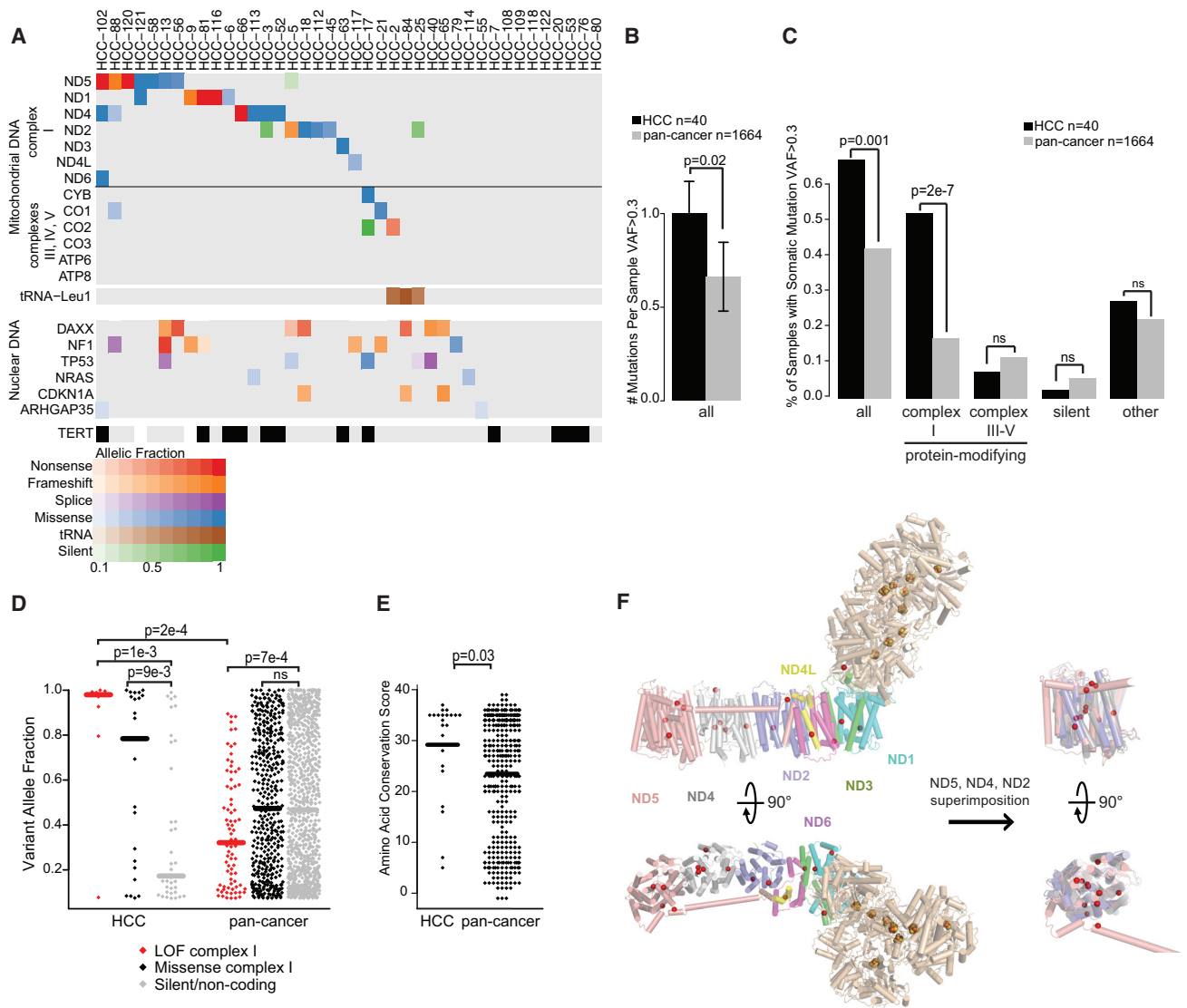
Figure 4. mtDNA Mutations in HCC

(A) Mutations in mtDNA genes (VAF ≥ 0.3) are shown, along with selected nuclear gene mutations. *TERT* events (shown in black) include promoter variants and gene amplifications. Each column represents one aggregated tumor.

(B) Barplot shows the mean number of somatic mutations in HCC samples and the pan-cancer analysis (Ju et al., 2014). Error bars show SE; heteroscedastic two-sided t test.

(C) Barplots show the percent of HCC samples containing different subsets of somatic mutations; Fisher's exact test.

(D and E) Comparison of variant allele frequency between somatic mutations (D) and amino acid conservation scores for sites with somatic missense mtDNA mutations (E) in HCC and the pan-cancer analysis. Horizontal lines show the median; two-sided Wilcoxon rank-sum test.

(F) Location of missense mutations observed in HCC within the complex I crystal structure from *Thermus thermophilus* (Baradaran et al., 2013). mtDNA-encoded genes (ND1-ND6) shown in colors, mutations shown as red spheres, Fe-S clusters as brown/golden spheres. Structurally similar proteins (ND2, ND4, and ND5) are shown superimposed to show mutation clustering.

(B)–(E) Data from one index sample per patient, rather than aggregated tumors (A). ns, not significant (p > 0.05). See also Figure S4 and Table S5.

tumor families obtained from each individual patient, as these were likely to represent early driver alterations. Widespread chromosomal losses were the most frequently observed truncal event, occurring in 13/14 tumor families (Figure 6B). WGD occurred in only 4/14 patients: two truncal events (HCC-81 and -7), one late event (HCC-102), and one convergent branch event (HCC-3), reiterating the stability of the near-haploid state during tumor evolution. Truncal complex I mtDNA mutation, present in

6/14 patients, was the next most frequent early alteration. *TERT* promoter alterations were observed in 8/14 tumor families, including 2/14 with clear evidence of truncal emergence (HCC-3 and -76) and 4/14 families with presumed truncal emergence based upon the presence of the event in all samples with adequate sequencing coverage. Among the putative driver mutations identified in this study, truncal alterations in *NF1* (2/14), *DAXX* (1/14), *PTEN* (1/14), and *ATM* (1/14) were also observed.
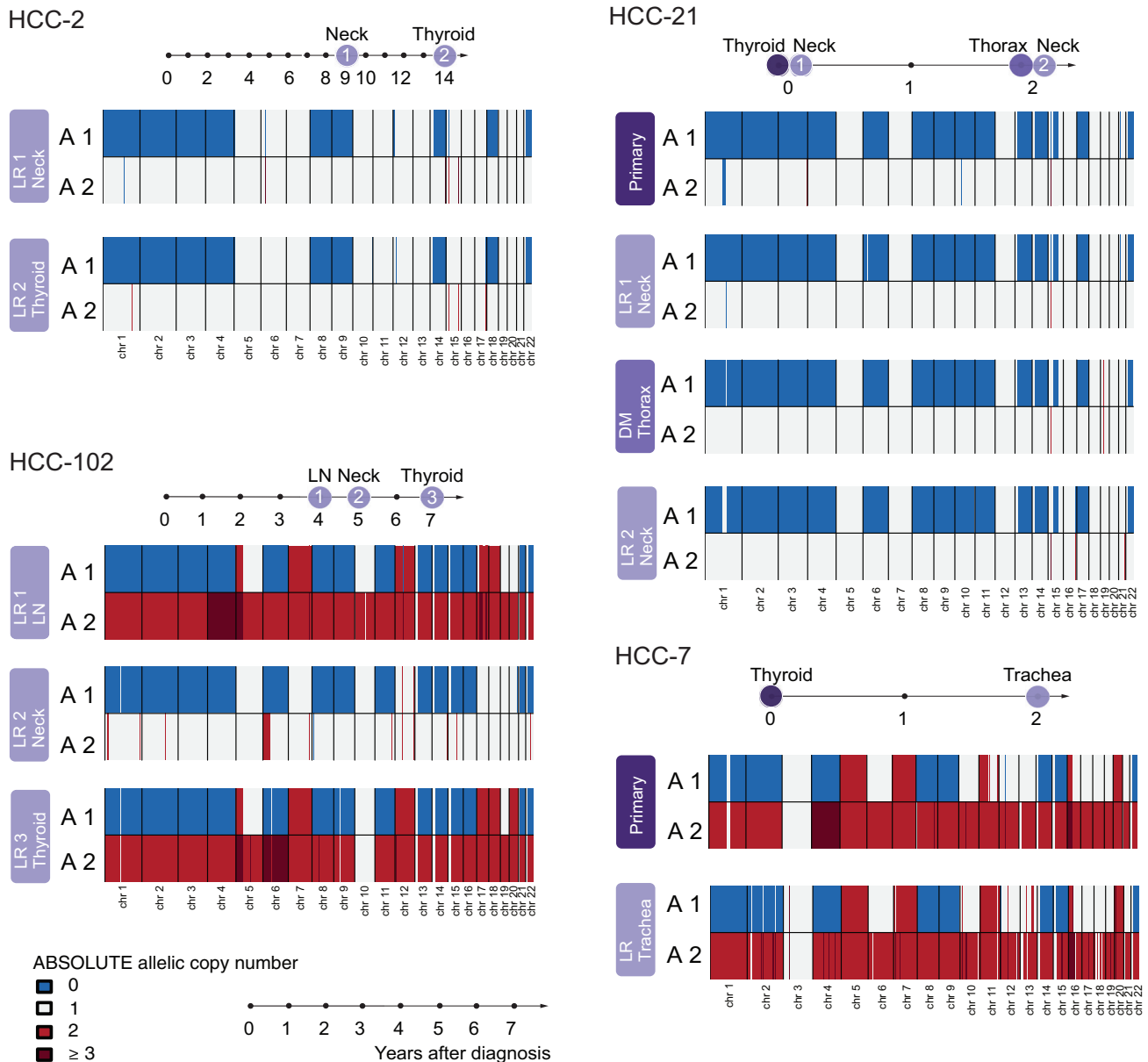
Figure 5. The Near-Haploid Genotype Is Stable during Metastatic Progression

ABSOLUTE-derived copy numbers for allele 1 (A1) (low copy numbers) and allele 2 (A2) (normal or high copy numbers) are shown for all tumors obtained from each patient. Timeline of resection for each patient is shown at the top. Chromosomal segments are shown horizontally by genomic position for each tumor sample, with color indicating the copy number. Blue indicates 0N, red indicates 2N, gray represents 1N, and dark red represents ≥3N (the color legend at the bottom applies to all panels). Tumor families depicted are representative of a stable near-haploid genotype (top) and complex genotypes with WGD (bottom).

We identified only one patient (HCC-17) with a classical pattern of tumorigenesis in which an initiating oncogene mutation (in this case, *TSHR*) was followed by inactivation of a tumor suppressor gene (*TP53*) and emergence of nDNA copy-number alterations (Figure S5A).

We also observed evidence for convergent evolution of putative driver events with independent mutations arising in distinct evolutionary branches in 5/14 patients. Independent alterations in *DAXX* and *CDKN1A* (HCC-84), *TP53* (HCC-25), *TERT* (HCC-81 and 102), mtDNA complex I genes (HCC-102),

and WGD (HCC-3) were observed (Figures 6A and S5A). As an example, we found one patient (HCC-84) harboring different *DAXX* and *CDKN1A* mutations in the primary and the recurrent tumor, consistent with convergent evolution and supporting a non-clonal origin of these events. In another patient (HCC-81), the primary and recurrent tumors harbored different *TERT* promoter mutations. The convergent evolution of mutations in these genes provided strong evidence of a driver role for these alterations, including at later stages of tumor evolution.
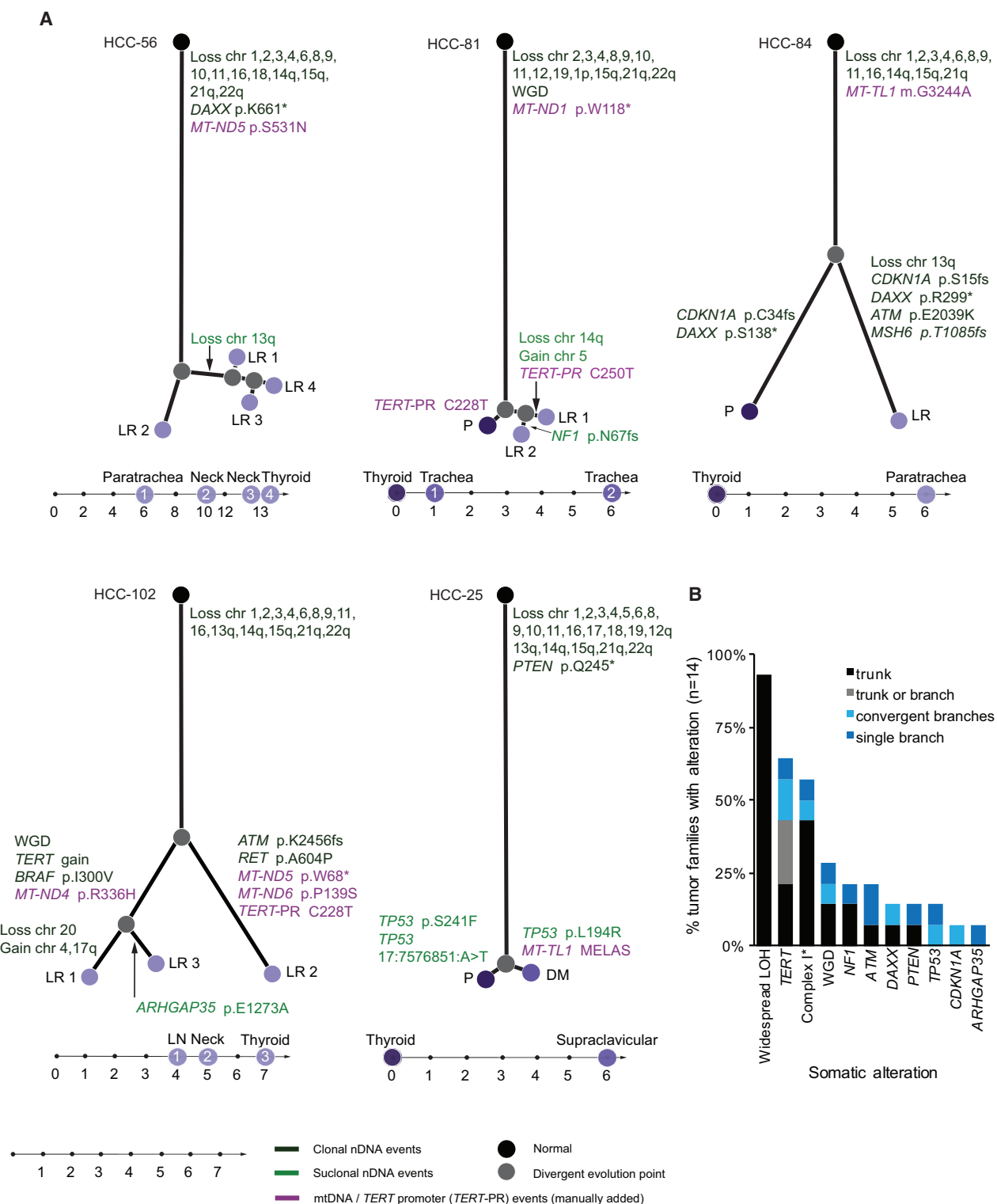
Figure 6. The Phylogeny of HCC Tumor Families

(A) Phylogenetic trees show relationships between distinct tumor samples obtained from each patient. The length of the phylogenetic branches is based on the Jaccard distance between clones taking into account clonal nDNA events (CCF ≥ 0.9). Clonal (black) and subclonal (green) gains, losses and oncogenic mutations are annotated onto the branches on which they occurred. High VAF (≥0.3) complex I mtDNA mutations and *TERT* promoter mutations were manually added to the trees (purple). The top black dot represents the germline (normal) genome. The initial vertical line from the normal genome to the first divergence of the lines depicts events shared by all samples; lines between nodes depict events shared among a few; purple dots depict events unique to the individual tumors

Although genomic analysis of related families of primary tumors and recurrent/metastatic lesions was prioritized over analysis of large numbers of primary tumors, we determined associations between primary tumor genotype and two important clinical outcomes, recurrence-free survival (RFS) (which included local recurrences and DM), and the time to development of DM. Although RFS was not statistically different between MIP and WIP cases (Figure S5B), the development of DM was significantly increased in WIP cases (Figure S5C), consistent with prior studies (Ghossein et al., 2006). A larger fraction of WIP tumors exhibited widespread LOH (8/11) compared with MIP (7/19); however, this was not statistically significant (Fisher's exact test, nominal p = 0.13, Figure S5D). *TP53* mutation was only observed in WIP tumors (3/11, Fisher's exact test, nominal p = 0.04). No other event was significantly enriched in MIP or WIP tumors. In contrast, distant metastatic disease was significantly associated with primary tumors harboring widespread LOH (Fisher's exact test, nominal $p = 2 \times 10^{-4}$, Figure S5E), and *TERT* alteration (promoter mutation or amplification, Fisher's exact test, nominal p = 0.03). We therefore further examined potential associations between widespread LOH and clinical outcomes. Widespread LOH significantly associated with decreased RFS (Figure S5F, Mantel-Cox test, p = 0.003) and the development of distant metastasis during the observation period (Figure S5G, Mantel-Cox test, p = 0.001). Widespread LOH was the most significant predictor of poor outcome in this cohort, being a stronger predictor of recurrence and distant metastasis than primary tumor invasiveness (compare Figures S5B–S5F and S5C–S5G).

DISCUSSION

This parallel analysis of the nuclear and mitochondrial genomes of HCC primary tumors, local recurrences, and metastases reveals three primary findings: (1) the near-haploid state is common in HCC, occurs early, and is stable throughout the course of the disease; (2) complex I mtDNA mutations are early clonal events maintained during tumor evolution, and therefore likely drivers and not simply markers; and (3) a distinct set of nDNA driver genes is expanded to include *DAXX*, *CDKN1A*, and *ARHGAP35*.

These alterations appear uniquely enriched in HCC compared with other cancers. We show that mtDNA complex I mutations are enriched in HCC compared with a pan-cancer analysis (Ju et al., 2014). In addition, widespread LOH has not been observed in other forms of thyroid cancer, including conventional FTC, of which HCC was considered a subtype (Jung et al., 2016). Similarly, a clinical sequencing study of poorly differentiated and anaplastic thyroid carcinoma, in which recurrent widespread LOH was not identified, further suggests widespread chromosomal losses are not simply a feature of progression to advanced disease in thyroid carcinoma (Landa et al., 2016). A small number of HCC cases

harbored alterations in oncogenes and TS associated with FTC and PTC, including *BRAF*, *NRAS*, *TSHR*, *EIF1AX*, and *PAX8-PPAR*γ, which are considered to be tumor initiating alterations. Therefore, these alterations also appear capable of driving tumorigenesis in the context of Hürthle cells. Alterations in *TP53* and amplifications or promoter mutations impacting *TERT* expression have been identified in clinically aggressive forms of thyroid cancer (Landa et al., 2016). The observance of these events in HCC suggests that the mechanisms driving the development of advanced disease in HCC overlap with other forms of thyroid carcinoma.

This study raises several questions—what are the functional effect(s) of mutations in complex I and widespread chromosomal losses, and how do these events confer fitness advantage? Disruptive mutations in complex I could in theory lead to the production of a metabolite with oncogenic function, similar to tumors with *IDH1* and other tricarboxylic acid cycle mutations (Sabharwal and Schumacker, 2014). Alternatively, mutations in complex I could protect from cell death, since complex I has been implicated in interferon- and p53-induced cell death, or facilitate the alternative use of metabolic intermediates to support anabolic processes required in HCC for proliferation (Compton et al., 2011; Huang et al., 2007; Shaham et al., 2010). The observation of complex I alterations in the two adenomas included in this study suggests that additional events are necessary to promote tumor progression. Whether complex I mutations and the near-haploid state co-evolve during tumor progression or arise from independent selective pressures remains an open question. Interestingly, haploid human embryonic stem cells exhibit a more than 30% increase in the mtDNA to nDNA ratio when compared with diploid cells, suggesting potential synchronization between nuclear and mitochondrial genomes (Sagi et al., 2016).

Widespread loss of chromosomes leading to the near-haploid state, an extensive, ordered form of aneuploidy distinct from most cancer genomes, was the most frequently observed early event by phylogenetic analysis (Beroukhim et al., 2010; Zack et al., 2013). It is therefore tempting to speculate that the near-haploid state arises early and unmasks a selective advantage for the metabolic consequences of complex I mutation or cooperates with additional mutations in genes encoding canonical TS. The latter hypothesis is congruent with the observation that mutations in TS genes that were recurrently altered in HCC were more frequently observed in near-haploid cases. In the near-haploid genome, complete TS gene inactivation requires merely a single hit. However, given the absence of widespread chromosomal losses in most cancer genomes, massive LOH is unlikely to be a prerequisite for biallelic inactivation of TS genes in HCC (Beroukhim et al., 2010). Therefore, enrichment for TS gene mutations in near-haploid HCC tumors might simply reflect the potential for single events to result in complete TS inactivation in this sensitized background. Alternatively, the near-haploid state may lead to gene dosage imbalances,

---

that are noted at the end. Below each tree, a timeline depicts the sequence of diagnosis and tissue sampling. P, primary tumor; DM, distant metastasis; LR, locoregional recurrence; LN, lymph node.

(B) Barplot shows a summary of genomic alterations observed in HCC evolution, based on 14 patients from whom multiple tumors were analyzed. Complex I * indicates the inclusion of variants in complex I genes and a tRNA variant associated with MELAS.

See also Figure S5.

or cumulative haploinsufficiencies (Davoli et al., 2013; Solimini et al., 2012). The strong association of widespread chromosomal loss with recurrence and metastatic disease is intriguing. However, whether widespread chromosomal loss represents an independent predictor of outcome will require additional studies considering the limitations of the current study, including retrospective design, potential for selection bias, duration of clinical follow-up, and limited sample size.

Our phylogenetic analyses suggest that widespread chromosomal losses generally appeared prior to other events, supporting the notion that selective pressures operative at early stages in HCC initiation and outgrowth drive the emergence of the near-haploid state. Although the driving force underlying widespread chromosomal loss is unclear, it is notable that other tumors prone to chromosomal losses, such as adrenal cortical carcinoma (ACC) and chromophobe kidney cancer, also acquire recurrent alterations in *DAXX*, *TP53*, and the *TERT* promoter (Assie et al., 2014; Davis et al., 2014; Zheng et al., 2016). In our cohort, *DAXX* and *TERT* promoter mutations were mutually exclusive, and *DAXX* mutations were more frequently observed in near-haploid tumors. This suggests that HCC employs multiple strategies of telomere maintenance and raises the possibility that ALT is favored in the context of a near-haploid genome. These findings reinforce the association between telomere stabilization and genome instability (Chiba et al., 2017; Fleisig et al., 2016; Heaphy et al., 2011; Maciejowski and de Lange, 2017; Marinoni et al., 2014).

The stability of the near-haploid state during clonal evolution of HCC is distinguished from ACC in which WGD was observed following widespread nDNA loss (Zheng et al., 2016). Furthermore, unlike colorectal and bladder cancers, in which nDNA copy-number alterations were discordant between primary tumors and metastases (Bambury et al., 2015; Mamlouk et al., 2017), the near-haploid state of HCC was maintained during metastatic outgrowth, in some cases over several years.

Our work provides a molecular roadmap for the identification of targetable genomic and metabolic vulnerabilities in HCC. Studies are warranted to determine whether the near-haploid state leads to a liability due to collateral copy loss of essential genes—a genetic state that might be leveraged by agents targeting such copy-number dependencies (Muller et al., 2012; Nijhawan et al., 2012). Whether this concept underlies occasional responses of HCC to cytotoxic chemotherapeutics remains an open question (Besic et al., 2012; Zhang et al., 2012). The development of therapies that target the metabolic consequences of complex I impairment or the potential dependency of HCC genomes on telomere-stabilizing mechanisms (via TERT or ALT pathways) also represent other directions of future research (Birsoy et al., 2014; Shaham et al., 2010; Vafai et al., 2016). Finally, the development of new model systems of HCC, upon which to interrogate hypotheses developed from the framework of alterations described here, represents an important future endeavor.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Human Tissue Studies
- METHOD DETAILS
  - ○ Nuclear Genome Analysis
  - ○ Somatic Copy Number Analysis
  - ○ Purity, Ploidy, and Phylogeny
  - ○ Validation of Nuclear Variants
  - ○ Detection of Gene Rearrangements
  - ○ mtDNA Variant Calling and Validation
  - ○ mtDNA Pan-Cancer Comparative Analysis
  - ○ Complex I Structural Analysis
  - ○ Fluorescence In-Situ Hybridization (FISH)
  - ○ Imaging Flow Cytometry and Gating Strategy
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Nuclear Mutation Significance Analysis
  - ○ mtDNA Mutation Significance Analysis
  - ○ Quantification of FISH Results
  - ○ Kaplan-Meier Analysis
- DATA AVAILABILITY

AUTHOR CONTRIBUTIONS

D.G.M. and G.H.D. conceived the study. R.K.G., K.K., D.D.-S., G.G., and D.G.M. designed the overall analysis and experiments and wrote the manuscript. R.K.G. and S.E.C. analyzed the mtDNA data. K.K. analyzed the nuclear data. P.P., P.F.A., C.S., L.Z.B., J.K., and J.M.H. contributed to data analysis. D.L., D.R., and I.L. performed the phylogenetic analysis. P.M.S., A.R.S., and F.C. reviewed the histopathology. D.D.-S. and A.J.I. performed the FISH analysis. Z.G. analyzed the complex I crystal structure. S.M. performed the IFC analysis. S.A., B.J.G., B.C., S.E.D., C.C.L., T.Z., L.W., G.H.D., D.G.M., and S.P. recruited patients, acquired samples, and analyzed clinical data. V.K.M., G.G., D.D.-S., and D.G.M. supervised the study. R.K.G. and K.K. contributed equally to this manuscript, and are listed in alphabetical order.

DECLARATION OF INTERESTS

A.J.I. reports ownership interests and has received honoraria from ArcherDx.

REFERENCES

Assie, G., Letouze, E., Fassnacht, M., Jouinot, A., Luscap, W., Barreau, O., Omeiri, H., Rodriguez, S., Perlemoine, K., Rene-Corail, F., et al. (2014). Integrated genomic characterization of adrenocortical carcinoma. Nat. Genet. *46*, 607–612.

Bambury, R.M., Bhatt, A.S., Riester, M., Pedamallu, C.S., Duke, F., Bellmunt, J., Stack, E.C., Werner, L., Park, R., Iyer, G., et al. (2015). DNA copy number analysis of metastatic urothelial carcinoma with comparison to primary tumors. BMC Cancer *15*, 242.

Baradaran, R., Berrisford, J.M., Minhas, G.S., and Sazanov, L.A. (2013). Crystal structure of the entire respiratory complex I. Nature *494*, 443–448.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Besic, N., Auersperg, M., Gazic, B., Dremelj, M., and Zagar, I. (2012). Neoadjuvant chemotherapy in 29 patients with locally advanced follicular or Hurthle cell thyroid carcinoma: a phase 2 study. Thyroid *22*, 131–137.

Birsoy, K., Possemato, R., Lorbeer, F.K., Bayraktar, E.C., Thiru, P., Yucel, B., Wang, T., Chen, W.W., Clish, C.B., and Sabatini, D.M. (2014). Metabolic determinants of cancer cell sensitivity to glucose limitation and biguanides. Nature *508*, 108–112.

Brandon, M.C., Lott, M.T., Nguyen, K.C., Spolim, S., Navathe, S.B., Baldi, P., and Wallace, D.C. (2005). MITOMAP: a human mitochondrial genome database–2004 update. Nucleic Acids Res. *33*, D611–D613.

Brastianos, P.K., Carter, S.L., Santagata, S., Cahill, D.P., Taylor-Weiner, A., Jones, R.T., Van Allen, E.M., Lawrence, M.S., Horowitz, P.M., Cibulskis, K., et al. (2015). Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. Cancer Discov. *5*, 1164–1177.

Brastianos, P.K., Nayyar, N., Rosebrock, D., Leshchiner, I., Gill, C.M., Livitz, D., Bertalan, M.S., D'Andrea, M., Hoang, K., Aquilanti, E., et al. (2017). Resolving the phylogenetic origin of glioblastoma via multifocal genomic analysis of pre-treatment and treatment-resistant autopsy specimens. NPJ Precis Oncol. *1*, 33.

Calvo, S., Jain, M., Xie, X., Sheth, S.A., Chang, B., Goldberger, O.A., Spinazzola, A., Zeviani, M., Carr, S.A., and Mootha, V.K. (2006). Systematic identification of human mitochondrial disease genes through integrative genomics. Nat. Genet. *38*, 576–582.

Cancer Genome Atlas Research Network (2014). Integrated genomic characterization of papillary thyroid carcinoma. Cell *159*, 676–690.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421.

Chiba, K., Lorbeer, F.K., Shain, A.H., McSwiggen, D.T., Schruf, E., Oh, A., Ryu, J., Darzacq, X., Bastian, B.C., and Hockemeyer, D. (2017). Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism. Science *357*, 1416–1420.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Cibulskis, K., McKenna, A., Fennell, T., Banks, E., DePristo, M., and Getz, G. (2011). ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics *27*, 2601–2602.

Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422–1423.

Compton, S., Kim, C., Griner, N.B., Potluri, P., Scheffler, I.E., Sen, S., Jerry, D.J., Schneider, S., and Yadava, N. (2011). Mitochondrial dysfunction impairs tumor suppressor p53 expression/function. J. Biol. Chem. *286*, 20297–20312.

Corver, W.E., Ruano, D., Weijers, K., den Hartog, W.C., van Nieuwenhuizen, M.P., de Miranda, N., van Eijk, R., Middeldorp, A., et al. (2012). Genome haploidisation with chromosome 7 retention in oncocytic follicular thyroid carcinoma. PLoS One *7*, e38287.

Corver, W.E., van Wezel, T., Molenaar, K., Schrumpf, M., van den Akker, B., van Eijk, R., Ruano Neto, D., Oosting, J., and Morreau, H. (2014). Near-haploidization significantly associates with oncocytic adrenocortical, thyroid, and parathyroid tumors but not with mitochondrial DNA mutations. Genes Chromosomes Cancer *53*, 833–844.

Costello, M., Pugh, T.J., Fennell, T.J., Stewart, C., Lichtenstein, L., Meldrim, J.C., Fostel, J.L., Friedrich, D.C., Perrin, D., Dionne, D., et al. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. *41*, e67.

Davis, C.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D., Shen, H., Buhay, C., Kang, H., Kim, S.C., Fahey, C.C., et al. (2014). The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell *26*, 319–330.

Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell *155*, 948–962.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Escobar, M.D., and West, M. (1995). Bayesian density estimation and inference using mixtures. J. Am. Stat. Assoc. *90*, 577–588.

Evangelisti, C., de Biase, D., Kurelac, I., Ceccarelli, C., Prokisch, H., Meitinger, T., Caria, P., Vanni, R., Romeo, G., Tallini, G., et al. (2015). A mutation screening of oncogenes, tumor suppressor gene TP53 and nuclear encoded mitochondrial complex I genes in oncocytic thyroid tumors. BMC Cancer *15*, 157.

Fleisig, H.B., Hukezalie, K.R., Thompson, C.A., Au-Yeung, T.T., Ludlow, A.T., Zhao, C.R., and Wong, J.M. (2016). Telomerase reverse transcriptase expression protects transformed human cells against DNA-damaging agents, and increases tolerance to chromosomal instability. Oncogene *35*, 218–227.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. *45*, D777–D783.

Ganly, I., Ricarte Filho, J., Eng, S., Ghossein, R., Morris, L.G., Liang, Y., Socci, N., Kannan, K., Mo, Q., Fagin, J.A., and Chan, T.A. (2013). Genomic dissection of Hurthle cell carcinoma reveals a unique class of thyroid malignancy. J. Clin. Endocrinol. Metab. *98*, E962–E972.

Gasparre, G., Porcelli, A.M., Bonora, E., Pennisi, L.F., Toller, M., Iommarini, L., Ghelli, A., Moretti, M., Betts, C.M., Martinelli, G.N., et al. (2007). Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. Proc. Natl. Acad. Sci. U S A *104*, 9001–9006.

Gasparre, G., Bonora, E., Tallini, G., and Romeo, G. (2010). Molecular features of thyroid oncocytic tumors. Mol. Cell Endocrinol. *321*, 67–76.

Ghossein, R.A., Hiltzik, D.H., Carlson, D.L., Patel, S., Shaha, A., Shah, J.P., Tuttle, R.M., and Singh, B. (2006). Prognostic factors of recurrence in encapsulated Hurthle cell carcinoma of the thyroid gland: a clinicopathologic study of 50 cases. Cancer *106*, 1669–1676.

Haugen, B.R., Alexander, E.K., Bible, K.C., Doherty, G.M., Mandel, S.J., Nikiforov, Y.E., Pacini, F., Randolph, G.W., Sawka, A.M., Schlumberger, M., et al. (2016). 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid *26*, 1–133.

Heaphy, C.M., de Wilde, R.F., Jiao, Y., Klein, A.P., Edil, B.H., Shi, C., Bettegowda, C., Rodriguez, F.J., Eberhart, C.G., et al. (2011). Altered telomeres in tumors with ATRX and DAXX mutations. Science *333*, 425.

Huang, G., Chen, Y., Lu, H., and Cao, X. (2007). Coupling mitochondrial respiratory chain to cell death: an essential role of mitochondrial complex I in the interferon-beta and retinoic acid-induced cancer cell death. Cell Death Differ. *14*, 327–337.

Hyytinen, E., Visakorpi, T., Kallioniemi, A., Kallioniemi, O.P., and Isola, J.J. (1994). Improved technique for analysis of formalin-fixed, paraffin-embedded tumors by fluorescence in situ hybridization. Cytometry *16*, 93–99.

Ju, Y.S., Alexandrov, L.B., Gerstung, M., Martincorena, I., Nik-Zainal, S., Ramakrishna, M., Davies, H.R., Papaemmanuil, E., Gundem, G., Shlien, A., et al. (2014). Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. Elife *3*, https://doi.org/10.7554/eLife.02935.

Jung, S.H., Kim, M.S., Jung, C.K., Park, H.C., Kim, S.Y., Liu, J., Bae, J.S., Lee, S.H., Kim, T.M., Lee, S.H., and Chung, Y.J. (2016). Mutational burdens and evolutionary ages of thyroid follicular adenoma are comparable to those of follicular carcinoma. Oncotarget *7*, 69638–69648.

Kasaian, K., Chindris, A.M., Wiseman, S.M., Mungall, K.L., Zeng, T., Tse, K., Schein, J.E., Rivera, M., Necela, B.M., Kachergus, J.M., et al. (2015). MEN1 mutations in Hurthle cell (oncocytic) thyroid carcinoma. J. Clin. Endocrinol. Metab. *100*, E611–E615.

Kirino, Y., Goto, Y., Campos, Y., Arenas, J., and Suzuki, T. (2005). Specific correlation between the wobble modification deficiency in mutant tRNAs and the clinical features of a human mitochondrial disease. Proc. Natl. Acad. Sci. USA *102*, 7127–7132.

Landa, I., Ganly, I., Chan, T.A., Mitsutake, N., Matsuse, M., Ibrahimpasic, T., Ghossein, R.A., and Fagin, J.A. (2013). Frequent somatic TERT promoter mutations in thyroid cancer: higher prevalence in advanced forms of the disease. J. Clin. Endocrinol. Metab. *98*, E1562–E1566.

Landa, I., Ibrahimpasic, T., Boucai, L., Sinha, R., Knauf, J.A., Shah, R.H., Dogan, S., Ricarte-Filho, J.C., Krishnamoorthy, G.P., Xu, B., et al. (2016). Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. J. Clin. Invest. *126*, 1052–1066.

Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell *152*, 714–726.

Landau, D.A., Sun, C., Rosebrock, D., Herman, S.E.M., Fein, J., Sivina, M., Underbayev, C., Liu, D., Hoellenriegel, J., Ravichandran, S., et al. (2017). The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. Nat. Commun. *8*, 2185.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, M., Schroeder, R., Ko, A., and Stoneking, M. (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. Nucleic Acids Res. *40*, e137.

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476–482.

Maciejowski, J., and de Lange, T. (2017). Telomeres in cancer: tumour suppression and genome instability. Nat. Rev. Mol. Cell Biol. *18*, 175–186.

Mamlouk, S., Childs, L.H., Aust, D., Heim, D., Melching, F., Oliveira, C., Wolf, T., Durek, P., Schumacher, D., Blaker, H., et al. (2017). DNA copy number changes define spatial patterns of heterogeneity in colorectal cancer. Nat. Commun. *8*, 14093.

Marinoni, I., Kurrer, A.S., Vassella, E., Dettmer, M., Rudolph, T., Banz, V., Hunger, F., Pasquinelli, S., Speel, E.J., and Perren, A. (2014). Loss of DAXX and ATRX are associated with chromosome instability and reduced survival of patients with pancreatic neuroendocrine tumors. Gastroenterology *146*, 453–460.e5.

Mariotti, C., Savarese, N., Suomalainen, A., Rimoldi, M., Comi, G., Prelle, A., Antozzi, C., Servidei, S., Jarre, L., DiDonato, S., and Zeviani, M. (1995). Genotype to phenotype correlations in mitochondrial encephalomyopathies associated with the A3243G mutation of mitochondrial DNA. J. Neurol. *242*, 304–312.

Maximo, V., and Sobrinho-Simoes, M. (2000). Hurthle cell tumours of the thyroid. A review with emphasis on mitochondrial abnormalities with clinical relevance. Virchows Arch. *437*, 107–115.

McFadden, D.G., Dias-Santagata, D., Sadow, P.M., Lynch, K.D., Lubitz, C., Donovan, S.E., Zheng, Z., Le, L., Iafrate, A.J., and Daniels, G.H. (2014a). Identification of oncogenic mutations and gene fusions in the follicular variant of papillary thyroid carcinoma. J. Clin. Endocrinol. Metab. *99*, E2457–E2462.

McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014b). Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell *156*, 1298–1311.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

Muller, F.L., Colla, S., Aquilanti, E., Manzo, V.E., Genovese, G., Lee, J., Eisenson, D., Narurkar, R., Deng, P., Nezi, L., et al. (2012). Passenger deletions generate therapeutic vulnerabilities in cancer. Nature *488*, 337–342.

Nijhawan, D., Zack, T.I., Ren, Y., Strickland, M.R., Lamothe, R., Schumacher, S.E., Tsherniak, A., Besche, H.C., Rosenbluh, J., Shehata, S., et al. (2012). Cancer vulnerabilities unveiled by genomic loss. Cell *150*, 842–854.

Nikiforova, M.N., Wald, A.I., Roy, S., Durso, M.B., and Nikiforov, Y.E. (2013). Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. J. Clin. Endocrinol. Metab. *98*, E1852–E1860.

Pathak, K.A., Klonisch, T., Nason, R.W., and Leslie, W.D. (2016). FDG-PET characteristics of Hurthle cell and follicular adenomas. Ann. Nucl. Med. *30*, 506–509.

Pryma, D.A., Schoder, H., Gonen, M., Robbins, R.J., Larson, S.M., and Yeung, H.W. (2006). Diagnostic accuracy and prognostic value of 18F-FDG PET in Hurthle cell thyroid cancer patients. J. Nucl. Med. *47*, 1260–1266.

Ramos, A.H., Lichtenstein, L., Gupta, M., Lawrence, M.S., Pugh, T.J., Saksena, G., Meyerson, M., and Getz, G. (2015). Oncotator: cancer variant annotation tool. Hum. Mutat. *36*, E2423–2429.

Rheinbay, E., Parasuraman, P., Grimsby, J., Tiao, G., Engreitz, J.M., Kim, J., Lawrence, M.S., Taylor-Weiner, A., Rodriguez-Cuevas, S., Rosenberg, M., et al. (2017). Recurrent and functional regulatory mutations in breast cancer. Nature *547*, 55–60.

Sabharwal, S.S., and Schumacker, P.T. (2014). Mitochondrial ROS in cancer: initiators, amplifiers or an Achilles' heel? Nat. Rev. Cancer *14*, 709–721.

Sabra, M.M., Ghossein, R., and Tuttle, R.M. (2016). Time course and predictors of structural disease progression in pulmonary metastases arising from follicular cell-derived thyroid cancer. Thyroid *26*, 518–524.

Safavi, S., and Paulsson, K. (2017). Near-haploid and low-hypodiploid acute lymphoblastic leukemia: two distinct subtypes with consistently poor prognosis. Blood *129*, 420–423.

Sagi, I., Chia, G., Golan-Lev, T., Peretz, M., Weissbein, U., Sui, L., Sauer, M.V., Yanuka, O., Egli, D., and Benvenisty, N. (2016). Derivation and differentiation of haploid human embryonic stem cells. Nature *532*, 107–111.

Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811–1817.

Schon, E.A., Rizzuto, R., Moraes, C.T., Nakase, H., Zeviani, M., and DiMauro, S. (1989). A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. Science 244, 346–349.

Shaha, A.R., Shah, J.P., and Loree, T.R. (1996). Patterns of nodal and distant metastasis based on histologic varieties in differentiated carcinoma of the thyroid. Am. J. Surg. 172, 692–694.

Shaham, O., Slate, N.G., Goldberger, O., Xu, Q., Ramanathan, A., Souza, A.L., Clish, C.B., Sims, K.B., and Mootha, V.K. (2010). A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. Proc. Natl. Acad. Sci. USA 107, 1571–1575.

Soding, J. (2005). Protein homology detection by HMM-HMM comparison. Bioinformatics 21, 951–960.

Solimini, N.L., Xu, Q., Mermel, C.H., Liang, A.C., Schlabach, M.R., Luo, J., Burrows, A.E., Anselmo, A.N., Bredemeyer, A.L., Li, M.Z., et al. (2012). Recurrent hemizygous deletions in cancers may optimize proliferative potential. Science 337, 104–109.

Stachler, M.D., Taylor-Weiner, A., Peng, S., McKenna, A., Agoston, A.T., Odze, R.D., Davison, J.M., Nason, K.S., Loda, M., Leshchiner, I., et al. (2015). Paired exome analysis of Barrett's esophagus and adenocarcinoma. Nat. Genet. 47, 1047–1055.

Stojadinovic, A., Hoos, A., Ghossein, R.A., Urist, M.J., Leung, D.H., Spiro, R.H., Shah, J.P., Brennan, M.F., Singh, B., and Shaha, A.R. (2002). Hurthle cell carcinoma: a 60-year experience. Ann. Surg. Oncol. 9, 197–203.

Taylor-Weiner, A., Zack, T., O'Donnell, E., Guerriero, J.L., Bernard, B., Reddy, A., Han, G.C., AlDubayan, S., Amin-Mansour, A., Schumacher, S.E., et al. (2016). Genomic evolution and chemoresistance in germ-cell tumours. Nature 540, 114–118.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14, 178–192.

Vafai, S.B., Mevers, E., Higgins, K.W., Fomina, Y., Zhang, J., Mandinova, A., Newman, D., Shaw, S.Y., Clardy, J., and Mootha, V.K. (2016). Natural product screening reveals naphthoquinone complex I bypass factors. PLoS One 11, e0162686.

Wada, N., Duh, Q.Y., Miura, D., Brunaud, L., Wong, M.G., and Clark, O.H. (2002). Chromosomal aberrations by comparative genomic hybridization in Hurthle cell thyroid carcinomas are associated with tumor recurrence. J. Clin. Endocrinol. Metab. 87, 4595–4601.

Wei, S., LiVolsi, V.A., Montone, K.T., Morrissette, J.J., and Baloch, Z.W. (2015). PTEN and TP53 mutations in oncocytic follicular carcinoma. Endocr. Pathol. 26, 365–369.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. 45, 1134–1140.

Zhang, H., Zeng, L., Liang, C., Qiu, H., Zhang, M., Zhu, Y., and Xie, C. (2012). Successful treatment of Hurthle cell thyroid carcinoma with lung and liver metastasis using docetaxel and cisplatin. Jpn. J. Clin. Oncol. 42, 1086–1090.

Zheng, S., Cherniack, A.D., Dewal, N., Moffitt, R.A., Danilova, L., Murray, B.A., Lerario, A.M., Else, T., Knijnenburg, T.A., Ciriello, G., et al. (2016). Comprehensive pan-genomic characterization of adrenocortical carcinoma. Cancer Cell 30, 363.

Zheng, Z., Liebers, M., Zhelyazkova, B., Cao, Y., Panditi, D., Lynch, K.D., Chen, J., Robinson, H.E., Shim, H.S., Chmielecki, J., et al. (2014). Anchored multiplex PCR for targeted next-generation sequencing. Nat. Med. 20, 1479–1484.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
AllPrep DNA/RNA FFPE Kit	QIAGEN	Cat# 80234
Agencourt FormaPure Kit	Beckman Coulter	Cat# A33342
Custom Design VariantPlex Kit	ArcherDX	Cat# CB0112
Archer® VariantPlex™ Protocol for Illumina®	ArcherDX	PR-SK0051-ILMN Rev A
FusionPlex® Solid Tumor Kit	ArcherDX	Cat# AB0005
Archer® FusionPlex® Protocol for Illumina®	ArcherDX	LA135.F
Deposited Data		
DNA sequencing of tumor samples	This paper	dbGaP: phs001580.v1.p1
Oligonucleotides		
Primers for mtDNA complex I genes (see Table S6)		
Kreatech™ PDGFRA (4q12) FISH probe, Red	Leica Biosystems	Cat# 04Q005B550
Vysis LSI EGFR SpectrumGreen Probe	Abbott Molecular	Cat# 07N98-020
Vysis CEP 8 SpectrumAqua Probe	Abbott Molecular	Cat# 06J54-018
Software and Algorithms		
Picard	Broad Institute, Cambridge, MA	http://picard.sourceforge.net/
Firehose	Broad Institute, Cambridge, MA	http://www.broadinstitute.org/cancer/cga/Firehose
ContEst	Cibulskis et al., 2011	N/A
deTiN	Rheinbay et al., 2017	N/A
MuTect	Cibulskis et al., 2013	N/A
Strelka	Saunders et al., 2012	N/A
Oncotator	Ramos et al., 2015	N/A
GATK version 3.3	McKenna et al., 2010	N/A
Variant effect predictor	McLaren et al., 2016	N/A
STAR	Dobin et al., 2013	N/A
ReCapSeg	Broad Institute, Cambridge, MA	http://gatkforums.broadinstitute.org/gatk/categories/recapseg-documentation
AllelicCapSeg	Broad Institute, Cambridge, MA	http://archive.broadinstitute.org/cancer/cga/acsbeta
ABSOLUTE	Carter et al., 2012	N/A
PhylogicNDT	Landau et al., 2017	N/A
Clustal Omega	Soding, 2005	N/A
PyMOL 2.0	Schrodinger	http://www.pymol.org/
INSPIRE v.200.1.388.0 and IDEAS 6.2.64.0	Millipore Sigma	N/A
Other		
COSMIC	Sanger Institute, Cambridgeshire, UK	http://cancer.sanger.ac.uk/cosmic

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David McFadden (david.mcfadden@utsouthwestern.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Tissue Studies

Cases of HCC were identified via a natural language search of the electronic medical record at Massachusetts General Hospital using the terms Hürthle and oncocytic to identify cases between 01/01/1990 – 05/21/2013. Patients with HCC were also prospectively identified after that time period leading to a collection of cases with surgery performed between 1994-2015. Individual pathology reports were reviewed by D.G.M. to identify carcinoma samples. This led to a cohort of n=70 samples derived from n=41 patients. Clinical annotation of each sample, including patient characteristics derived from the medical record, can be found in Table S1. Formalin-fixed paraffin-embedded histopathology blocks were obtained from tissue archives and slides were reviewed by an expert endocrine pathologist (P.M.S.). Regions of high tumor cellularity were circled and excised. DNA was then isolated from the excised tissue using the AllPrep DNA/RNA FFPE Kit (QIAGEN) and submitted for whole exome sequencing (Broad Institute). All specimens were obtained and sequenced with approval of the Partners Human Research Committee (protocol #2013P000811).

METHOD DETAILS

Nuclear Genome Analysis

Whole exome sequencing was performed using the Illumina exome with all targets from the Agilent exome design (Agilent SureSelect All Exon V2), all coding regions of Gencode V11 genes, and all coding regions of RefSeq gene and KnownGene tracks from the UCSC genome browser (http://genome.ucsc.edu). Pooled libraries were normalized to 2nM and denatured using 0.2 N NaOH prior to sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using either the HiSeq 2000 v3 or HiSeq 2500. Each run was a 76bp paired-end with a dual eight-base index barcode read. Data was analyzed using the Broad Picard Pipeline which includes de-multiplexing and data aggregation. Data derived from whole exome sequencing were processed using established analytical tools. The Picard pipeline (http://picard.sourceforge.net/) aligned sequencing reads to the hg19 human genome build. These BAM files were uploaded into the Firehose platform (http://www.broadinstitute.org/cancer/cga/Firehose), which coordinates the flow of datasets through distinct algorithms. Tumor and normal genotypes were compared for concordance. ContEst was used to estimate the level of cross-contamination (Cibulskis et al., 2011). OxoG artifacts, which could arise due to oxidation of a G base pair on only one strand during fragmentation, and FFPE-induced artifacts were filtered (Costello et al., 2013). Joint re-alignment was performed using the Novoalign and blat filters. A panel of normal samples (PoN) was used to filter the remaining false positives, including potential sequencing artifacts and germline sites missed in the matched normal tissue (Cibulskis et al., 2013). Additionally, to recover mutations lost to tumor-in-normal contamination from adjacent tissue controls we applied the deTiN pipeline (Rheinbay et al., 2017). The latest version of MuTect (Cibulskis et al., 2013) was used to call somatic single-nucleotide variants (SNVs) and Strelka (Saunders et al., 2012) to call small insertions or deletions (indels) on each tumor-normal pair. The consequence of all coding was annotated using Oncotator (Ramos et al., 2015). Manual review of mutations was performed using Integrated Genomics Viewer (Thorvaldsdóttir et al., 2013) and alterations that were invalid based on this IGV review were excluded from the final results (including mutations called only on duplicate reads, mutations with orientation bias, mutations called on reads that also contained indels and other low allelic fraction SNVs, mutations called in regions with poor mapping).

Somatic Copy Number Analysis

Copy number alterations were estimated using ReCapSeg (http://gatkforums.broadinstitute.org/gatk/categories/recapseg- documentation). For each tumor sample, the read coverage profile of each target segment was first normalized with the total number of aligned reads. Then, a panel of normal samples (PoN) sequenced at the same target regions was used to normalize the read depth at each capture probe. These normalized target regions were segmented and gave rise to copy number data. The allelic copy number ratio was estimated by integrating allele fraction data from heterozygous germline variants in the tumor sample that were detected in the matched normal sample. The AllelicCapSeg tool was used to combine copy number and allele fraction data (http://archive.broadinstitute.org/cancer/cga/acsbeta). The contribution of each homologous allele to copy number alterations (including copy neutral events) was inferred from allelic copy numbers (Stachler et al., 2015; Brastianos et al., 2015; Brastianos et al., 2017).

Purity, Ploidy, and Phylogeny

We performed a 'force-calling' method to identify mutations shared across several samples from the same individual (Stachler et al., 2015). This process allowed us to aggregate all SNVs found in each tumor sample from a patient. These force-called mutant and reference allele counts for each mutation as well as allelic copy numbers were used as input to ABSOLUTE (Carter et al., 2012) to provide ploidy and purity estimates for each tumor. ABSOLUTE solutions were manually curated. Subsequently, total allelic copy numbers and cancer cell fractions (CCF) were generated. Using ABSOLUTE allelic copy numbers clonal focal copy number changes were identified and manually reviewed. These segments were analyzed for regions exhibiting significant SCNA alterations that have been reported previously (Zack et al., 2013). Cases with altered genomic locations described to have a clear target gene in this previous pan-cancer analysis were summarized.

To analyze the phylogenetic relationship between populations across tumors from one individual we used PhylogicNDT (Landau et al., 2017), an extension of Phylogic (Landau et al., 2013; McFadden et al., 2014b), that takes the cancer cell fraction (CCF) distributions provided by ABSOLUTE to determine the phylogenetic tree that best represents the clonal evolution among multiple tumor

samples from the same patient. PhylogicNDT is an *N*-dimensional Bayesian clustering framework based on mixtures of Dirichlet Processes (Escobar and West, 1995), where the number of clusters is inferred over many Markov chain Monte Carlo (MCMC) iterations. Phylogenetic trees were drawn for each individual using the CCF estimates of these clusters for clonal mutations, post-assigning clonal SCNAs. This was done using a neighbor-joining algorithm (Biopython) based on the jaccard distance metric (SciPy) for each sample (Cock et al., 2009).

Validation of Nuclear Variants

Targeted next-generation sequencing was performed using anchored multiplex PCR (AMP) and previously described methodologies (Zheng et al., 2014). Total nucleic acid was isolated from FFPE tumor tissue using the Agencourt FormaPure Kit (Beckman Coulter). A customized version of the VariantPlex™ Solid Tumor Kit (Custom Design VariantPlex Kit, ArcherDX Inc.) was used to validate a subset of nuclear variants identified by WES, and for the detection of hotspot mutations in the *TERT* gene promoter (Table S3). The assay detects single nucleotide variants (SNV) and insertion/deletions present at ≥5% allelic frequency in regions of genomic DNA with sufficient read coverage and targets the following genes (exons): ABL1 (4-7), AKT1 (3,6), ALK (21-23,25), APC (16), ARID1A (1-20), ATM (1-63), ATRX (1-35), AURKA (2,5-8), BRAF (11,15), BRCA1 (2-23), BRCA2 (2-27), CCNE1 (3-8,10,12), CDH1 (1-16), CDK4 (2-7), CDKN2A (1-3), CIC (1-20), CSF1R (7,22), CTNNB1 (3), DAXX (1-8), DDR2 (12-18), DDX3X (1-17), EGFR (3,7,15,18-21), ERBB2 (8,10,19-21,24), ERBB3 (2-3,7-8), ERBB4 (3-4,6-9,15,23), ESR1 (8), EZH2 (16), FBXW7 (1-11), FGFR1 (4,7-8,13,15,17), FGFR2 (7,9,12,14), FGFR3 (7-9,14-16,18), FLT3 (11,14,16,20), FOXL2 (1), GNA11 (5), GNAQ (4-5), GNAS (6-9), H3F3A (2), HNF1A (3-4), HRAS (2-3), IDH1 (3-4), IDH2 (4), JAK2 (11,13-14,16,19), JAK3 (4,13,16), KDR (6-7,11,19,21,26-27,30), KEAP1 (2-6), KIT (2,8-11,13-15,17-18), KRAS (2-5), MAP2K1 (2,3,6-7), MAP3K1 (1-20), MDM2 (2-4,6,8,10), MEN1 (2-10), MET (2,11,14,16,19,21), MLH1 (12), MPL (10), MSH6 (1-10), MSI, MYC (1-3), MYCN (3), NF1 (1-58), NF2 (1-15), NOTCH1 (25-27,34), NPM1 (11), NRAS (2-5), PIK3CA (2,5,7-8,10,14,19,21), PIK3R1 (1-10), POLE (9-14), PTCH1 (1-23), PTEN (1-9), PTPN11 (3,13), RB1 (1-27), RET (10-11,13-16), RHOA (2-3), RNF43 (2-10), ROS1 (38), SDHB (1-8), SMAD2 (7), SMAD4 (2-12), SMARCA4 (3-36), SMARCB1 (2,4,5,9), SMO (3,5-6,9,11), SRC (14), STAG2 (3-34), STK11 (1-9), SUFU (1-12), TERT (1), TP53 (1-11), TP63 (1-14), TSC1 (3-23), TSC2 (2-42), TSHR (10), VHL (1-3). We attempted validation of all nuclear variants that were detected by exome sequencing and targeted by the VariantPlex genotyping assay. A total of 59 samples from 40 patients were tested (Table S3). Of these, 7 samples had unavailable data (indicated as N/A, column D, Table S3). The remaining 52 tumors (from 37 individuals) harbored 31 unique nuclear variants. We confirmed 24/29 (83%) of those variants by genotyping or Sanger sequencing (column D, Table S3). Visual inspection of VariantPlex sequencing data for each individual locus revealed that 4 out of the 5 variants that we were unable to validate (column E, Table S3), were not adequately covered by unique sequencing reads, and were thus inconclusive. Therefore, out of the 25 nuclear variants that could be appropriately assessed in the region of interest, 24 (96%) were confirmed by genotyping and Sanger sequencing, indicating high-fidelity mutation calling by our WES pipeline.

Detection of Gene Rearrangements

Assessment of gene rearrangements was accomplished using the FusionPlex® Solid Tumor Kit (ArcherDX Inc.), a targeted fusion assay designed to detect fusion transcripts in reverse transcribed double-stranded complementary DNA (cDNA), obtained from samples with ≥5% tumor cellularity. The fusion assay detects rearrangements in the following genes (exons): *ADCK4* (1-2, 4-6, 9-10, 12-15), *AKT3* (1-2, 13), *ALK* (1, 3, 17, 19-22, 29), *AR* (1-4, 6, 7-8), *ARHGAP6* (1-3), *ARHGAP26* (10- 13), *AXL* (14-15), *BRAF* (1-2, 8-11, 17), *BRD4* (1, 10-12), *CCDC6* (1-8), *CD74* (1-8), *CHTOP* (2- 6), *EGFR* (7-9, 14-18, 23-28), *ERBB2* (2-4), *ERBB4* (17-18, 20), *ESR1* (3-5), *EWSR1* (1, 3-8, 12-13), *FGFR1* (1, 7-13, 16-18), *FGFR2* (3-4, 17), *FGFR3* (3, 7-12, 15-18), *FGR* (2-3), *INSR* (13-18, 21-22), *INSRR* (13-18, 21-22), *JAK1* (1-7, 9-25), *JAK2* (1, 6, 9, 11-12, 16-17, 19, 24), *MAML2* (2-4), *MAST1* (2, 8, 19-20, 26, 29), *MAST2* (1, 5), *MET* (2, 11-16, 20-21), *MUSK* (8-9, 11-14), *NFIB* (1, 7-9), *NOTCH1* (2, 27-28, 34), *NOTCH2* (1, 27, 33), *NRG1* (2-4, 6), *NTRK1* (1, 8-17), *NTRK2* (9-11, 13-20), *NTRK3* (1, 11-16, 18-19), *NUMBL* (3-7, 9-10), *NUTM1* (2-3), *PDGFB* (1-2, 6), *PDGFRA* (1, 9-11, 13-14, 20-23), *PIK3CA* (2-3), *PKN1* (9-14), *PLAG1* (2-4), *PPARG* (3-8), *PRKACA* (2-4), *PRKCA* (3-7), *PRKCB* (3-7), *RAF1* (1, 9-11, 17), *RET* (1, 8-13, 19), *RHOA* (1-5), *ROS1* (1, 31-37, 43), *TMPRSS2* (1-5). Double stranded genomic DNA (VariantPlex™) or reverse-transcribed cDNA (FusionPlex®) was enzymatically sheared, end-repaired, adenylated, and ligated with a half-functional adapter. The sequencing library was generated using two hemi-nested PCR reactions following the manufacturer's VariantPlex™ and FusionPlex® Protocols for Illumina, respectively. Next-generation sequencing was performed using the Illumina NextSeq 2x150 (Illumina) (Zheng et al., 2014).

mtDNA Variant Calling and Validation

All WES reads mapping either to GRCh37 chromosome MT or to 1049 nuclear sequences of mitochondrial origin (NUMTs) (Li et al., 2012) were re-aligned to mtDNA revised Cambridge Reference Sequence (NC_012920 equivalent to GRCh37 chromosome MT) using GATK version 3.3 (McKenna et al., 2010), BWA version 0.7.10 (Li and Durbin, 2009), and Picard Tools version 1.119 (broadinstitute.github.io/picard/). Variants were detected via the GATK HaplotypeCaller assuming a mixture of 20 chromosomes (HaplotyperCaller -ploidy 20 -L MT:1-16569 –heterozygosity 0.01 -contamination 0). The HaplotypeCaller's estimate of variant allelic fraction (VAF; equivalent to heteroplasmy in normal samples) was used to identify somatic mutations using the following seven filters: (1) VAF ≥ 0.05 in tumor; (2) VAF ≤ 0.05 in matched normal; (3) if variant detected at low VAF in matched normal (i.e. 0<VAF ≤ 0.05), then tumor VAF ≥ 0.25; (4) Variant supported by >5 reads in tumor; (5) Depth ≥ 15X in tumor; (6) Depth ≥ 10X in matched normal; (7) StrandBias FS score<20. Depth filters were derived from published power calculations (Cibulskis et al., 2013). Manual inspection

of reads was performed on all variants VAF ≥ 0.3. Six priority variants VAF ≥ 0.3 that were excluded by depth criteria (m.4830G>A, m.5032G>A, m.6352T>C, m.6745G>A, m.12769GA>G, m.12778G>A) were included in our analysis with 2 of these variants being validated by Sanger sequencing, while 3 others were found in multiple tumors from a single patient. Variants were annotated using online Variant Effect Predictor (VEP) (McLaren et al., 2016) (June 2016, Gencode basic transcripts). Very similar somatic mutation calls were obtained using the Mutect2 tool from GATK version 3.6 (McKenna et al., 2010) (parameter initial_normal_lod=0.0, no filtering applied); specifically, 40/46 variants VAF ≥ 0.3 by either pipeline were detected in both pipelines. Analysis of variants VAF ≥ 0.3 was limited to mutations that exceeded 0.3 both using HaplotypeCaller's estimate of allelic fraction as well as point estimates using all reads passing quality filters. To identify mtDNA deletions, we searched for split-reads using STAR (Dobin et al., 2013) (parameters –outSJfilterOverhangMin 12 –alignSJoverhangMin5 –scoreDelOpen -10000 –scoreInsOpen -10000), filtering for break points supported by at least 3 split-reads on each strand.

Ten randomly selected non-silent variants with VAF ≥ 0.3 in mitochondrial complex I genes were validated by Sanger sequencing in tumors. The primers utilized for mtDNA complex I genes are provided in Table S6.

mtDNA Pan-Cancer Comparative Analysis

HCC mtDNA variants were compared to published mtDNA somatic mutations detected in a pan-cancer analysis of 1664 tumor samples (Ju et al., 2014), including both single nucleotide variants and indels. Variant allele frequency and gene annotations were derived from published tables. LOF mutations included stop-gained (nonsense) and frameshift indels. Silent and non-coding variants included synonymous coding variants, intergenic regions, D-Loop regions, tRNAs and rRNA mutations. MELAS comparisons included all 19 known tRNA mutations associated with the syndrome annotated in MitoMap (Brandon et al., 2005) (positions 583, 1630, 1642, 1644, 3243, 3244, 3252, 3256, 3258, 3260, 3271, 3291, 4332, 5541, 8316, 12147, 12206, 12299, 14693). Evolutionary conservation scores were calculated for each position corresponding to a missense mutation. The Amino Acid Conservation Score indicates the number of species with identical residues across a pre-computed alignment of the mitochondrial genome across 44 vertebrate species (Lindblad-Toh et al., 2011), as described previously (Calvo et al., 2006). To search for LOF germline variants, we applied VEP to all published germline variants, which included only single nucleotide variants with VAF ≥ 0.3 and did not include indels or low heteroplasmy variants.

Complex I Structural Analysis

We mapped the human missense HCC mutations onto the 3.3 Å resolution X-ray structure of complex I from *Thermus thermophilus* (PDB: 4HEA) (Baradaran et al., 2013). Amino acid sequences of human complex I subunits (ND1, ND2, ND3, ND4, ND4L, ND5, and ND6) were aligned with the corresponding subunits of *Thermus thermophilus* (Nqo8, Nqo14, Nqo7, Nqo13, Nqo11, Nqo12, and Nqo10, respectively) using Clustal Omega (Soding, 2005). With the exception of S531N mutation in ND5, for which the sequence alignment was uncertain, all other mutations were found in evolutionarily conserved segments of the protein. Figures illustrating the location of the mutations in the structure were generated with PyMOL 2.0 (http://www.pymol.org/). The proton pumping domains of ND5, ND4 and ND2 were superimposed using "align" function of PyMOL.

Fluorescence In-Situ Hybridization (FISH)

Triple-color FISH was used to simultaneously assess the copy number of chromosomes 4 and 8 (that frequently displayed LOH and appeared to be haploid in our cohort), and of chromosome 7 (which always preserved heterozygosity, and was frequently present in either a diploid or a tetraploid state). Triple-color FISH was performed on interphase nuclei from 5-μm sections of paraffin-embedded tumor tissue, following the Vysis manufacturer's protocol, and using probes for *PDGFRA* (Kreatech Red, Leica Biosystems) to evaluate chromosome 4, *EGFR* (Vysis Spectrum Green, Abbot Molecular) specific for chromosome 7, and CEP8 (Vysis Spectrum Aqua, Abbott Molecular) corresponding to centromere 8. FISH analysis was pursued to determine absolute chromosomal copy numbers in tumor cells, however, the evaluation of hybridization results is complicated due to the nuclear truncation occurring in tissue sections. To validate the method, we performed FISH on a subset of tumors spanning the three main ploidy classes identified by ABSOLUTE: 5 "diploid", 5 "near-haploid", and 2 "complex" cases ([V]FISH validation cohort, Figure S3).

Imaging Flow Cytometry and Gating Strategy

Disaggregated nuclei for flow cytometry were prepared from formalin fixed paraffin-embedded samples per protocol (Hyytinen et al., 1994) with minor modifications. Two 50 μM sections for each tumor were deparaffinized in xylenes and rehydrated in an ethanol series ending in dH₂0. Samples were digested in Carlsberg protease (Proteinase type VIII, Sigma Aldrich), 0.1% in 0.1M Tris/ 0.07 M NaCl, for one hour at 37°C with frequent agitation and vortexing to release nuclei. The nuclear suspension was separated from undigested material and clumps by either filtration or repeated standing/decanting/rinsing. The final suspension was centrifuged and resuspended in 70% ice cold ethanol for 24 hrs. Cells were washed in PBS then incubated in 20ug/mL Propidium Iodide (PI) (Sigma P-4170) in PBS plus 1 mg/mL Type IIa Ribonuclease-A (Sigma R-5000) for 15 minutes at RT and then filtered using 35 uM cell strainers (BD Falcon 352235) and run on a SORP 8 laser LSRII. Events were first gated on FSC-H vs SSC-A (log). PI-A vs PI-H for doublet correction. PI intensity was measured using PI-A. For imaging flow cytometry, 50k event (RIF) files were collected on an Amnis ImageStream MKII Imaging Flow Cytometer (Amnis EMD Millipore, now a part of Sigma) using INSPIRE (v.200.1.388.0) Data Acquisition software. Data was collected using a Brightfield (BF) Area histogram gate < 50 microns² to gate out beads and small debris, plus a low stringency BF Gradient RMS histogram gate <35 to gate out any out of focus events due to potential clogs

or bubbles. 5 mW 488 nm laser power was used to illuminate PI and 2mW 785nm laser power was used for side scatter (SSC). Data analysis was performed using IDEAS 6.2.64.0 software. A hierarchical gating strategy was used to identify PI positive, single cell nuclei as a parent population for the analysis. Comparison of various stringencies of the gating hierarchy (ALL; PI positive; single cells by Area vs. Aspect ratio of PI; with/without gating out saturated events by Raw Max Pixel of PI) showed no appreciable shift in the modal PI intensity peak. The higher stringency gating enriched the overall percent of events under the peak (see Supplemental Information). Representative images were chosen from (+/- 500) of the modal PI intensity peak. PI Intensity and Raw Max Pixel measurements used a Combined Mask (MC). Mean Pixel PI values used a Ch03 Morphology Mask. Area BF measurements used a Ch09 AdaptiveErode 95% mask.

QUANTIFICATION AND STATISTICAL ANALYSIS

Nuclear Mutation Significance Analysis

Mutational significance analysis was performed using the combination of three MutSig tools (MutSig2CV): (i) MutSigCV (Lawrence et al., 2013) to detect genes with increased burden of mutations by comparing the number of non-silent mutations to the number that is expected by chance; (ii) MutSigCL (Lawrence et al., 2014) to detect genes with clustering of mutations along the gene sequence that is unlikely by chance; (iii) MutSigFN (Lawrence et al., 2013) to detect genes with mutations that are enriched in likely functional positions identified by evolutionary conservation. To detect candidate cancer genes p values using Fisher's p value combination were integrated. We corrected for multiple-hypothesis testing and identified candidate cancer genes as those with Benjamini-Hochberg False Discovrey Rate (BH FDR) q-value $\leq$ 0.1. We used an approach of aggregated tumors that represent the union of unique SNVs and indels found across all tumors from one individual for analyzing mutational significance. In addition, we used restricted hypothesis testing using a panel of 326 previously published cancer genes (Forbes et al., 2017; Lawrence et al., 2014), to identify additional recurrently mutated genes. Sample size and description of significance testing can be found in the results section. To test if tumor suppressor mutations were enriched in cases with widespread chromosomal loss, a Fisher's exact test was used on data from each patient where chromosomal state was determined by the degree of LOH (<20 or >20 arms) and each patient was counted as an event if at least one TS mutation was present.

mtDNA Mutation Significance Analysis

Prevalence comparisons between our cohort and the pan-cancer analysis were calculated as the percent of samples having at least one somatic mutation VAF $\geq$ 0.3 in each category (e.g. non-silent mutation in complex I). The mutation rate between HCC and the pan-cancer cohort was compared using a heteroscedastic two-sided t-test. Enrichment analysis for protein-modifying mutations between HCC and the pan-cancer cohort was done using a Fisher's Exact Test. Enrichment for protein-modifying alterations in complex I genes was based on a two-sided test of proportions. VAF comparisons for complex I variants between HCC and the pan-cancer samples were made using a two-sided Wilcoxon rank sum test. Sample size, mean and median designations, error bar definitions, and p values for all mtDNA data can be found in Figure 4 and its legend.

Quantification of FISH Results

At least 50 nuclei per sample were assessed and scored for the number of FISH signals present for each targeted chromosome in a blinded fashion. To enrich for whole nuclei (i.e. minimize scoring of truncated nuclei), FISH assessment focused on nuclei that: (1) had at least one signal corresponding to each individual probe, and (2) had a minimum of 2 signals for at least one of the probes. FISH analysis was then extended to tumors with Allelic CapSeq data consistent with massive LOH and either "near-haploid" or "complex" genomes. In Figure S3, the chr 4: chr 7: chr 8 FISH signal ratio was used to distinguish tumors displaying a near-haploid (NH) genome (1:2:1 ratio), from samples arising from a NH tumor that had undergone Whole Genome Duplication (WGD) (2:4:2 ratio).

Kaplan-Meier Analysis

Cumulative survival analysis was performed using the Kaplan-Meier method; curves were compared with the log-rank test. Patients were classified as having widespread LOH if at least 9 chr arms had undergone LOH.

DATA AVAILABILITY

The whole exome sequencing data is available in dbGAP, the accession number is phs001580.v1.p1.

Supplemental Information

Widespread Chromosomal Losses and Mitochondrial

DNA Alterations as Genetic Drivers

in Hürthle Cell Carcinoma

Raj K. Gopal, Kirsten Kübler, Sarah E. Calvo, Paz Polak, Dimitri Livitz, Daniel Rosebrock, Peter M. Sadow, Braidie Campbell, Samuel E. Donovan, Salma Amin, Benjamin J. Gigliotti, Zenon Grabarek, Julian M. Hess, Chip Stewart, Lior Z. Braunstein, Peter F. Arndt, Scott Mordecai, Angela R. Shih, Frances Chaves, Tiannan Zhan, Carrie C. Lubitz, Jiwoong Kim, A. John Iafrate, Lori Wirth, Sareh Parangi, Ignaty Leshchiner, Gilbert H. Daniels, Vamsi K. Mootha, Dora Dias-Santagata, Gad Getz, and David G. McFadden
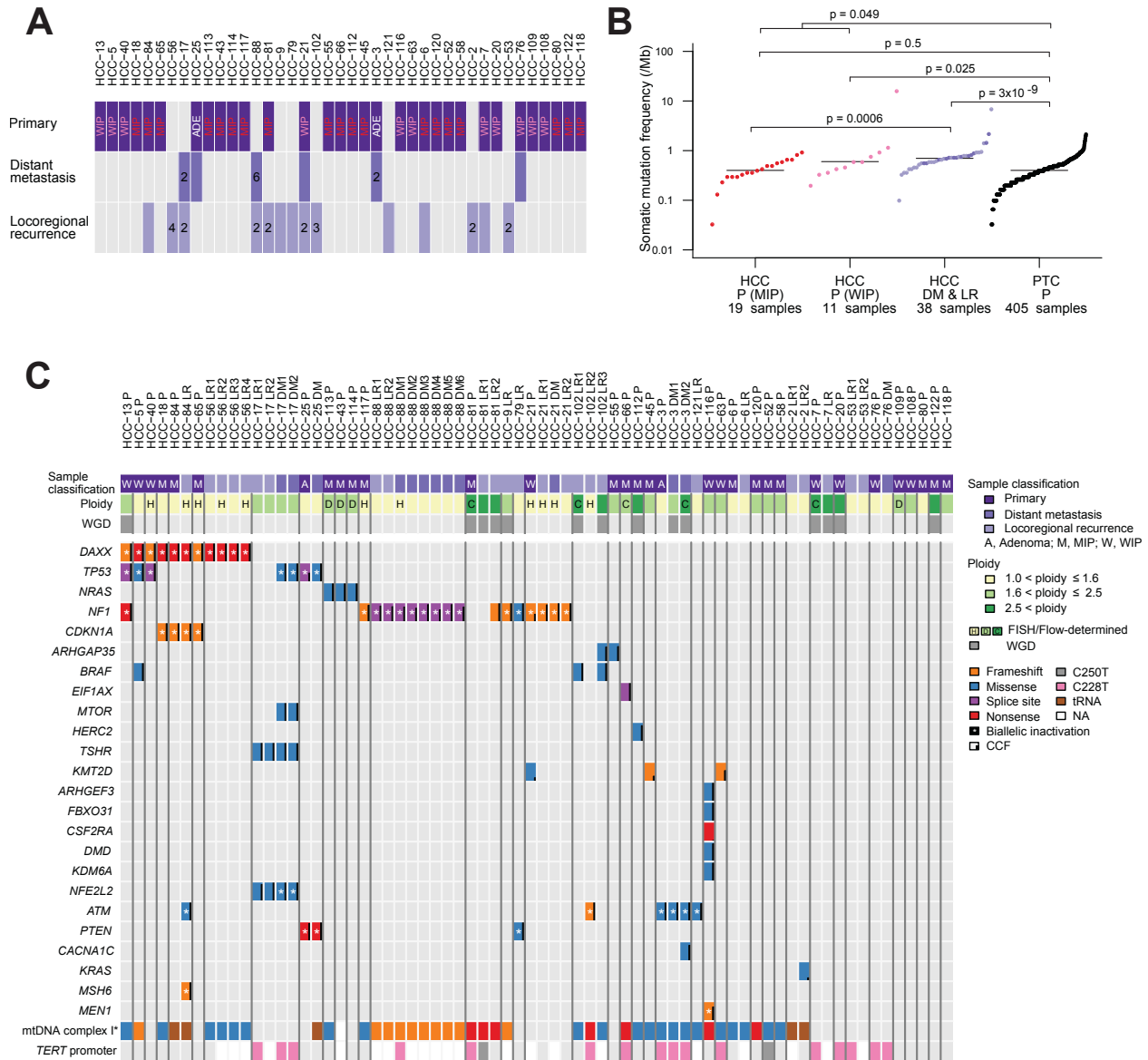
Figure S1, related to Figure 1. (A) HCC samples are summarized for each patient. Purple shading indicates the presence of a tumor sample, either primary, distant metastasis, and/or locoregional recurrence, for that patient. Grey color indicates no sample was present. Numbers indicate quantities of sequenced samples; primary samples are differentiated into widely invasive primaries (WIP), minimally invasive primaries (MIP) and adenoma (ADE) according to the pathological report. (B) The somatic mutation frequency is shown for HCC and papillary thyroid carcinoma cases (annotated as PTC) on logarithmic scale. Primary samples (P) are divided into widely invasive primaries (WIP) and minimally invasive primaries (MIP); two adenoma cases were omitted. Distant metastases (DM; dark purple) and locoregional recurrences (LR; light purple) were grouped together. The median is depicted; the Wilcoxon rank-sum test was used to test for differences. (C) The CoMutPlot is grouped according to cluster of nuclear mutations identified. Ploidy results of FISH and Imaging Flow Cytometry (Flow) are indicated by letters for near-haploid (H), diploid (D), and complex (C) tumors. The presence of WGD is depicted. Each column represents one tumor; black vertical lines separate different patients and white

vertical lines separate tumors from the same individual. Below, nuclear mutations and mtDNA mutations in complex I and tRNA genes (mtDNA complex I*) as well as *TERT* promoter mutations are shown. White stars indicate the presence of LOH in the second allele (i.e., biallelic inactivation). White boxes show tumors with insufficient DNA coverage for analysis.

Figure S2, related to Figure 2. Clonal focal copy number changes of oncogenes and tumor suppressor genes are delineated by their genomic position along the 22 chromosomes (top to bottom). Alterations are indicated by rectangles that cover the chromosomes in which alterations have been identified. Sample identifiers and classifications are shown at top with purple shading indicating the tumor type (first row). Each column represents one tumor with dark grey vertical lines separating different patients, and white vertical lines separating tumors from the same individual. Letters indicate the minimally invasive (M) vs widely invasive (W) primary samples as well as adenomas (A). Patients are ordered according to mean ploidy of 'aggregated tumors' (second row), and individual tumor samples from the same patient are ordered based on date of resection. Ploidy results of FISH and Imaging Flow Cytometry (Flow) are indicated by letters for near-haploid (H), diploid (D), and complex (C) tumors. Colors indicate loss or gain at a genomic locus. Note this figure depicts only focal copy number events.

A

Tumor ID	PDGFRA (chr 4) : EGFR (chr 7) : CEN8 (chr 8)			
	Median	Average	Top Call	
HCC-3 P	2:4:2	2.4:4.6:2.1	2:4:2	26%
HCC-3 DM2	2:4:2	2.1:3.9:1.8	2:4:2	19%*
HCC-7 P[v]	2:3:1	2.0:3.1:1.5	2:3:1	12%*
HCC-21 P	1:2:1	1.2:2.3:1.2	1:2:1	61%
HCC-21 LR1[v]	1:3:1	1.4:2.6:1.4	1:2:1	23%
HCC-21 LR2	1:2:1	1.1:2.3:1.2	1:2:1	58%
HCC-25 P	1:2:1	1.3:2.7:1.2	1:2:1	40%
HCC-25 DM	2:4:2	1.6:4.1:1.6	2:4:2	20%
HCC-40 P[v]	1:2:1	1.2:2.0:1.1	1:2:1	30%
HCC-43 P[v]	2:2:2	2.0:1.9:1.9	2:2:2	32%
HCC-56 LR2	1:2:1	1.1:2.5:1.1	1:2:1	63%
HCC-56 LR4	1:2:1	1.1:2.3:1.1	1:2:1	55%
HCC-65 P[v]	1:2:1	1.2:2.1:1.1	1:2:1	66%
HCC-66 P[v]	3:2:2	2.8:2.9:1.9	3:3:2	10%*
HCC-81 P[v]	3:2:2	2.5:3.3:1.7	3:4:2	16%*
HCC-84 LR[v]	1:2:1	1.2:2.2:1.3	1:2:1	50%
HCC-88 DM1	1:2:1	1.1:2.5:1.0	1:2:1	55%
HCC-102 LR1	3:4:2	2.5:3.8:1.8	3:4:2	24%
HCC-102 LR2	1:2:1	1.2:2.6:1.2	1:2:1	55%
HCC-109 P[v]	2:2:2	1.9:1.7:1.9	2:2:2	18%*
HCC-113 P[v]	2:2:2	1.8:1.9:1.8	2:2:2	18%*
HCC-114 P[v]	2:2:2	1.9:1.9:1.8	2:2:2	24%
HCC-117 P[v]	1:2:1	1.3:2.5:1.3	1:2:1	44%

[v] FISH validation cohort: * Heterogeneous population (Top Call detected in <20% of nuclei)

Near Haploid (HCC-40P)

PDGFRA (chr 4)	EGFR (chr 7)	CEP8 (chr 8)
1	2	1
2	2	1
1	2	2
1	3	1
1	2	1
1	1	1
2	1	2
1	5	1
2	5	2
2	2	3
3	2	2

1:2:1 → 30/50 nuclei (60%)
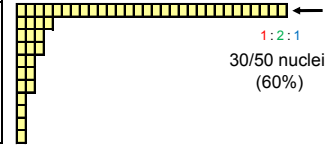
Diploid (HCC-43 P)

PDGFRA (chr 4)	EGFR (chr 7)	CEP8 (chr 8)
2	2	2
1	2	1
2	2	2
2	1	2
2	2	2
3	1	3
1	2	2
1	1	1
2	2	2
1	1	2
1	1	3
1	3	2
2	1	3
3	2	3
3	3	1
3	2	3
3	3	2
4	4	3

2:2:2 → 16/50 nuclei (32%)

Complex (HCC-81 P)

PDGFRA (chr 4)	EGFR (chr 7)	CEP8 (chr 8)
3	4	2
2	3	1
2	4	1
3	4	2
3	3	2
3	3	1
3	2	2
2	2	1
2	2	2
1	2	2
1	1	2
1	3	2
1	3	3
1	4	1
1	4	2
1	5	2
2	4	2
2	4	1
3	4	3
3	5	1
4	4	2
4	5	2
5	7	3

3:4:2 → 8/50 nuclei (16%)

B

Near-haploid — HCC-21 LR1 (59,074); HCC-40 P (57,979)

Diploid — HCC-43 P (93,011); HCC-109 P (72,032)

Complex/WGD — HCC-7 P (129,912); HCC-81 P (157,645)

Histograms: Normalized frequency vs PI Intensity (6e4, 1.2e5, 1.8e5)

Tumor ID	Count	%Gated	Mean PI Intensity	Modal PI Intensity	Mean Area (μM²)	Mean Pixel Intensity	Ploidy
HCC-21 LR1	18,363	53.7	59,185	59,074	79.8	250.6	Near-haploid
HCC-40 P	21,446	84.1	56,779	57,979	65.1	291.6	Near-haploid
HCC-43 P	34,689	87.3	89,736	93,011	79.2	402.5	Diploid
HCC-109 P	22,896	74.7	72,138	72,032	81.4	332.3	Diploid
HCC-7 P	10,923	54.8	130,311	129,912	192.7	429.5	Complex/WGD
HCC-81 P	5,783	59.3	160,711	157,645	142.2	468.6	Complex/WGD

Figure S3, related to Figure 3. (A) Summary of FISH Results. Left panel, copy numbers of chromosomes 4, 7 and 8 are depicted as the median and average (of at least 50 nuclei). The most prevalent FISH result in each cell population (Top Call) and its relative frequency is also shown. Right panel, the distribution of chr 4, chr 7, and chr 8 probe counts in representative examples of near-Haploid (HCC-40 P), diploid (HCC-43 P), and complex (HCC-81 P) tumors are shown. The Top Call by FISH and its prevalence is marked by an arrow. (B) Upper panel, frequency distribution histograms of HCC samples from each ploidy category obtained by Imaging Flow Cytometry (IFC). All plots show the normalized frequency for propidium iodide (PI) intensity of gated single cells. Horizontal lines in each plot indicate the R4 gated population that was selected from the modal PI intensity (value shown in each plot) and verified to represent single cells by imaging. Brightfield and fluorescent images of nuclei are shown in the upper right of each histogram with the area (μM^2; blue) and mean pixel intensity (yellow) shown along with a 10 μM scale bar. Aberrant PI peaks outside the R4 gate in complex/WGD samples represent tissue debris visually confirmed by imaging. Lower panel, summary table of metrics from flow cytometry with PI labeling. All values shown are from the R4 gated population of cells with the exception of the modal PI intensity which corresponds to the R3 parent population.
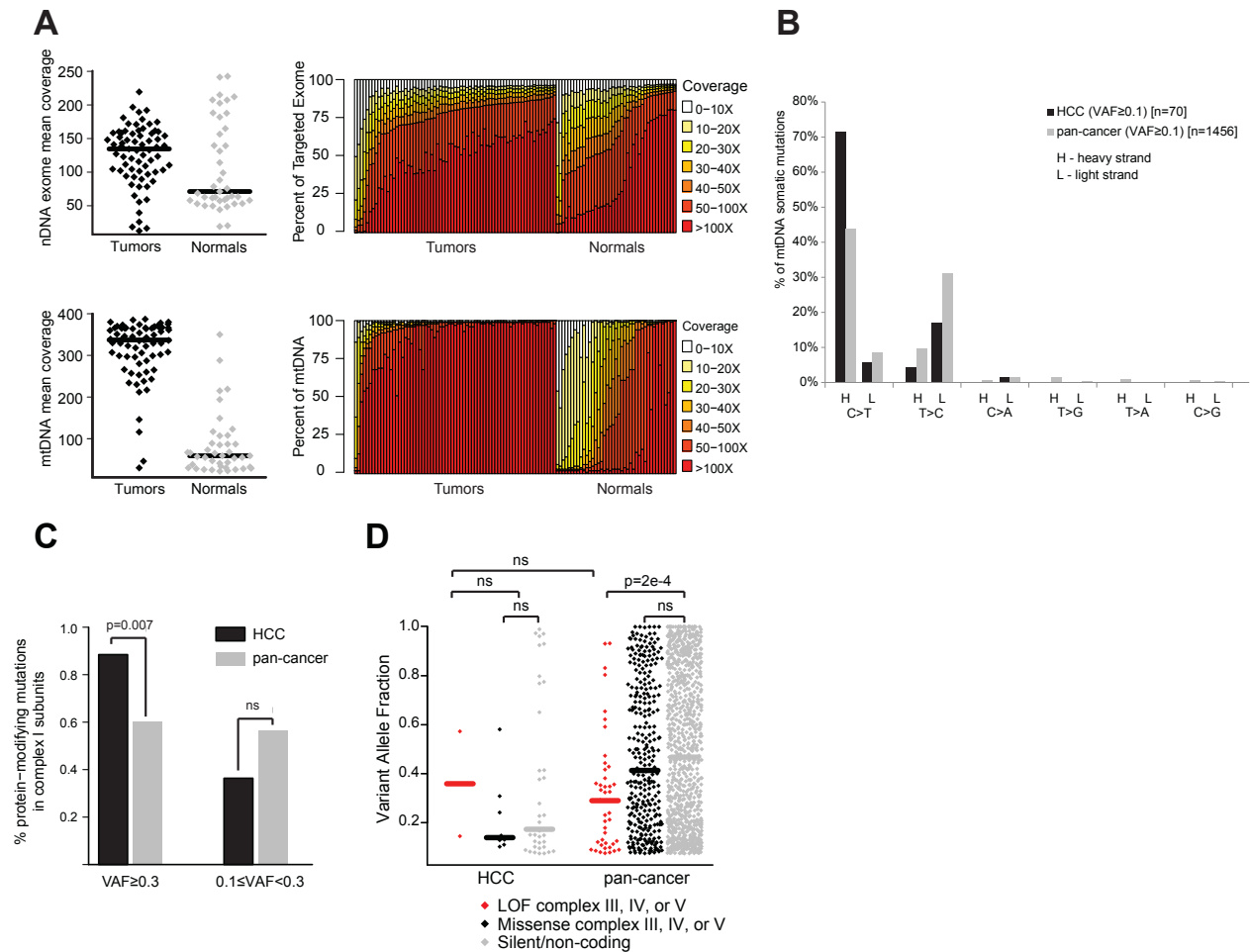
Figure S4, related to Figure 4. (A) Mean nDNA and mtDNA coverage is shown for tumor and normal samples. For each sample, distribution of different coverage depths is shown for the targeted exome and the mitochondrial genome. Samples are sorted separately for the percent of targeted exome and percent of mtDNA panels. (B) Mutational strand bias in mtDNA in HCC and the ICGC pan-cancer analysis with C>T mutations occurring preferentially on the heavy strand and T>C mutations occurring preferentially on the light strand. (C) Barplots show the percent of all protein-modifying somatic mutations in complex I subunits according to variant allele fraction (VAF) in HCC compared to the pan-cancer analysis; two-sided test of proportions. (D) Protein-modifying variants in complexes III, IV, or V do not show higher VAF compared to silent or non-coding variants in HCC and the pan-cancer analysis; two-sided Wilcoxon rank sum test; ns corresponds to a p value >0.05.

A

Years after diagnosis: 0 1 2 3 4 5 6 7

Legend:
- Clonal nDNA events (black line)
- Suclonal nDNA events (green line)
- mtDNA / *TERT* promoter (*TERT*-PR) events (manually added) (magenta line)
- Normal (black circle)
- Divergent evolution point (grey circle)

HCC-2
Loss chr 1,2,3,4,6, 8,9,18,14q,22q
ALPK2 p.R1915H
PPM1D p.Q510*

LR 1 — LR 2
Neck (1) — Thyroid (2)
0 2 4 6 8 9 10 12 14

HCC-3
Loss chr 1,2,3,4,5,6,8, 9,10,11,16,17,18,21, 13q,14q,15q
ATM *p.I323V*
MT-ND4 p.E123K
TERT C228T

Gain chr 4p,12p,15q
WGD
Loss chr 20
Gain chr 1q,6q,17q,21q
CACNA1C p.A879T
WGD
Loss chr 19q
Loss *CDKN1B*

P — DM 1 — DM 2
Thyroid — Rib Spine (1)(2)
0 1 2 3 4 5 6

HCC-6
Loss chr 1,2,3,4,6,8, 9,11,16,14q,15q,22q
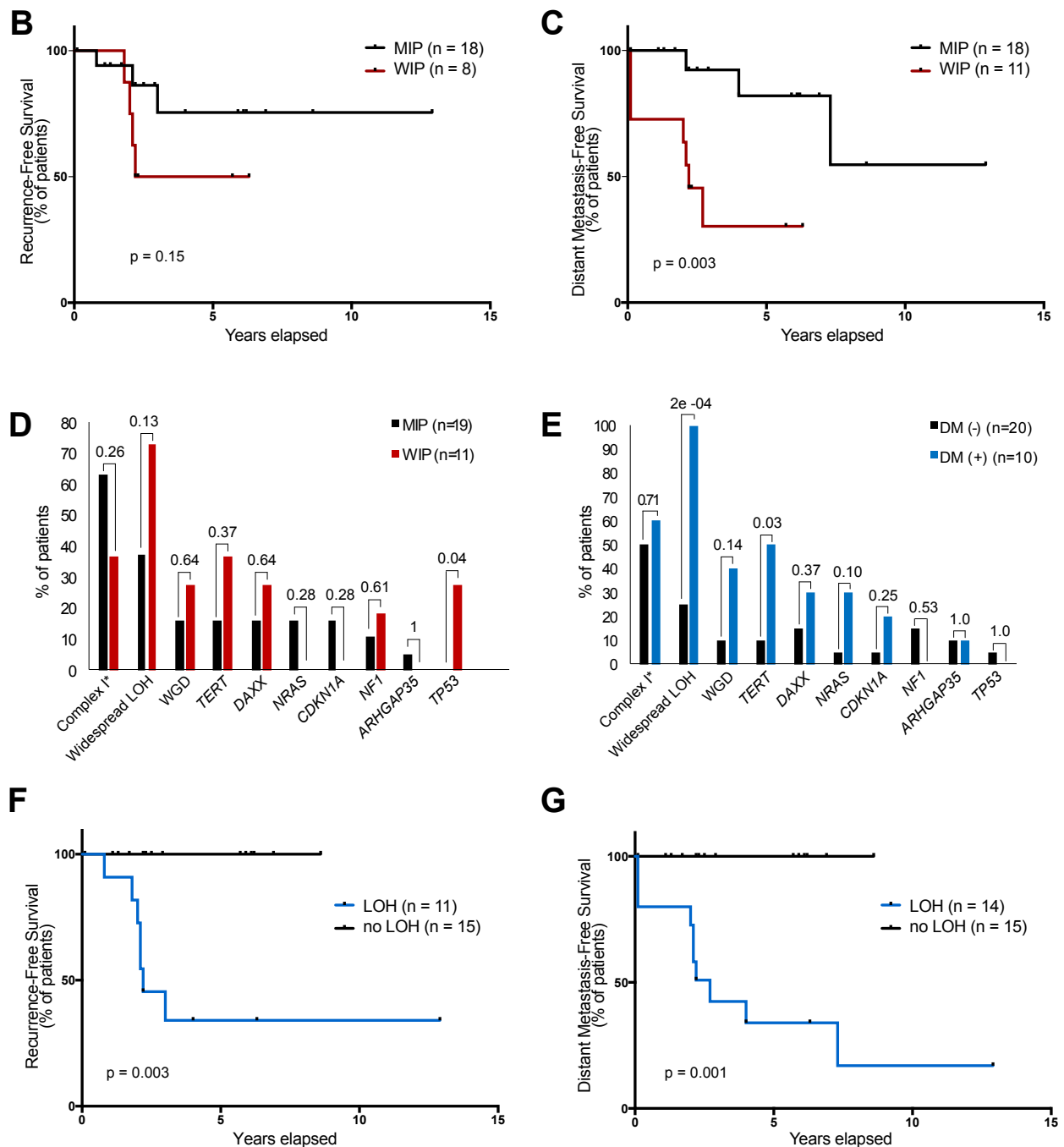*MAP2K2* p.R231L
MT-ND1 p.A147T

TERT gain

P — LR
Thyroid — Neck
0 1 2 3

HCC-7
Loss chr 1,2,3,4,6,8, 9,10,12,17,18,13q,14q, 15q,16q,19p,21q,22q
Gain chr 16p
WGD
(*TERT*-PR C228T)

Loss chr 19q
Gain chr 4q

P — LR
Thyroid — Trachea
0 1 2

HCC-17
TSHR p.I568T
TG p.P33L
NFE2L2 p.G81D
(*TERT*-PR C228T)

Loss chr 2,3
Gain chr 2,3,7,8,12, 16,17,19,20,1q
TP53 p.G245S
MTOR p.L2138F

Loss chr 19
NOTCH2 p.S1110F
ALB p.R0G

TG p.Q1809E

DM 2 — DM 1 | LR 1 — LR 2
Trachea (1)(2) — Breast (1)(2)
0 1 2 3 4 5 6 7

HCC-21
Loss chr 1,2,3,4,6,8,9, 10,11,17,13q,14q,22q
NF1 p.TY2284fs

MGA p.LKY745f
PTEN loss
Loss chr 19
Gain chr 2p,3q,4,6 8,9p,10p,13q,15q, 18,21q
KMT2D p.P1221S

P — LR 1 — DM — LR 2
Thyroid Neck (1) — Thorax Neck (2)
0 1 2

HCC-53
Loss chr 1,2,3,4,6,8,9,11, 16,18,14q,15q,21q,22q
Gain chr 16
(*TERT*-PR C228T)

LR 1 — LR 2
Neck (1) — Paratrachea (2)
0 1 2 3 4 5 6 7 8 9

HCC-76
Loss chr 1,2,3,4,6,8,9,10, 11,16,18,20,1q,13q,14q,15q, 21q,22q
TERT-PR C228T

DM — P
Thyroid Shoulder
0 1 2

HCC-88
Loss chr 1,2,3,6,8,9,11, 4q,10q,14q,15q,
MAP3K1 p.E1441Q
NF1 17:29587534:insT
MT-ND5 p.E145fs
(*TERT*-PR C228T)

DM 1 — DM 4
DM 3
DM 2
DM 6
MT-ND4 p.S35N
DM 5 — LR 1 — LR 2
Neck (1)(2) — Soft Tissue (1)(2)(3)(4)(5)(6) — Spine
0 1 2 3 4 5 6 7 8

Figure S5, related to Figure 6. (A) Phylogenetic trees show relationships between distinct tumor samples obtained from each patient. The length of the phylogenetic branches is based on the Jaccard distance between clones taking into account clonal nDNA events (CCF≥0.9). Clonal (black) and subclonal (green) gains, losses and oncogenic mutations are annotated onto the branches on which they occurred. High VAF complex I mtDNA mutations and *TERT* promoter mutations were manually added to the trees (purple). The top black dot represents the germline (normal) genome. The initial vertical line from the normal genome to the first divergence of the lines depicts events shared by all samples; lines between nodes depict events shared among a few; leaves depict events unique to the individual tumors that are

noted at the end. Below each tree, a timeline depicts the sequence of diagnosis and tissue sampling. Purple circle for the *MT-ND4* variant indicates its presence in DM1, DM2, DM3, DM4, DM5, and DM6, but not LR1 and LR2. (B) Kaplan-Meier curve showing recurrence-free survival (RFS, which includes local recurrences and distant metastases) of patients with minimally invasive (MIP, black) and widely invasive (WIP, red) primary tumors (n = 26). Patients were excluded if metastasis was diagnosed at initial presentation; only cases with primary tumors analyzed were included. Mantel-Cox test used to determine p value. (C) Kaplan-Meier curve showing distant metastasis-free survival of patients with MIP (black) and WIP (red) primary tumors (n = 29). Patient were not excluded if metastasis was diagnosed at initial presentation. Bar plots of specific alterations observed in MIP (black) and WIP (red) tumors, n = 30 (D), and in patients with distant metastases (DM+, blue) versus those without distant metastases (DM-, black), n = 30, (E) with nominal p values from Fisher's exact test above each bar. (F) Kaplan-Meier curve showing RFS based on primary tumors exhibiting widespread chromosomal losses (LOH, blue) and those without widespread chromosomal losses (no LOH, black). The same inclusion criteria as in panel B were used. Mantel-Cox test used to determine p value. (G) Kaplan-Meier curve showing distant metastasis-free survival based on primary tumors exhibiting widespread chromosomal losses (LOH; blue) and those without widespread chromosomal losses (no LOH; black). The same inclusion criteria as in panel C were used. Mantel-Cox test used to determine p value.