Mutational, transcriptional and viral shedding dynamics of the marine turtle fibropapillomatosis tumor epizootic

Kelsey Yetsko[1]*, Jessica Farrell[1]*, Maximilian R. Stammnitz[2], Liam Whitmore[3], Jenny Whilde[1], Catherine B. Eastman[1], Devon Rollinson Ramia[1], Rachel Thomas[1], Aleksandar Krstic[4], Paul Linser[1], Simon Creer[5], Gary Carvalho[5], Brooke Burkhalter[1], Elizabeth P. Murchison[2], Christine Schnitzler[1,6] and David J. Duffy[1,3,4,5].

[1] The Whitney Laboratory for Marine Bioscience and Sea Turtle Hospital, University of Florida, St. Augustine, Florida 32080, USA.

[2] Transmissible Cancer Group, Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK

[3] Department of Biological Sciences, School of Natural Sciences, Faculty of Science and Engineering, University of Limerick, Limerick, Ireland.

[4] Systems Biology Ireland & Precision Medicine Ireland, School of Medicine, University College Dublin, Belfield, Dublin 4, Ireland.

[5] Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, UK.

[6] Department of Biology, University of Florida, Gainesville, Florida, 32611, USA.

Correspondence to: David J. Duffy. Email: duffy@whitney.ufl.edu

Running title: Molecular origins and transmission of fibropapillomatosis

Key words: Environmental DNA; eDNA; Viral shedding; Whole genome sequencing; Precision oncology; Cancer; Transcriptomics; Fibropapillomatosis; green sea turtles; fibropapilloma-associated turtle herpesvirus (FPTHV); Chelonid herpesvirus 5 (ChHV5); Chelonid fibropapilloma-associated herpesvirus (CFPHV) tumor; Precision Wildlife Medicine; Conservation Medicine; Wnt signaling; MAPK Signaling; β-catenin; *Chelonia mydas;* mRNA sequencing (mRNA-seq); precision medicine; genome medicine; cancer; systems medicine; reptile tumors; immune activation; wildlife cancer models; pathogen-induced tumors; oncogenic mutations; eDNA-based pathogen monitoring.

Abstract

Sea turtle populations are directly and indirectly under threat from a range of anthropogenic processes. Perhaps the most visibly apparent of these is the disfiguring tumor disease epizootic (animal epidemic) known as fibropapillomatosis. Fibropapillomatosis continues to spread geographically, with prevalence of the disease also growing at a number of affected sites globally. Environmental exposures seem key to inducing tumor development, possibly through weakening host immune systems to the point of enabling pathogen-induced tumor formation. However, we do not yet understand the precise molecular and mutational events driving fibropapillomatosis tumor formation and progression. Similarly, many open questions remain about the role of the herpesvirus (chelonid herpesvirus 5, ChHV5) associated with the disease as a potential co-trigger, and whether its occurrence within tumors is causative or

opportunistic. Without improved understanding of the basic biology of this disease epizootic, treatment, containment and mitigation options are severely hampered.

To address fundamental questions relating to the oncogenic signaling, mutational spectrum, viral load, viral transcriptional status (lytic or latent) and spread, we employed transcriptomic profiling, whole genome sequencing, immunohistochemistry and environmental (e)DNA-based monitoring of viral shedding. In particular we focused on the mutational landscape of tumors and assessing the transcriptional similarity of external (skin) and internal (visceral organs) tumors, and the oncogenic signaling events driving early stage tumor growth and post-surgical tumor regrowth. These analyses revealed that internal fibropapillomatosis tumors are molecularly distinct from the more common external tumors. However, our molecular analyses also revealed that there are a small number of conserved potentially therapeutically targetable molecular vulnerabilities in common between internal and external tumors, such as the MAPK, Wnt, TGFβ and TNF oncogenic signaling pathways. We also determined that the tumor genomes can harbor copy number gains, indicating potentially viral-independent oncogenic processes. Genes within such mutated genomic regions have known roles in human skin cancer, including MAPK-associated genes. Turtles attempt to mount an immune response, but in some animals this appears to be insufficient to prevent tumor development and growth. ChHV5 was transcriptionally latent in all tumor stages sequenced, including early stage and recurrent tumors. We also revealed that the tumors themselves are the primary source of viral shedding into the marine environment and, if they are surgically removed, the level of ChHV5 in the water column drops.

Together, these results offer an improved understanding of fibropapillomatosis tumorigenesis and provide insights into the origins, therapeutic treatment, and appropriate quarantine responses for this wildlife epizootic. Furthermore, they provide insights into human pathogen-induced cancers, particularly mechanisms which are difficult to study in the human and terrestrial context, such as time-course quantification-based monitoring of viral shedding.

Introduction

Sea turtle fibropapillomatosis (FP) is potentially a canary in the coalmine, indicating that continued human-induced environmental damage may be an alternative route by which oncogenicity is conferred on normally well-tolerated viruses. This is particularly worrying as long-lived reptiles have normally robust anti-cancer defenses[1,2], and as there are already a range of human viruses known to be capable of inducing tumor formation when the host immune system is compromised[3,4]. In addition to improving wildlife conservation medicine, fibropapillomatosis precision oncology can reveal the precise mechanisms through with environmental triggers, viral dynamics and host cell transformation can rapidly induce novel cancer incidence on an epidemic scale, thereby simultaneously informing human cancer research[2,5].

Fibropapillomatosis (Fig. 1A) is a tumor disease event of epizootic (animal epidemic) proportions, affecting wild populations of endangered marine turtles circum-globally[6-9].

Fibropapillomatosis leads to sea turtle fatalities by direct (affecting internal organs) and indirect means (blocking vision and feeding or restricting movement and immune functioning)[7,10]. This sea turtle tumor epizootic is particularly striking given that long-lived turtle species (over 80 years) tend to have more robust anti-cancer defenses than humans[1]. Green sea turtles (*Chelonia mydas*) are most affected by fibropapillomatosis, providing a clear species to focus on for experimental studies, but the disease also occurs in all other marine turtle species[11-14]. Sea turtle fibropapillomatosis continues to spread geographically[7,10] throughout equatorial and subequatorial regions[7,10,11,15,16]. Fibropapillomatosis has been reported in every major ocean basin in which green turtles are found, particularly in near-shore habitats[13,17](www.cabi.org/isc/datasheet/82638). In addition to spreading globally, fibropapillomatosis rates continue to increase in many affected sites, posing serious conservation challenges. Currently, of all green sea turtles stranding, over 40% in Florida, 30% in Hawaii, 35% in Texas, 35% in northeastern Brazil, 34% at Príncipe Island in the Gulf of Guinea and 50% in Puerto Rico are FP-afflicted [10,18-22]. Many of these sites have seen a rapid increase in disease incidence over recent years, for instance 13.3% in 2005 compared with 42% in 2016 in Florida, 13.2% in 2012 compared with 35.3% in 2015 in northeastern Brazil, and 0% in 2009 compared with 35.2% in 2018 in Texas (the occurrence of fibropapillomatosis in Texas was first reported in 2010 at a rate of 0.6%) [10,18-21].

A chelonian-specific herpesvirus (chelonid herpesvirus 5, ChHV5), has been implicated in driving this disease epizootic, although Koch's postulates to confirm its causative role have yet to be definitively confirmed [7,12,23-26]. Even so, ChHV5 infection alone is not sufficient to induce FP tumor growth. ChHV5 is globally distributed, including being present in sea turtle populations in which fibropapillomatosis tumors have never been reported[27]. An environmental co-trigger(s) seems to be the required key to both the development of fibropapillomatosis tumors and the geographic spread of the disease, with the environmental trigger correlating with anthropogenic activity[7,26,28,29]. Much uncertainty remains about the specific environmental trigger(s) and how they interact with ChHV5 and the host immune-system to give rise to FP tumorigenesis[7]. It is highly likely that environmental exposures induce immunosuppression in marine turtles enabling viral load to increase to the point of crossing an oncogenic threshold, similar to a number of human virally-induced cancers [3,7,26,30]. There is a paucity of knowledge concerning the molecular signaling events underpinning fibropapillomatosis tumor initiation, development and growth, with even less known about the viral and host transcriptional and mutational landscape driving FP tumorigenesis. Additionally, the precise relationship between internal and the more common external fibropapillomatosis remains to be elucidated.

Similarly, open questions remain regarding whether ChHV5 is lytic or latent in FP tumors and the occurrence of viral shedding and transmission[7,26,31-34]. Novel environmental DNA (eDNA) approaches can help resolve the dynamics of viral shedding. Environmental DNA is a forensics approach to the extraction and identification of organismal DNA fragments (genetic material) released into the environment and is a rapidly advancing, non-invasive approach capable of improving endangered species detection and early pathogen detection[35-43]. Environmental samples can be analyzed for micro- and macro-organisms by several

eDNA methods including metabarcoding and species-specific quantitative PCR (qPCR)[38,40,44]. The development of a rapid and high throughput sampling scheme to detect virus shedding into the marine environment would benefit pathogen surveillance efforts, and consequently the performance of wildlife health status monitoring could be improved[31,45]. Here we applied qPCR and whole genome sequencing/shotgun sequencing ("unbiased", nonbarcoded) eDNA approaches to quantify ChHV5 viral shedding from rehabilitating patients. Such novel approaches to environmental viral shedding are a particularly beneficial feature of aquatic models of virally-induced tumors, such as fibropapillomatosis, and will enable greatly improved understanding of the dynamic relationship between viral load and viral shedding, not readily measurable in terrestrial species.

While advances in our understanding of the fibropapillomatosis tumor disease epizootic in sea turtles continue to be made[7,26,27,30,46-53], many questions remain unanswered in relation to this enigmatic disease. There is virtually no molecular information about the relationship (e.g. primary/metastatic) between the numerous tumors (which can range from tens to hundreds of tumors per individual) arising on an individual. Similarly, the drivers of early stage, internal and post-surgical regrowth tumors remain to be elucidated. Determining the contribution of each facet of this multifactorial disease will be key to combatting this anthropogenic-implicated disease epidemic both at the level of individual clinical treatment and for population level management/mitigation strategies [2,30]. The relative contributions of ChHV5, environmental trigger(s), immune suppression, host genome mutation and altered gene expression to fibropapillomatosis oncogenesis have never been determined. Indeed, it is not known whether any mutational events occur within fibropapillomatosis tumor genomes. Here we applied a combination of extensive transcriptomics, genomics, precision oncology[30,54,55], immunohistochemistry and eDNA-based viral shedding monitoring to determine the host and viral molecular events underpinning fibropapillomatosis tumorigenesis, growth, post-surgical recurrence and transmission.

Results

Fibropapillomatosis tumors oncogenic signaling networks, immune response and therapeutic vulnerabilities

To determine the molecular events governing FP tumor growth, and whether different oncogenic signaling networks drive varying tumor types in different tissues we conducted transcriptomics of fibropapillomatosis tumors and patient matched non-tumor tissue (90 RNA-seq samples, Supplemental Table 1). Differentially expressed genes were then analyzed at the gene, pathway, network, and systems level. When the mRNA transcripts differentially expressed between external new growth, external established, and external post-surgical regrowth (recurring) fibropapillomatosis tumors were compared, there was a high degree of overlap (Fig. 1B). This suggests that the molecular events driving fibropapillomatosis tumor formation, growth and regrowth are broadly similar.

Next we compared the mRNA transcripts differentially expressed (DE) between established external tumors and those of internal visceral tumors, lung FP and kidney FP (Fig. 1C, Supplemental Table 2). In contrast to the three types of external tumors (new growth, established, re-growth, Fig. 1B), the internal visceral tumor DE transcripts were extremely divergent to the external tumor DE transcripts (Fig. 1C). Furthermore, lung and kidney tumors also showed minimal overlap with each other (Fig. 1C) suggesting different oncogenic signals are driving internal FP compared with external FP tumors, and also kidney FP tumors compared with lung FP tumors. The diverging transcriptional profiles of internal and external tumors can also be seen at the whole transcriptome level (Fig. 1D,E).

A large number of genes were differentially expressed between all tumor samples and all non-tumor samples (Fig. 2A). Highly upregulated genes included those associated with oncogenic signaling, such as CTHRC1 (Wnt and adenocarcinoma signaling) and CRABP2 (retinoic acid signaling), while highly downregulated where associated with skeletal muscle (AACTA1, MYL6B and KLHL41, Fig. 2A). To better understand the biological processes involved in each tumor type we examined the top 20 Gene Ontology (GO) terms (called by IPA, ranked by p-value) for each set of differentially expressed transcripts (Fig. 2B,C and Supplemental Fig.1A-C). Eighteen of the top 20 GO terms associated with established external FP tumors were cancer or neoplasia-associated terms, with 'activation of skin tumors' and 'melanoma' associated terms featuring prominently (Fig. 2B). The two non-cancer GO terms, 'formation of muscle' and 'muscle contraction', were inhibited, likely indicating an absence of muscle in FP tumors compared with non-tumored skin controls. For new growth external tumors, 14 of the top 20 GO terms were cancer-associated (Supplemental Fig.1A). These 14 GO terms were not as highly activated as in established tumors, likely due to the early stage at which these tumors were sequenced, and the paucity of early stage (sequenced human) tumors in the IPA knowledgebase compared with established tumors (the stage more frequently diagnosed and biopsied/sequenced in human samples). In terms of level of activation, the GO terms of regrowth external tumors were intermediate between new growth and established FP. Eleven of the top 20 regrowth GO terms were cancer-related, again with a number of muscle-related biological processes being inhibited (Supplemental Fig.1B).

Twelve of the kidney fibropapillomatosis's top 20 GO terms were cancer-associated, while all of the lung fibropapillomatosis's GO terms were cancer-associated (Fig. 2C,D). Unusually, almost all of the lung fibropapillomatosis's top 20 GO terms that had direction of activation/inhibition called were inhibited (Supplemental Fig.1C). This is in complete contrast to the external tumor types analyzed. A number of the kidney fibropapillomatosis's cancer-associated GO terms were also inhibited (Fig. 2C). This is surprising, as the internal FP's were not compared directly with external tumors, rather differential expression was called based on comparison with their patient matched non-tumor internal tissue samples (non-tumored lung and non-tumored kidney). To ensure no non-tumor sample related bias, we also directly called transcripts differentially expressed (DESeq2) between established external FP samples and internal lung FP or kidney FP samples (Supplemental Fig. 2A-C). This showed 11 cancer-associated GO terms (of the top 20 GO terms) to be activated in lung

FP (Supplemental Fig. 2C). Only four of the top 20 GO terms were inhibited cancer-associated terms (Supplemental Fig. 2C). This suggests that the inhibition of cancer-associated GO terms seen when lung FP was compared with non-tumor lung (Supplemental Fig.1C) may be due to non-tumor tissue specific expression variation, rather than cancer-associated processes being inhibited in lung FP tumors. As demonstrated by the PCA plot, the expression profiles of the internal non-tumor tissue samples differed widely from the skin and external FP tumor samples (Fig. 1E). Overall, of the top 20 GO terms from each tumor compared with their matched non-tumor tissue type, six were common to every FP type sequenced, although their activation/inhibition status differed between internal and external tumors, as did their extent of activation across the three external tumor types (Fig. 2D).

To further determine what signaling events are prominent in driving early external tumor formation (new growth and post-surgical regrowth), we examined the top 200 (ranked by p-value) GO terms in early external tumors compared with established external tumors (Fig. 3A). Similar to the gene level analysis there was a very high degree of overlap between the three external tumor types. However, of the 200 GO terms, 22 were uniquely common to early new and regrowth tumors (Fig. 3A). Of these 22 GO terms, eight of them were associated with Leukocyte/lymphatic processes (Fig. 3B), suggesting a crucial role of immune response during initiation of FP tumor growth. Immune activation may be in response to the ChHV5 virus, or to the tumor cells themselves. All eight leukocyte/lymphatic process GO terms were strongly activated across the three external FP types (Fig. 3B), although for established external tumors these terms fell outside of the top 200 GO terms. Interestingly, two of these eight terms (cell movement of leukocytes and leukocyte migration) were called as statistically significant in internal tumors (although outside of the top 200. Fig. 3B). In kidney FP tumors both terms were strongly activated, mirroring the external tumors. Conversely however, in lung FP tumors both of these GO terms were inhibited (Fig. 3B), again highlighting differences between the molecular signaling events driving internal and external FP tumors. Interferon gamma, a cytokine critical to both innate and adaptive immunity, was also called as an inferred transcriptional regulator (ITR) in all five tumor types, being activated in every tumor type barring lung tumors where it was inhibited (Supplemental Fig. 2D)

Given the strong leukocyte/lymphocyte infiltration results from the tumor transcriptomics, we next assessed the host adaptive immune response by identifying CD3 positive T lymphocytes[56] within fibropapillomatosis tumor tissue sections. This confirmed the transcriptomics findings, revealing that even in early stage FP tumors there is a high level of T lymphocyte infiltration (Fig. 3C). CD3 is an immunophenotypic cell marker, which is found only in T lymphocytes and is central to the formation of antigen-receptor interactions through the T cell receptor/CD3 complex [56,57]. CD3 positive staining was strongest in epidermal regions, where inclusion bodies (presumably due to lytic ChHV5) most commonly occur within FP tumors[12]. Together the transcriptomics and CD3 staining demonstrate that an immune response is mounted by the host (*C. mydas*), either to the tumor cells themselves, and/or more likely given the localization pattern to ChHV5 infection.

The disparate signaling events detected by the transcriptomics between external, lung and kidney tumors, potentially make it less likely that a single systemic anti-cancer therapeutic would prove effective against both external and internal tumors. However, to investigate whether any common therapeutically targetable oncogenic pathways exist between these tumor types we next compared their top 100 ITRs. IPA analysis infers the upstream transcriptional regulators responsible for the observed transcriptomic signatures by comparing the differential gene expression profiles to known regulator induced changes in its knowledgebase. Mirroring the gene level analysis, ITR analysis also showed very little overlap between the top 100 ITRs of established external, lung and kidney FP tumors (Fig. 4A). However, if a common therapeutically targetable vulnerability exists it should be located in the overlapping ITRs of these three FP tumor types. Therefore, we further investigated the 16 ITRs which all three FP types had in common (Fig. 4 A-C). These 16 ITRs represented nine genes and seven pharmaceutical compounds (drugs). Of the nine gene ITRs, almost all were activated across all five FP tumor types sequenced (external: established, new and regrowth, and internal: lung and kidney), with new growth external tumors tending to be the exception (Fig. 4C). These nine genes form a highly interconnected regulatory network (Fig. 4B), with 32 edges between the nodes and a protein-protein interaction enrichment p-value of 3.82e-07 (String). Interestingly, retinoic acid (RA) signaling was activated strongly in established external, external regrowth and kidney FP tumors (Fig. 4C). Retinoid therapy is widely used as an anti-cancer therapeutic and maintenance therapy for a number of human cancers (such as the pediatric cancer neuroblastoma[58]), although conversely RA signaling is known to be activated in some cancers. Retinoic acid is a widely available and inexpensive anti-cancer therapeutic, which would make it ideally suited for use in sea turtle rehabilitation facilities. Unfortunately the findings here suggest RA would be ineffective against FP. These transcriptomic findings explain the failure of early attempts to treat FP tumors and prevent regrowth using ectopic RA application (Supplemental Fig. 3A).

The 16 ITRs (Fig. 4C) tended to fall into three main categories, canonical Wnt signaling, MAPK signaling and immune-related signaling (CD44, IL6, TNF, TGFβ). These pathways form part of an interlinked signaling network (Fig. 4B). Amyloid precursor protein (APP) may also be immune-related, as in addition to being implicated in neuron synapses and Alzheimer's disease, it has also been implicated to be involved in antimicrobial activity and iron export. As predicted by the transcriptomics, β-catenin was located at the cellular membrane in new growth external FP tumors, indicating inactivation of Wnt/β-catenin signaling (Fig. 4D). Again in line with the ITR analysis, nuclear localization of β-catenin was present within other tumor types (Fig. 4D and Supplemental Fig. 4A), which is indicative of Wnt/β-catenin pathway activation. While nuclear β-catenin did occur in external established, regrowth, and internal tumors it was far from ubiquitous, suggesting intra-tumor heterogeneity in terms of Wnt signaling activation. Generally the transcriptomics revealed that each of the three main shared signaling pathways (Wnt signaling, MAPK signaling and immune-related signaling) were activated; inversely, but logically, the drug inhibitors of these pathways were inhibited (U0126, SB203580 and dexamethasone, Fig. 4C). This suggests that FP tumors (both external and internal) may be susceptible to treatment with inhibitors of these pathways.

Transcriptionally-inferred molecular origins of fibropapillomatosis

To gain further insights into FP's origins we employed larger scale network clustering analysis. Established tumors demonstrated clusters of 'viral and inflammatory responses', 'inhibition of anoikis' (programmed cell death of anchorage-dependent cells that detach from the extracellular matrix), 'cellular senescence' and 'miRNA regulation' (Fig. 5A). Interestingly, given FP's environmental trigger(s), established external tumor's highly interconnected ITR network (protein-protein interaction enrichment p-value <1.0E-16) also had signatures related to 'cellular responses to organic substances and chemical stimulus' (Supplemental Fig. 5A). Furthermore, pathways related to 'Kaposi sarcoma-associated herpes virus infection' (KEGG pathway analysis 1.20E-32) were also detected in established external FP (Supplemental Fig. 5A). Kaposi sarcoma is a human cancer which arises in immunocompromised patients, with lesions developing on skin, lymph nodes, or other organs, and is associated with a herpesvirus (human herpesvirus 8, HHV8). Like HHV8, the chelonid herpesvirus, ChHV5, is similarly postulated to drive FP tumor formation in sea turtles compromised by an external environmental or immunomodulatory trigger, a hypothesis which is strengthened by the viral and immune response signaling dynamics revealed by the transcriptomics. Immune-related processes featured in all five tumor types (Fig. 5A,B and Supplemental Fig. 6A-C). Although transcriptionally divergent to established external FP tumors, lung FP also showed an interconnected network (protein-protein interaction enrichment p-value <1.0E-16) with cellular immune response and organic substance and chemical response nodes (Supplemental Fig. 5B). Furthermore, clusters related to response to inorganic substances, metal ions, viruses and radiation were also detected (Fig. 5B). Indeed, 'Quantity of Metal' was called as a GO term and was activated for all tumor types, with the exception of kidney tumors in which it was not called (Supplemental Fig. 6D).

Pathway analysis (Fig. 5C) and GO term analysis (Fig. 5D) revealed a graded activation of metastatic-related signaling across the FP tumor types. FP tumors have been described as primary, despite the numerous tumors which regularly develop on each afflicted individual. However, to date no in-depth molecular analysis has been conducted to determine if all tumors on an individual are indeed primary, or if all or some of them (particularly internal visceral tumors) occur due to metastatic spread of a primary tumor. Our analysis suggests that the situation may not be so clear-cut, and therefore systematic phylogenetic/phylogenomic analysis of numerous tumors upon the same individual should be conducted. Our transcriptomics suggests that at a minimum, fibropapillomatosis tumors have a propensity to mutate towards the activation of metastatic pathways, with internal tumors showing stronger activation than external tumors (Fig. 5C,D). Early external FP tumors do not display metastatic signaling, rather such pathways are mildly inhibited (Fig. 5C,D). Established external FP show mild metastatic signaling activation, while internal tumors, particularly kidney tumors, show elevated activation of these pathways (Fig. 5C,D). It should be determined whether this activation is due to metastasis having occurred, or whether the propensity to metastatic activation falls short of complete metastasis.

The genomic mutational landscape of disseminated FP tumors

Genome instability and mutation (host) is a hallmark of cancer [59], DNA replication and repair machinery malfunction frequently during oncogenesis leading to cancer cells having aberrant genomes. Such genome instability and mutations can range from single base pair changes to whole chromosome gains and deletions [59]. Indeed, initial mutations within the genes controlling DNA replication and repair can be self-selecting leading to further cancer promoting genomic alterations. The mutational landscape of the host genomes of marine turtle fibropapillomatosis has yet to be explored, we therefore, next conducted whole genome sequencing (WGS, DNA-seq) of three fibropapillomatosis tumors and their patient-matched non-tumor tissue, to investigate whether they harbor genome instability in the form of copy number variations (CNVs), regions of the chromosome which have additional copies (gains) or which have been deleted (losses). Copy number variation analysis revealed that fibropapillomatosis tumors can harbor mutations, i.e. may not solely be driven by non-genomic events (e.g. viral, epigenetic or environmental). However, inter-tumor heterogeneity was observed, even between two tumors (one lung and one kidney) from the same individual (Figs 6A, Supplemental Fig. 7A). Copy number gains were present in the lung tumor (Fig. 6A), while no copy number variations were observed in the other two tumors (Supplemental Fig. 7A). The heterogeneous CNV landscape of the lung and kidney tumor from the same individual (patient 27-2017-Cm), suggests that these two tumors are not clonal (to each other) and did not arise as metastatic spread from the same original primary tumor. Although, it remains a possibility that the kidney tumor metastasized to the lung, only gaining CNVs after leaving the kidney.

Next the CNVs occurring in Patient 27-2017-Cm's lung tumor were mapped to the gene level revealing 228 genes (174 of which were recognized by IPA [human filter]) with copy number gains. When this CNV gene list was subjected to IPA analysis it was highly enriched for cancer-associated genes, with non-melanoma solid tumor (a human skin cancer) being the most enriched pathway (p-value 1.14E-10). This indicates that these CNVs may be involved in driving FP tumorigenesis and growth/progression (Fig. 6B), rather than solely being passenger mutations. Furthermore, 22 Wnt/β-catenin-related and 14 MAPK associated genes harbored copy number gains (Fig. 6C), with p38 MAPK signaling, ultraviolet radiation (UV)-induced MAPK signaling, MAP2K1, MAP3K4 and MAP3K8 called as pathways and ITRs of the CNV harboring gene list (IPA). That UV-induced MAPK signaling (particularly the SMPD3 gene) was called is particularly interesting as UV exposure is a putative environmental trigger of FP [26,30]. Additionally, there were epigenetic-related genes harboring copy number gains in the lung tumor, such as BRD7 which is involved in chromatin remodeling complexes. Genes harboring copy number gains in Patient 27-2017-Cm's lung FP genome had corresponding expression gains when dosage analysis was performed using the corresponding RNA-seq data for this tumor (Fig. 6D, Supplemental Table 3), confirming that these mutational gains induced changes in gene expression. These mutated genes are also involved in the regulation of cell proliferation by regulating the mitotic cell cycle regulation and DNA replication checkpoint (Supplemental Figure 7B), dysregulation of these cycles is a key hallmark of cancer cells [59].

ChHV5 is latent in FP tumors, including early stage and internal tumors

While further investigation of the genomic and environmental drivers of FP is warranted, the suspected causal relationship between ChHV5 and FP also requires further study. We demonstrated previously that ChHV5 was transcriptionally latent in established FP tumors [26]. However, that study only examined a small number of established external tumors (seven) and did not investigate whether ChHV5 was lytic or latent in early stage, regrowth or internal tumors. We postulated that ChHV5 might be latent in established tumors, but more active during crucial early stage tumor initiation events, akin to the 'hit and run' hypothesis of viral oncogenesis [26]. To investigate this hypothesis, we aligned the reads from the 90 RNA-seq samples above to the ChHV5 genome.

Across all FP samples, ChHV5 transcripts were low, and we detected no major switch to active (lytic) virus in either new growth FP or post-surgical regrowth FP (Fig. 7A). In fact, levels of ChHV5 transcripts were only marginally higher in FP tumors than they were in non-tumored tissue controls (Fig. 7A), though this was not significant for any of the FP tumor types compared with the non-tumor tissues. The only significant difference found in the level of viral RNA transcripts between any of the groups was between regrowth and established growth external tumors (Kruskal Wallis with Dunn-Bonferroni post hoc, $p = 0.013$). Furthermore, we detected no switch to active (lytic) viral transcription in internal tumors (Fig. 7A). Together this suggests that the role of lytic ChHV5 in FP is minimal and that if ChHV5 is contributing to FP oncogenesis, it is either occurring transiently, during very early tumorigenesis (before visible tumor appears), or through the expression of ChHV5 miRNAs (not assessed here), or that it is the latently expressed ChHV5 transcripts that are driving oncogenesis. We therefore next examined the individual gene level transcripts to determine which ChHV5 genes were transcriptionally active in FP tumors (albeit at relatively low levels). Samples were grouped into four types: non-tumor, external FP (including established, new growth and regrowth), kidney FP, and lung FP. Across all four sample types quite a consistent pattern of ChHV5 gene expression emerged (Fig. 7B, Supplemental Fig. 8A), with only 22 of ChHV5's 104 genes showing levels of expression above 1 transcripts per kilobase million (TPM, Supplemental Table 2). Latency-associated genes such as F-LANA formed part of this 22 gene group, being consistently expressed across all tumor types (Figs 7B and Supplemental 8A). Given the paucity of lytic ChHV5, these 22 ChHV5 genes represent the most likely viral drivers able to contribute to FP initiation and ongoing tumor development and growth; therefore they warrant further functional investigation.

Almost all samples with above 200 ChHV5 viral reads per 10 million total reads originated from just three patients (Supplemental Fig. 9A). Interestingly all three of these higher viral transcript patients were successfully rehabilitated and released. This prompted us to investigate whether there was any relationship between number of viral transcripts and rehabilitation outcome. Counter-intuitively, patients with positive outcome (release) on average had samples with statistically significantly higher ChHV5 transcripts (Mann-Whitney U Test, $p = 0.03$), while those patients that died in care or were euthanized due to advanced disease actually had lower ChHV5 transcripts (Supplemental Fig. 9B). Even when internal tissue samples were removed from the analysis (as all internal samples originated

only from the deceased/euthanized category), there remained a significant difference between the two groups (Mann-Whitney U Test, $p = 0.0143$).

High ChHV5 viral loads, predominantly arise from latent virus

Given the low level of ChHV5 transcripts in FP tumors, we next used whole genome sequencing (WGS/DNA-seq) to determine the viral load of ChHV5 to determine whether the low number of transcripts arises due to a lack of virus, or whether large quantities of virus are present within FP, but that the majority of these are latent (not undergoing active viral replication and transcription). By not conducting any viral enrichment steps, the resulting read numbers give a more reliable indication of the relative abundance of viral DNA compared with host DNA (all within a single sample/library). Viral DNA sequencing reads in the FP tumors covered a broad range. Interestingly, the quantity of viral DNA reads in FP tumors was not significantly different when compared to any other group, with the exception of the plasma samples (Fig. 7C). The lung and kidney tumor DNA-seq samples (also used for host genome CNV analysis) from the same Patient 27-2017-Cm had dramatically different viral loads (Table 1). The early growth FP tumor from 'Lilac' (Patient 25-2018-Cm) had an intermediate viral load 1,036 ChHV5 per 10 million total reads (RPTMT, Table 1 and Fig. 7C). Despite the high level of ChHV5 DNA sequencing reads in the lung tumor (3,673 RPTMT), this same tumor only had a very low level of ChHV5 RNA sequencing reads (68 RPTMT), more closely resembling the read numbers of the non-tumor tissues (Table 1). This suggests that rapid proliferation of host tumor cells, infected with latent virus is the primary driver of the high viral loads sometimes observed in FP tumors (some FP tumors can double in size in less than two weeks[46]).

The 27-2017-Cm kidney tumor had minimal host CNVs and minimal viral load (Table 1 and Fig 7C) suggesting that the tumor must be driven by other oncogenic mechanisms such as point mutations, epigenetic changes, or transcriptional/translational aberrations. The viral load of this kidney tumor was within the range of non-tumored tissue, plasma from FP-afflicted turtles and tissue from non-tumored hatchlings (Table 1 and Fig. 7C).

When DNA-based viral reads (Fig. 7C) are compared with viral reads from the RNA transcripts (Fig. 7A) of the 90 RNA-seq samples, the highest viral transcript load observed was approximately 600 RPTMT (one FP and one non-tumor sample), while the highest viral DNA load observed was 3,673 RPTMT. Although a high viral DNA load did not equate to high viral transcription (RNA) in the same sample (Table 1). As expected, WGS reads were dispersed across the entire ChHV5 genome (Supplemental Fig. 10A), not being restricted to ChHV5's transcriptionally active regions (Supplemental Fig. 8A). This confirms that the limited transcriptional signature is genuine and not due to a sequencing artifact. Together the whole ChHV5 genome-level and gene-level TPM analysis highlights the marked difference in reads between viral DNA presence and viral RNA transcription in FP tumors. Conversely, the range of viral DNA and viral RNA within non-tumored tissue was largely overlapping suggesting that the ChHV5 present in non-tumored samples may be more likely to be transcriptionally active (Table 1 and Figs 7A, 7C).

We next examined the ChHV5 viral loads (viral DNA-based qPCR, ChHV5 UL 30 assay) in a wider cohort of internal tumor and matched non-tumor types. Lung and bladder FP tumors consistently had the highest viral loads (Fig. 7D), while of the non-tumor tissue types assayed, bladder tissue also showed some of the highest viral loads, with detected levels being higher than in some of the FP tumors (Fig. 7D). These results have implications for the spread of ChHV5, as urine is a potential source of ChHV5 transmission [32,33].

ChHV5 may be vertically transmitted from mother to offspring

Interestingly, the range of ChHV5 detected in hatchlings (one *C. mydas*, one loggerhead and one leatherback) overlapped the range of ChHV5 in FP-afflicted non-tumor tissue, the FP kidney tumor and FP-afflicted blood plasma samples (Fig. 7C). Of the three species, the leatherback sample had the highest number of ChHV5 reads. If hatchlings are already infected with ChHV5 (vertical transmission), this has serious implications for any potential population-level vaccination based mitigation strategies. On beach (immediate nest emergence[60]) sampling and sequencing-based ChHV5 detection should be conducted to confirm this finding and to determine the prevalence of ChHV5 infected hatchlings.

Detection of ChHV5 shedding into the water column by environmental DNA (eDNA) approaches, reveals novel shedding dynamics and tumor burden correlations

A number of potential routes for ChHV5 transmission have been postulated, including via vectors such as *Ozobranchus* leeches[34], which are commonly found on fibropapillomatosis-afflicted turtles, often at high density within the crevices of external fibropapillomatosis tumors (Fig. 8A). In line with previous studies[34], we confirmed that ChHV5 could specifically be detected from DNA extracted from leeches which had fed on FP-afflicted turtles (Fig. 8A). Another potential mode of transmission is direct ChHV5 shedding into the environment. Direct viral shedding as a route of transmission is thus far only supported by indirect evidence, i.e. cloacal swabs, urine, ocular, oral and nasal secretions[32,33,61], viral inclusion bodies near the surface of external tumors [31] and the elevated levels of ChHV5 detected in bladder tissue (Fig. 7D). Direct detection of ChHV5 in the water column is lacking. Therefore, we employed environmental DNA (eDNA)-based approaches coupled with qPCR and next-generation sequencing to detect ChHV5 in patient tank water. Such novel approaches can aid in answering previously intractable questions about direct ChHV5 transmission such as presence, abundance, and persistence in the marine environment. First, we determined that the presence of ChHV5 could be readily detected in eDNA extracted from tank sea water, using either an established[25] UL30 qPCR assay or next-generation sequencing (Figs 7C, 8B-E). Furthermore, not only was ChHV5 detectible, it was quantifiable allowing us to make comparisons between individual tanks and across time (Fig. 8A-D). The level of detectible virus in patient tank water was positively correlated (R^2 = 0.5431, p = 0.0002, df = 19) to the tumor burden (as assessed by total tumor surface area) of the patient(s) housed in that tank (Fig. 8A). Patients with large well-established tumors shed more virus into tank water than those with small new-growth tumors. As FP tumors were surgically removed, so did the level of ChHV5 in patient tanks drop (Fig. 8C,D), suggesting that the tumors are the primary source of environmental ChHV5 (either through direct tumor shedding, or migration of virus throughout the body and excretion in bodily fluids). Additionally, the quantity of

ChHV5 eDNA in tank water was positively correlated with the *C. mydas* (green turtle) eDNA level (R^2 = 0.66, p = 0.00001, df = 19), as detected by a custom *C. mydas* 16srRNA DNA assay (Fig. 8E).

Discussion

Transcriptional and mutational origins and therapeutic vulnerabilities of fibropapillomatosis

Wildlife pathogens have been shown to exacerbate the effects of environmental degradation, habitat loss and the climate emergency on population levels, potentially leading to local and global extinctions[2,62-65]. Furthermore, as the risk of extinction increases for a given species the detrimental effect of disease on the population worsens[66]. Anthropogenic activities are not only stressing habitats, but the rapid environmental changes induced by these activities are likely increasing cancer rates in wild populations[67]. Human-induced perturbations of inshore marine environments have also been implicated as a co-trigger of the fibropapillomatosis tumor epizootic in green sea turtles[7,29]. It is thought that environmental changes are key to conferring oncogenicity upon ChHV5, potentially through compromising or modulating the turtle's ability to respond to the viral infection. We have demonstrated activation of immune-related signaling in FP tumors and shown localized CD3+ T-cell infiltration within new growth tumors. That the host still mounts an immune response within tumors is interesting given the previous links between FP-afflicted turtles and immunosuppression [7,11,68,69]. However, varying studies have found contradictory immunosuppression findings between FP and non-FP-afflicted individuals, for example no differences were observed in leukocyte activity between *C. mydas* turtles with and without FP, when blood samples were tested via flow cytometry[7,10,11,70,71]. Furthermore, not all turtles with FP are immunosuppressed and it is currently not clear whether immunosuppression in FP-afflicted turtles, when it does occur, is a cause or consequence of the disease [7,10,11,71]. Our transcriptomics and immunohistochemistry show that lymphocytes do infiltrate and mount an immune response within FP tumors. It will be important to elucidate why such a response is not sufficient at preventing initial tumorigenesis or preventing continued tumor growth. It should also be determined in the cases where spontaneous tumor regression does occasionally occur in non-heavily afflicted FP individuals whether regression is a consequence of elevated lymphocyte infiltration and activation[7,10,12,72]. Further elucidation of the immune status of FP-afflicted individuals and the temporal dynamics and extent of immune responses will be crucial to determining whether the environmental co-trigger(s) of FP promote tumorigenesis through immunosuppression, be it systemic or localized.

Currently, no treatments exist for internal tumors with such turtles having to be euthanized regardless of their health status otherwise. Furthermore, surgical excision of FP tumors often results in high rates of tumor regrowth/recurrence[11,26,46]. Therefore, it is imperative that chemotherapeutic approaches are developed to augment surgical removal of external tumors and to provide first-line therapy for applicable internal tumors. We show that the molecular drivers of external and internal FP tumors are largely transcriptionally distinct. However,

transcriptomic profiling and whole genome sequencing (CNV analysis) of FP tumors revealed that MAPK, Wnt and TGFβ pathway inhibitors are putative therapies for both external and internal FP, which could be targeted to simultaneously reduce tumor burdens (e.g. systemic targeting via oral dosing).

Although kinase inhibitors have been used extensively in human cancers, their application in veterinary medicine in mostly restricted to dogs and cats and only incidentally in other animals[73]. While Wnt signaling is heavily implicated in a wide range of human cancers, no clinically approved therapeutic yet exists, although clinical trials are ongoing[74,75]. Similarly, TGFβ targeting therapeutics are undergoing clinical trials[76]. The MAPK/ERK pathway is involved in cell proliferation and survival and is implicated in carcinogenesis[77]. As such, a number of MAPK pathway drugs have been developed, such as MEK and BRAF inhibitors which are most commonly used to treat melanoma. MAPK/ERK inhibitors have greatly improved the life expectancy of patients with malignant melanoma, but acquired resistance almost inevitably occurs. Additionally, kinase inhibitors may cause serious adverse events for therapy recipients, with up to 46% of patients reported with grade 3-4 serious adverse events[78]. Worrying aspect of adverse effects with BRAF/MEK inhibitors is the occurrence of skin toxicities that also include secondary malignancies, such as keratoacanthoma and cutaneous squamous cell carcinoma[79]. RAF inhibitors may also cooperate with human papilloma virus to promote the initiation of cutaneous tumors[80,81]. Given the complexities of patient responses to MAPK targeting therapeutics, other common pathways between internal and external tumors could be prioritized, or therapeutics selected to separately target external and internal FP tumors.

The novel findings of this study offer somewhat conflicting evidence about the nature of the relationship between internal and external tumors (i.e. whether they arise from primary or metastatic processes). Whole transcriptome profiling indicated that internal tumors are more closely-related to the tissue type in which they are found (presumptive tissue of origin), suggesting that they arise from the *de novo* transformation of these tissues and not as disseminated metastases from external tumors. However, skin cancer-related pathways/GO terms were called for internal tumors both from the differentially expressed genes (RNA-seq) and from the genes harboring copy number variation at the DNA level. Additionally, there was a trend towards increasing activation of metastasis-associated signaling particularly for kidney tumors. If ChHV5 is responsible for inducing *de novo* tumor formation in internal organs, then the limited expression and load of ChHV5 in some internal tumors is highly irregular. Similarly, the paucity of elevated levels of viral transcripts across all tumor samples suggests that lytic virus may not be driving the establishment of numerous primary tumors within the same individual. If ChHV5 truly is essential to the initiation of FP tumors and not just an opportunistic pathogen, it may be acting only through latently expressed genes (or microRNAs), or there may be an extremely limited temporal window during tumor initiation (tumorigenesis) in which lytic virus is more abundant (i.e. the viral hit and run hypothesis[82]), something we found no evidence for in any of the early stage tumors that were sequenced.

It is possible that a more complex situation exists with some internal tumors arising *as de novo* tumors while others within the same individual are due to metastasis. Given the

presence of mutations within FP tumor genomes and their heterogeneity revealed here, in depth phylogenomics (clone history reconstruction) should be highly feasible, and its application could establish the clonal relationship between multiple tumors within the same individuals, including internal tumors. This would definitively address whether any FP tumor types arise as a consequence of metastatic spread. Given the heightened propensity of kidney tumors towards metastatic signaling pathway activation, an intriguing hypothesis would be that lung tissue receives sufficient air contact to enable lytic viral activation and *de novo* tumorigenesis, whereas kidney tissue does not. It was shown that in laboratory conditions ChHV5 would only undergo lytic reproduction within infected host cells when in direct contact with air [48]. Supporting this, our whole genome sequencing and qPCR revealed that lung tumors tended to have much higher viral loads (DNA-based assessment) than kidney tumors (Fig. 7C,D), although there was no significant difference between the number of viral transcripts (RNA-seq) from lung and kidney tumors (Fig. 7A). Expanding similar genomics and transcriptomics approaches to tumors of other internal organ types would help elucidate whether there is a relationship between the level of the tissue's air exposure and viral load or lytic virus level.

The role of ChHV5 in tumor progression

The propensity of ChHV5 to only reproduce lytically (under laboratory settings) when infected host cells were in direct contact with air[31], may also account for the localization of inclusion bodies within the epidermis of FP tumors[12]. It is possible that this feature could be an advantageous shedding strategy, as air contacting tissues are predominantly the skin surface and lungs, where shed virus would have a higher likelihood of successfully dispersing to other individuals. In line with this, we have shown that the level of ChHV5 in the water column positively correlates with level of external tumor surface area of patients.

Interestingly, when assessed in relation to patient survival and release, the presence of higher levels of ChHV5 RNA transcripts was actually prognostic for successful outcomes (Supplemental Fig. 9B). Patients having fewer viral transcripts were more likely to die or be euthanized during the course of their rehabilitation. Currently, due to the lack of treatment options (surgery is hampered by *C. mydas's* hard plastron and carapace) any turtles found to be afflicted with internal tumors in Florida's rehabilitation facilities are normally humanely euthanized[18]. Persistent aggressive post-surgical regrowth can also lead to the decision to euthanize a FP-afflicted patient. If ChHV5 is acting as an opportunistic pathogen in FP one might expect higher viral transcription in poor outcome patients, yet if ChHV5 is a driver of FP tumors it would also be expected that poor outcome patients would have higher levels of viral transcription (either lytic or latent associated gene transcription). Neither of these was the case for the patients whose tumors we sequenced. It would be informative to establish whether a similar relationship exists for viral load (DNA-based assessment) and outcome, or whether it is unique to viral transcription.

We reveal here that tumor-specific genome level mutations do occur within FP tumors, demonstrating that inter-tumor heterogeneity exists between two internal tumors from the same patient. Tumors harbored copy number gains in a variety of gene types and these were strongly enriched for cancer-associated functions. The mutational landscape of

fibropapillomatosis, has important implications for understanding the disease's molecular drivers and identifying potential therapeutic targets[2,54], and the identification of its environmental drivers (particularly single nucleotide polymorphism mutations)[83,84]. Furthermore, better understanding of FP's mutational landscape can help elucidate inter- and intra-tumor heterogeneity, the relationship between tumors within the same individual, and reveal the human cancer types which FP's mutational spectra most resembles, informing the disease's etiology and future treatment strategies[30]. Expanded whole genome-based mutational analysis may help explain how host cells can evolve and or rewire to become tumorigenic in the absence of high lytic virus loads, observed here.

Our findings demonstrate that latent ChHV5 predominantly accounts for the viral loads detected in FP tumors. Our genomic results show that the majority of ChHV5 replication in external and internal FP tumors is most likely driven through passive latent viral replication mechanisms, with the viral genomes being copied as the host tumor cells divide. Fibropapillomatosis tumors have been shown to grow rapidly [46]. Our genomic analysis is consistent with the relative lack of lytic virus found within FP tumors by histological and single gene approaches[31,34].

Mechanisms of viral exposure
ChHV5 DNA was identified in *C. mydas*, loggerhead (*Carretta carretta*), and leatherback (*Dermochelys coriacea*) hatchlings, and their abundance overlapped the range (no significant difference) seen in FP-afflicted non-tumor tissue, an FP kidney tumor and FP-afflicted blood plasma samples. Although it should be noted that ChHV5 genome coverage was not as complete in these hatchling samples as for tumor samples (data not shown), a key objective should be to confirm whether vertical transmission of the virus occurs from mother to offspring. This finding conflicts with the predominant hypothesis that ChHV5 infection is acquired only after juvenile turtles recruit to nearshore habitats[53,85]. The occurrence of *in ovo* vertical transmission would have serious implications for the consideration of any potential future population-level vaccination-based mitigation strategies. On beach (immediate nest emergence[60]), or pre-hatching sampling and sequencing-based ChHV5 detection should be conducted to confirm this finding and to determine the prevalence of ChHV5 infected hatchlings.

Accurate detection and monitoring of wildlife pathogens with the capacity to impact species survival is essential to devise and implement appropriate mitigation policies[45]. Pathogenic viruses have the ability to infect susceptible species without presenting obvious symptoms[45]. Consequently, accurate, sensitive, non-invasive techniques are required to detect and monitor pathogens and pathogenic vectors outside of host systems[45,86]. eDNA approaches have been shown to detect aquatic pathogens earlier than traditional methods and provide advanced warning of infection and mass mortality events[40,45]. Recently, *Ranaviruses*, have been detected by eDNA analysis in environmental water samples from susceptible host species habitats, and demonstrated a strong relationship between environmentally-shed viral load and host tissue viral load[45]. Our application of eDNA approaches to detect ChHV5 in sea water is particularly significant given that unlike more immediate acting pathogens, there is generally a long lag time (years or decades) between infection and tumor formation[87-89] for pathogen-

induced cancers. This delay can hamper our understanding of infection dynamics and initiating events, making the ability for ongoing monitoring and surveillance critical.

ChHV5 can infect endangered sea turtle species without presenting signs of infection until environmental co-factors trigger transformation and oncogenesis; therefore diagnostic techniques capable of detecting the pathogen outside hosts is vital[33]. It has previously been asserted that only a small percentage of tumors shed virus (7% of tumors in 35% of individuals)[31]. However, our eDNA-based monitoring of viral shedding demonstrated that ChHV5 could be detected even in the tanks of turtles with low tumor burdens and that ChHV5 shedding positively correlated to tumor burden. This suggests that previous attempts to estimate shedding using localized approaches such as number of epidermal intranuclear inclusion bodies in tumor sections may underestimate the level of viral shedding occurring, with in-water ChHV5 quantification having improved detection limits than previous approaches. The future use of sensitive, novel molecular technologies such as eDNA may allow for the early detection of ChHV5 presence in the environment of vulnerable populations, and enable further research into the etiology, host species transmission and disease ecology of FP.

Our ChHV5 eDNA findings likely represent the first instance of time-course environmental DNA quantification of marine pathogen release from individual patients. This unique tool to dynamically track environmental viral shedding in response to disease progression and clinical interventions such as surgery and drug treatment will greatly enhance our ability to address fundamental questions relating to this disease, as well as to design evidence-based rational management strategies (e.g. containment/isolation policies). Such approaches can also improve our understanding of how environmental and clinical factors influence viral release and spread. The ability to perform eDNA-based quantitative viral shedding monitoring will also further enhance the utility of fibropapillomatosis as a model not just to better understand this wildlife epizootic, but also to address unresolved questions relating to viral shedding for other animal and human pathogen-induced cancers.

Summary
Taken together, our results provide genome-level evidence for the complex relationship between external, new growth, established, post-surgical regrowth, and internal visceral tumors. They reveal the paucity of lytic ChHV5 replication within tumors, even those harboring high viral loads, as well as the heterogeneous mutational landscape of FP genomes and the host immune responses mounted within FP tumors. The application of precision oncology, genomic approaches, and novel technologies such as eDNA-based pathogen monitoring can assist in determining the molecular events underpinning FP tumor development, viral shedding and transmission dynamics, and enable the rational design of novel therapeutic interventions (such as pharmacological disruption of MAPK and TGFβ oncogenic signaling) and clinical management strategies. Importantly, the adoption of such approaches can elucidate the specific triggers of FP and the precise mechanisms through which these viral and environmental triggers are driving the fibropapillomatosis disease epizootic in marine turtles.

Materials and Methods

Tissue sampling

Sampling was carried out under permit number MTP-20–236 from the Florida Fish and Wildlife Conservation Commission and with the ethical approval from the University of Florida's Institutional Animal Care and Use Committee (IACUC). External fibropapillomatosis (FP) tumors were surgically removed and punch biopsies taken of non-tumored areas as previously described [26]. Internal tissue samples (tumors and non-tumor tissue) were obtained from animals during necropsies conducted immediately after euthanasia. Note that no animal was euthanized for the purposes of this study, but rehabilitating sea turtles determined to harbor internal tumors are currently euthanized in Florida as no treatment yet exists for internal tumors, and additional complications arising from surgery and other health concerns sometimes necessitates the humane euthanization of marine turtles in rehabilitation. Internal tissue samples were treated the same as the external samples. All samples were obtained from juvenile *C. mydas*, as this stage is the most commonly afflicted by the disease. Sex is not readily determinable in juveniles, but was provided for individuals that were euthanized due to internal tumors or other complications and in which necropsies were performed, or for individuals that were endoscoped (KARL STORZ, Multi-Purpose Rigid Endoscope for small animals) as part of their rehabilitative care (see Supplemental Table 1). Samples were stored until extraction in RNA-later (Qiagen) at -80°C, according to manufacturer's instructions.

RNA and DNA Isolation, library preparation and sequencing from tissue samples

For RNA-Seq samples, total RNA was extracted using either an RNeasy Fibrous Tissue kit (Qiagen, Cat No. 74704) or RNeasy Plus kit (Qiagen, Cat No. 74134) with column-based genomic DNA removal, according to manufacturer's instructions. Ninety RNA samples, comprising 70 fibropapillomatosis tumor samples and 20 non-tumor samples from 12 juvenile green turtles which had stranded in Northern Florida, were used for sequencing. Samples were further categorized by tissue type, as well as growth profile for the external tumors only (see Supplemental Table 1). Sequencing libraries were generated from 500 ng of total RNA using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Cat No. E7530), including polyA selection, according to manufacturer's protocol. Size and purity of the libraries were analyzed on a Bioanalyzer High Sensitivity DNA chip (Agilent). The RNA samples used for library construction had a RIN value range of 7.2 to 9.8, with the median RIN value of all samples being 9.1. Libraries were sequenced as paired end reads with a read length of 100 bp on a HiSeq 3000 (Illumina). ERCC Spike-In Mix (ThermoFisher) was used as an internal control: 2 µL of 1:400 diluted ERCC Spike-In Mix with 500 ng of total RNA input.

For DNA-Seq samples, DNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen, Cat No. 69504). Libraries were generated using Illumina TruSeq DNA PCR-Free Library Prep kit including fragmentation with a Covaris S220 sonic disruptor. Size and purity of the libraries were analyzed on a Bioanalyzer High Sensitivity DNA chip (Agilent). Libraries were

sequenced as paired end reads with a read length of 100 bp on an Illumina HiSeq 3000. Six whole genomic DNA samples, comprising three fibropapillomatosis tumor samples and patient-matched non-tumor samples from two juvenile green turtles which had also stranded in Northern Florida, were used for sequencing. These samples were further categorized by tissue type, with one tumor and patient-matched healthy tissue sample each coming from an external, kidney, and lung tissue source (see Supplemental Table 1).

In addition, whole genomic DNA was extracted from 10 plasma samples taken from six individual turtles during the course of their rehabilitation (Supplemental Table 1). Excess blood was taken from blood samples drawn for routine medical care by our veterinarians and plasma was separated from the red blood cells by centrifugation. Only 100 µl of plasma was collected and stored as allowed by permit number MTP-20–236 from the Florida Fish and Wildlife Conservation Commission. DNA was extracted from 60 μl of plasma using a DNeasy Blood & Tissue Kit (Qiagen, Cat No. 69504) and used to produce low input Illumina Fragment libraries using a Low Input Library Prep kit v2 (Clontech Laboratories, Inc., Catalog No. 634899). DNA was fragmented using the Covaris S220 sonic disruptor and libraries were then sequenced as paired end reads with a read length of 100bp on an Illumina HiSeq 3000.

Furthermore, three whole genomic DNA samples were also taken from ground flipper tissue samples from three deceased hatchling turtles, each separate species: a green (*C. mydas*), loggerhead (*Caretta caretta*), and leatherback turtle (*Dermochelys coriacea*), and processed following the methods for tissue as detailed above. The leatherback sequence length was the only sample that differed, with paired end reads of length 150 bp instead of 100 bp.

Finally, one sample of environmental DNA (eDNA) taken from holding tank water from the Whitney Laboratory Sea Turtle Hospital facility of the University of Florida was also used for sequencing. For eDNA sampling, seawater from 5 tanks (4 housing juvenile green sea turtles and 1 housing loggerhead post-hatchling washbacks, 500 ml seawater per tank) was filtered (EMD Millipore PES 0.22um sterivex filter) and DNA was extracted from the filter using a Qiagen DNeasy Blood & Tissue Kit (Qiagen, Cat No. 69504). Libraries were generated using a NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Cat No. E7645), including fragmentation with a Covaris S220 sonic disruptor. Fragment size and purity of the libraries were analyzed on a Bioanalyzer High Sensitivity DNA chip (Agilent). Libraries were sequenced as paired end reads with a read length of 100 bp on an Illumina HiSeq 3000.

Quality control and read trimming

The software FastQC - https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ - was used to assess data quality. Reads were then trimmed with trim_galore (The Babraham Institute, version 0.5.0) to remove ends with a Phred quality score less than 30, to remove adaptor sequences, and to remove sequences fewer than 25 bp after trimming. For any samples that contained over-represented sequences according to FastQC, the trimmomatic tool [90] (version 0.36) was then used to remove these sequences from reads and any sequences

less than 25 bp after trimming. The number of raw reads per sample and reads remaining after trimming can be found in Supplemental Table 1.

Read alignment and read counts

Reads from all samples (RNA-seq, DNA-seq, and eDNA) were first aligned to the ChHV5 genome [GenBank accession number: HQ878327.2][91] to determine the level of viral RNA and DNA present in each sample using bowtie2[92](version 2.3.5.1). The overall alignment rate to the ChHV5 genome was low for both RNA-seq and DNA-seq samples, with < 0.001% of reads aligning on average for all samples (Supplemental Table 1). Reads for all RNA-seq samples were then aligned to the draft genome for *C. mydas* [GenBank assembly accession number: GCA_000344595.1][93] using hisat2[94] (version 2.0.4). The overall alignment rate to the green turtle genome for RNA-seq samples was $82 \pm 7\%$ (mean $\pm$ SD) (Supplemental Table 1). One sample, an external established growth fibropapillomatosis tumor, had an extremely low alignment rate of 26% to the green turtle genome, and was therefore removed from further analysis. For DNA-seq samples, only tissue, plasma, and the eDNA sample were aligned to the green turtle genome. The overall alignment rates were $91 \pm 0.005\%$, $88 \pm 0.02\%$, and 0.31% (mean $\pm$ SD) for tissue samples, plasma samples, and the single eDNA sample, respectively.

Transcript abundance for both ChHV5 virus specific and *C. mydas* specific transcripts was generated using htseq-count[95] (version 0.6.1p1) with the following parameters: not strand-specific, feature type 'gene', and union mode for *C. mydas* specific transcripts, and not strand-specific, feature type 'gene', intersection non-empty mode, and a minimum aQual of 0 for ChHV5 virus specific transcripts. Count tables for viral and turtle transcripts were merged for all RNA-seq samples and counts were normalized for gene length and sequencing depth by transcripts per million (TPM) (Supplemental Table 2).

Differential expression analysis

Prior to differential expression analysis, the raw counts were processed with the RUVseq Bioconductor package [96] (version 0.99.1) using the RUVs method to remove low abundance genes, normalize the RNA-seq data, and remove unwanted variation among replicates. PCA plots (see Fig. 1a) were generated using the PtR script in the Trinity toolkit (Haas et al. 2013) both before and after RUVseq normalization. The RUVseq-processed matrix was then used to identify differentially expressed (DE) transcripts using the run_DE_analysis.pl script for the DESeq2 Bioconductor package (Love et al. 2014) and available through the Trinity toolkit (Haas et al. 2013). The run_DE_analysis.pl script was adjusted to also filter out low abundance genes by removing genes with a mean count ≤ 10 across all samples prior to differential expression analysis. The resulting lists of DE genes were sorted and filtered to include only those transcripts with an adjusted p-value of < 0.05 and a $\log_2$ fold change of > 2 or < -2. A list of upregulated and downregulated transcripts that overlapped from different sample types was generated and used to create area-proportional Venn diagrams of overlap using BioVenn [97].

Pathway analysis and annotation

Gene lists were analyzed for overrepresented pathways, biological functions, and upstream regulators using Ingenuity Pathway Analysis (IPA, Ingenuity Systems, Qiagen). The p-values reported for IPA results were generated by IPA using a right-sided Fisher exact test for over-representation analysis, Benjamini-Hochberg correction for multiple hypothesis testing correction, and a z-score algorithm for upstream analysis; p-values < 0.05 were considered significant. For the systems level analysis, only *C. mydas* DE transcripts that could be annotated to their closest characterized human homolog were included as input.

To better annotate DE transcripts that had turtle-specific gene identifiers [GenBank assembly accession number: GCA_000344595.1], which cannot be used with IPA, the sequence file containing all amino acid sequences for the green turtle genome was re-annotated using PANNZER2 with the –PANZ_FILTER_PERMISSIVE option [98]. When protein and product descriptions for the annotated DE transcripts agreed between PANNZER2 and the original green turtle genome annotation, the PANNZER2 annotation was used if it provided the name of the closest characterized human homolog instead of a turtle-specific identifier. However, it was often the case that the protein and product descriptions for the annotated DE transcripts were not in agreement, so a random subset of the protein sequences of 11 genes were blasted (blastp, https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) against the NCBI non-redundant protein database (nr) to determine which annotation method was most accurate. Since 10 out of the 11 protein sequences tested had the original green turtle genome annotation as the top hit, the genome annotation was used for instances in which the two annotations disagreed. If there was no human homolog available in this case, the genome protein description was checked against the STRING database for human homologs [99]. Human annotation was used to enable the most comprehensive systems-level analysis, as human genes have been the most extensively annotated and characterized. Out of all of the unique transcripts identified as differentially expressed in all pairwise comparisons, 63% were annotated using the available green turtle genome, 18% were annotated using the PANNZER2 re-annotation of the green turtle genome amino acid sequences, 13% were annotated using the protein description and the STRING database, and 6% of the transcripts either remained unannotated or the annotation was too ambiguous to use in downstream analyses.

Copy number analysis

To identify potential consistent mutations involved in driving tumorigenesis, copy number variations (CNVs) were identified between paired sets of whole genome DNA sequencing of tumor and healthy tissue each from kidney, lung, and external sources. To assign copy numbers (CNs) to genomic segments of these six genomes, the 140,023 genomic scaffolds of the *C. mydas* draft genome assembly were first combined into a single mega-chromosome. Then, the read-depth based algorithm cn.MOPS [100] was used in R to assign copy number, following the methods of Stammnitz et al. (2018)[55] and using a custom R script produced by the University of Cambridge Transmissible Cancer Group. The script has been made publically available at GitHub (https://github.com/MaximilianStammnitz/turtle-FP-cancer). Samples 27K4H, 27L5H, and liCSVS2 served as normal controls compared to 27K2, 27L1F,

and liRRF4 respectively, and read-depths were counted in 5000 bp bins across the combined scaffolds. Segmentation, which determines the length and position of a CNV, was performed using the cn.MOPS algorithm 'fastseg' with the referencecn.mops function to join consecutive segments with large or small expected fold changes to make a candidate segment. It then supplies candidate segments that show variations along the mega-chromosome and also across the supplied samples. The custom R script was used to model copy number posterior likelihoods for CN states CN0 through CN128. The locations of the segmented CNVs per scaffold from 27L1F (the only sample with clear copy number gains) were extracted and compared back to the *C. mydas* genome annotation file to compile a list of likely oncogenic genes. Gene lists were analyzed for overrepresented pathways, biological functions, and upstream regulators using IPA as detailed above. For this analysis, only the list of genes was input into IPA due to no available fold-change or p-value for CNVs.

To determine how copy number gains in the genome influence gene expression, RNA-Seq TPM normalized count data was compared between 27L1F (lung tumor with copy number gains) and 27K2F (kidney tumor that was copy number neutral) for each gene found within 27L1F's higher copy number regions. The expression for each tumor sample was first compared to the expression from its patient-matched healthy tissue sample (27L4H and 27K4H for lung and kidney, respectively) to produce an expression ratio to account for potential tissue specific differences in expression. The expression ratios were then compared between 27-2017-Cm's lung and kidney tumor samples. If the lung expression ratio was 100% greater than that of the kidney for a specific gene, that gene was considered highly dosage sensitive, in that expression was much higher in the sample with a copy number gain than one that was copy number neutral. If the lung expression ratio was 50% greater than that of the kidney for a specific gene, that gene was considered still slightly dosage sensitive. Any genes that did not fit these criteria, either due to similar expression ratios between the kidney and lung tissues, or due to expression not fitting the expected pattern (i.e. genes being downregulated in the lung tumor relative to healthy lung tissue, which would be expected in genomic regions with copy number losses), were classified as dosage resistant.

qPCR assays

qPCR assays were conducted on FP tissue samples to quantify viral load within a range of tumor types, as well as eDNA samples to look at viral shedding dynamics within rehabilitation tanks. DNA was extracted from 79 tissue samples from 13 juvenile green turtles which had stranded in Northern Florida as detailed above. DNA was also extracted from 60 tank water samples (20 samplings with three replicates per sampling event). ChHV5 viral load was quantified using TaqMan Fast Advanced Mastermix (ThermoFisher, Cat No. 4444557) according to the manufacturer's protocol by amplifying the ChHV5 virus-specific DNA polymerase (UL30) gene[25] (Supplemental Table 4). A species-specific assay was also developed to target the 16S ribosomal RNA gene (rRNA) for Atlantic populations of *C. mydas* to serve as positive controls and to compare the level of viral DNA present against the level of *C. mydas* DNA present within each sample (Supplemental Table 4). This sea turtle species was the primary focus as they are the ones that are most commonly housed at the Whitney Sea Turtle Hospital and most commonly afflicted by FP. A LightCycler490

Instrument II (Roche) was used for amplification and cycling parameters were as follows: 95 °C 10 min, 45 cycles of 95 °C 10 s, 60 °C 20 s, and 72 °C 20 s. All samples were run in triplicate with negative controls. Absolute quantification was utilized in this study. Standard curves using synthetic fragments of the UL30 gene and *C. mydas* 16S rRNA gene (see Supplemental Table 5) were generated to calculate the amount of DNA of these two genes present within each sample (in pg of DNA).

Histology methodology, embedding, sectioning and staining
Turtle tissue samples surgically removed using a CO_2 laser and stored in 4% paraformaldehyde at 4°C overnight. Samples washed twice in 1x PBS for 10 minutes; once in Milli-Q H_2O for 10 minutes; twice in 50% ethanol for 15 minutes; twice in 90% ethanol for 15 minutes; twice in 100% ethanol for 15 minutes. Samples stored in 100% ethanol at 4°C for 3 nights. Samples washed in 100% aniline for 1 hour; 50:50 aniline:methyl salicylate for 1 hour; twice in 100% methyl salicylate for 1.5 hours. Samples stored in 50:50 methyl salicylate:paraffin at 60°C overnight. Samples washed twice in 100% paraffin at 60°C for 3 hours. Samples stored in 100% paraffin overnight. Samples embedded in 100% paraffin and stored at 4°C.

Paraffin blocks sectioned into 6μm ribbons of six on charged Fisherbrand Superfrost Plus microscope slides using an AO Spencer "820" microtome and stored at room temperature

Tissue sections rehydrated by a series of washes: xylene A for 10 minutes; xylene B for 5 minutes; 50:50 xylene:alcohol for 5 minutes; 100% alcohol A for 5 minutes; 100% alcohol B for 5 minutes; 95% alcohol A for 5 minutes; 95% alcohol B for 5 minutes; 80% alcohol for 5 minutes; 70% alcohol for 5 minutes; 50% alcohol for 5 minutes; distilled H_2O for 5 minutes; 1x PBS. Tissue sections incubated at room temperature for 1.5 hours in 200μl PBS pre-incubation medium (1% normal goat serum + 0.1% bovine albumin serum + 0.1% Tritonx100 + 0.02% sodium azide + PBS). Tissue sections incubated at room temperature overnight in primary antibody medium or control medium (1:100 primary anti β-catenin antibody from rabbit Sigma C2206 + 1:100 primary anti β-actin antibody from mouse Sigma A5441 + PBS pre-incubation medium, or PBS pre-incubation medium only, respectively). Tissue sections washed twice in 1x PBS for 20 minutes. Tissue sections incubated at 37°C for 2 hours in 1:250 FITC GAR (goat anti-rabbit) + TRITC GAM (goat anti-mouse) + PBS pre-incubation medium. Secondary antibodies were from Jackson ImmunoResearch Labaoratires Inc. (West Grove, PA) and were affinity purified and selected for very low cross-reactivity with other animal sources of Ig. Tissue sections washed with 300ml 1x PBS and 2μl Hoechst 33342, trihydrochloride, trihydrate (Life Technologies Corp., Eugene OR) for 10 minutes. Tissue sections washed twice in 1x PBS for 10 minutes. Three drops of 60% glycerol in PBS containing PPD (p-phenylenediamine, 0.3mg/ml) as a fluorescence quench inhibitor were applied to the sections and a cover slip then added to each slide (6 tissue sections per slide). A Leica SP5 confocal microscope was used to visualize and capture images of the fluorescent staining in each tissue section.

For CD3 staining, tissue sections where sent to the University of Florida Veterinary Diagnostic Laboratories core facility, and were stained with rabbit anti-human CD3 ε chain antibody clone LN10 (RM-9107-S1 Thermosfisher, Labvision) and an alkaline-phosphatase based red chromogen detection kit. and co-stained with hematoxylin. This CD3 anit-body has previously been validated as also specifically recognizing green sea turtle CD3[56].

Retinoic Acid Therapeutic Methodology
Photos with a scale bar were taken of patients undergoing ectopic retinoic acid treatment using an Olympus Tough TG-5, bi-weekly, for the duration of their treatment. This allowed the surface area of each tumor to be analyzed using ImageJ. Direct measurements were also taken bi-weekly using iGaging digital calipers to record the length and width of each tumor. A topical retinoic acid therapeutic (Spear Tretinoin Cream 0.1%) was applied for a 6 - 8 week course depending on the veterinary determination of patient status. Each treated tumor was coupled with a control tumor in the same anatomical location on the opposite side of the body. Tumor length, width, and surface area were analyzed to determine the overall effectiveness of topical retinoic acid treatments for inhibiting FP tumor growth.

Data availability
The RNA-Seq and DNA-Seq data including raw reads are deposited in NCBI (https://www.ncbi.nlm.nih.gov/) under BioProject ID: PRJNA449022 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA449022). Code for the copy number analysis has been deposited on Github (https://github.com/MaximilianStammnitz/turtle-FP-cancer/).

Competing financial and non-financial competing interests statement
The authors declare no competing interests. Additionally, the funding agencies had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Table

Table 1. Comparison of viral sequencing reads per tumor and non-tumor tissue at the DNA and RNA level. Number of reads per 10 million total reads (RPTMT) are shown for each sample.

Sample	DNA viral load (RPTMT)	RNA viral reads (RPTMT)
Patient 27-2017-Cm:		
Lung FP tumor	3,673	68
Lung non-tumor	68	36
Kidney FP tumor	75	30
Kidney non-tumor	66	26
Patient 25-2018-Cm ('Lilac'):		
External new growth FP tumor	1,036	-
Skin non-tumor	81	-

Figure Legends

Figure 1. Fibropapillomatosis tumors and differential transcript expression. **A)** Top, left: Fibropapillomatosis-afflicted green sea turtles (*Chelonia mydas*) in one of the hospital's seawater tanks, awaiting tumor removal surgery. Established tumors are visible as large pinkish outgrowths. Top, middle: Numerous new growth tumors occurring around the ventral tail and rear flipper area of patient 25-2018-Cm 'Lilac'. Top, right: Post-surgical regrowth tumor imaged after surgical resection. Re-growing tumor is pinkish tissue surrounded by paler non-tumored skin.Bottom: A computed tomography (CT) scan of fibropapillomatosis-afflicted *C. mydas*. CT is one of the approaches used for diagnosing internal tumors. In all images arrows indicate selected examples of tumors. **B, C)** Overlap of transcripts significantly differentially expressed (DE) (as called by DESeq2) in fibropapillomatosis from the RNA-seq data. Area-proportional Venn diagrams were generated using BioVenn (http://www.biovenn.nl/)[97]. Transcripts were considered significant if passing the following cut-offs: adjusted p-value of < 0.05 and $\log_2$ fold change of > 2 or < -2. **B)** Overlap of DEs from the following comparisons: established external FP, new growth external FP, and regrowth external FP, when all are compared to healthy skin for differential expression analysis. **C)** Overlap of DEs from the following comparisons: established external FP, kidney FP, and lung FP, each compared to their non-tumored tissue sources for differential expression analysis (healthy skin, healthy kidney, and healthy lung, respectively). **D)** Principle component analysis (PCA) of all internal tumor samples compared to external tumor samples, RNA-seq. **E)** PCA of all samples, includes all tumor and non-tumor samples, RNA-seq.

Figure 2. Gene Ontology (GO) term analyses of differentially expressed transcripts in each tumor type. **A)** Volcano plot of genes differentially expressed between fibropapillomatosis tumors (all types) and non-tumor tissue (all types) as determined by DESeq2 analysis of the RNA-seq samples. Transcripts were considered significant if passing the following cut-offs: adjusted p-value of < 0.05 and $\log_2$ fold change of >2 and <-2. Plot generated using Degust: interactive RNA-seq analysis (http://degust.erc.monash.edu/)[101]. **B, C)** Activation/inhibition

z-scores of the top 20 disease-associated GO terms associated with transcripts differentially expressed in different types of fibropapillomatosis tumors (RNA-seq), as detected by IPA, ranked by p-value (calculated by right-tailed Fisher's Exact Test, with Benjamini-Hochberg correction). **B)** Established external FP versus non-tumored skin. A total of 689 of the significant DEs were recognized by IPA and used in the analysis. **C)** Kidney FP versus non-tumored kidney tissue. A total of 618 of the significant DEs were recognized by IPA and used in the analysis. **D)** Activation/inhibition z-scores of the six GO terms common to the top 20 GO terms of all sample comparisons, plus the 'Cancer' and 'Melanoma' GO terms, as detected by IPA, ranked by p-value (calculated by right-tailed Fisher's Exact Test, with Benjamini-Hochberg correction).

Figure 3. Transcriptomic- and histological-based immune profiling of fibropapillomatosis tumors. **A)** Overlap of the top 200 disease-associated GO terms associated with transcripts differentially expressed in different growth stages of external FP tumors (new growth, regrowth, established), as detected by IPA, ranked by p-value (calculated by right-tailed Fisher's Exact Test, with Benjamini-Hochberg correction). Both activated and inhibited GO terms for each tumor type when it was compared to its healthy tissue source were included. Area-proportional Venn diagrams were generated using BioVenn (http://www.biovenn.nl/). **B)** Activation/inhibition z-scores for eight (the leukocyte/lymphatic-associated GO terms) out of the 22 disease GO terms uniquely common to the top 200 ranked GO terms of early fibropapillomatosis (new growth and regrowth, see Fig. 3A), shown for all tumor types. Some of these GO terms were called for lung, kidney and/or established external tumors (as shown), although they fell outside of the top 200 ranked GO terms called for these three tumor types. **C)** CD3 anti-body based staining of T lymphocyte infiltration in new growth tumor tissue, with alkaline-Phosphatase based secondary anti-body staining producing reddish/purple positive staining. For ease of visualization nuclei are counterstained with Hematoxylin (blue staining). Selected positive CD3 stained areas are indicated by black arrows.

Figure 4. Transcriptional regulator analysis revealing the cellular signaling driving fibropapillomatosis and potential therapeutic targets. **A)** Overlap of the top 100 inferred transcriptional regulators (ITR) of the transcripts differentially expressed in different types of FP tumors (kidney FP, lung FP, external established FP) when compared to their respective non-tumored tissue sources, as detected by IPA, ranked by p-value (calculated by right-tailed Fisher's Exact Test, with Benjamini-Hochberg correction). Area-proportional Venn diagrams were generated using BioVenn (http://www.biovenn.nl/). The 16 ITRs that were common to lung, kidney, and established external FP were selected for further analysis. **B)** Interaction networks of the top 16 ITRs (9 genes shown, 7 drugs excluded) common between lung, kidney and established external tumors. Core network generated by String[99] (https://string-db.org/), expanded network generated by HumanBase (https://hb.flatironinstitute.org/). **C)** Activation/inhibition z-scores for the 16 ITRs common between lung, kidney, and established external tumors, shown for all tumor types. ITRs are segregated according to functional class,

i.e. genes and drugs **D**) Anti-body based immunohistochemistry of new growth and established external tumor tissue. Tissue sections are stained with Anti β-catenin (an ITR called for 3 of the 5 tumor types, see Fig. 4C) and counter stained with Hoechst 33342 to visualize nuclei and Anti β-actin. Selected cells with nuclear (activated) β -catenin staining are indicated by white arrows.

Figure 5. Network analysis of inferred transcriptional regulator analysis of fibropapillomatosis tumors. **A, B)** Network-based functional module discovery of the top 200 ranked ITRs (called by IPA) of **A)** established external and **B)** lung tumors generated by HumanBase (https://hb.flatironinstitute.org/). **C)** Activation/inhibition z-scores of the 'Colorectal Cancer Metastasis Signaling' gene ontology (GO) term associated with transcripts differentially expressed in different types of fibropapillomatosis tumors (kidney FP, lung FP, external FP) when compared to their respective non-tumored tissue sources, as detected by IPA. **D)** Activation/inhibition z-scores of 11 invasion and metastasis associated GO terms from transcripts differentially expressed in different types of fibropapillomatosis tumors (kidney FP, lung FP, external FP) when compared to their respective non-tumored tissue sources, as detected by IPA.

Figure 6. Genome-wide copy number variation (CNV) analysis of whole genome sequencing of patient-matched tumor and non-tumor tissue reveals possible oncogenic drivers in a lung tumor. **A)** Copy number variation analysis using the cn.MOPS package (Klambauer et al. 2012) implemented in R (https://www.r-project.org/). Tumored tissue was compared to non-tumored tissue from the same tissue source and same patient (27-2017-Cm). Each point shows a genomic segment 5000 bp long and its associated copy number within the tumor relative to its patient-matched non-tumor tissue, with a copy number of 2 being the null hypothesis for a diploid organism. Red data points indicate where the segmentation algorithm called a potential significant CNV. **B)** Top 25 disease-associated gene ontology (GO) terms, as detected by IPA, ranked by p-value (calculated by right-tailed Fisher-s Exact Test, with Benjamini-Hochberg correction) of the genes found within the potential significant CNV regions identified in 27-2017-Cm's lung tumor. While 220 genes were extracted from the putative CNV regions, only 174 were mapped (human-centric) and included in the IPA analysis. **C)** Functional interaction network of the MAPK-associated genes harboring CNVs in 27-2017-Cm's lung tumor. Network generated using HumanBase (https://hb.flatironinstitute.org/). **D)** Comparison of gene expression differences between kidney tumor (27K, copy number neutral) and lung tumor (27L, copy number gains) for potential highly dose sensitive genes found within the lung tumor CNV regions of 27-2017-Cm. Count ratios were generated by taking the transcripts per million (TPM) normalized read counts per gene from tumor samples and dividing that value by the normalized counts per gene from the patient-matched non-tumor tissue samples.

Figure 7. ChHV5 transcriptomics and genomics across tumor and non-tumor samples. **A)** ChHV5 expression across each sample type. Dot plot of the number of ChHV5 reads (RNA-seq) per 10 million total reads per sample. Line denotes mean, error bars denote standard error. Significant differences in averages between sample types were determined by a Kruskal Wallis with Dunn-Bonferroni post hoc and are denoted by an asterisk (*). Per sample type, non-tumor n = 20, established external tumor n = 35, regrowth external tumor n = 12, new growth external tumor n = 9, kidney tumor n = 6, lung tumor n = 7. **B)** Total number of samples in which reads (RNA-Seq) for each ChHV5 gene were detected. A gene was counted as detected if a sample had TPM-normalized counts > 0 for said ChHV5 gene. **C)** ChHV5 abundance (DNA-based) across each sample type. Dot plot of the number of ChHV5 reads (DNA-seq) per 10 million total reads across sample types. Line denotes mean, error bars denote standard error. Per sample type, non-tumor n = 3, fibropapillomatosis tumor n = 3, plasma from FP-afflicted turtles n = 10, tissue from non-tumor hatchlings n = 3, tank environmental DNA (eDNA) n = 1. **D)** ChHV5 viral quantification of a range of FP types (n = 43 samples) and non-tumor tissue (n = 36) samples from 13 individual patients using the UL30 DNA qPCR assay[25] (Supplemental Table 4). Error bars denote standard deviation. Absolute quantity of ChHV5 was determined through a standard curve of known amounts (in picograms) of a UL30 gene fragment (Supplemental Table 5). Truncated x-axis label, internal sub-cutan. = internal sub-cutaneous.

Figure 8. Environmental DNA (eDNA)-based detection, quantification and monitoring of ChHV5 viral shedding into patient tank water. **A)** Top, image of a marine leech removed from surface of a fibropapillomatosis tumor, with dark red *C. mydas* blood pellet visible. Bottom, detection and quantification of ChHV5 UL30 gene DNA by qPCR, using leeches as proxy eDNA samples (whole leech lysis and DNA extraction). Error bars denote standard deviation. **B) C**orrelation of individual patient tumor surface area (mm^2) and the concentration of ChHV5 virus being shed into their tank water, as detected by qPCR (positive correlation, R^2 = 0.5431, p = 0.0002, df = 19). Background water drop image licensed under CC BY 2.0**C** ("Water Sculptures" by Tom Bullock, https://ccsearch.creativecommons.org/photos/6ccaf9a6-6a31-4aff-82fc-a071066fb689). **C)** Time-course of tumor surface area changes (surgical removal) and concentration of ChHV5 shed into tank water, as detected by qPCR. **D)** Time-course of ChHV5 viral shedding into four individual patient tanks, as detected by qPCR. Tumor removal surgery events are denoted by an asterisk. Note: the ChHV5 eDNA detected in tank 1 in week 4 was due to a second patient (FP-positive) being added to that tank for a single week, due to the rehabilitation needs of the hospital. Error bars denote standard deviation. **E)** Correlation of *Chelonia mydas* eDNA shedding and ChHV5 eDNA shedding (positive correlation, R^2 = 0.66, p = 0.00001, df = 19), both *C. mydas* (16srRNA gene assay) and ChHV5 (UL30 gene assay) eDNA were detected by qPCR.

Supplemental Figure 1. Additional gene ontology (GO) term analyses of differentially expressed transcripts. **A - C)** Activation/inhibition z-scores of the top 20 disease-associated

GO terms associated with transcripts differentially expressed in different types of fibropapillomatosis tumors (RNA-seq), as detected by IPA, ranked by p-value (calculated by right-tailed Fisher's Exact Test, with Benjamini-Hochberg correction). **A)** New external FP versus non-tumored skin. A total of 698 of the significant DEs were recognized by IPA and used in the analysis. **B)** Regrowth external FP versus non-tumored skin. A total of 514 of the significant DEs were recognized by IPA and used in the analysis. **C)** Lung FP versus non-tumored lung tissue. A total of 653 of the significant DEs were recognized by IPA and used in the analysis.

Supplemental Figure 2. Additional fibropapillomatosis transcriptomic comparisons. **A)** Overlap of transcripts from RNA-seq data significantly differentially expressed (DE) (as called by DESeq2) in fibropapillomatosis between either kidney FP compared to established external FP, or lung FP compared to established external FP. Area-proportional diagrams were generated using BioVenn (http://www.biovenn.nl/). Transcripts were considered significant if passing the following cut-offs: adjusted p-value of < 0.05 and $\log_2$ fold change of > 2 and < -2. **B, C)** Activation/inhibition z-scores of the top 20 disease-associated gene ontology (GO) terms associated with transcripts differentially expressed in different types of fibropapillomatosis tumors (RNA-seq), as detected by IPA, ranked by p-value (calculated by right-tailed Fisher's Exact Test, with Benjamini-Hochberg correction). **B)** Kidney FP versus established external FP tumors. **C)** Lung FP versus established external FP tumors. **D)** Activation z-scores of the Interferon gamma (IFNG) inferred transcriptional regulator (ITR) associated with transcripts differentially expressed across the five different types of fibropapillomatosis tumors, when compared to their respective non-tumor tissue sources.

Supplemental Figure 3. A) Time-course of relative tumor growth profiles of retinoic acid (RA, Tretinoin cream 0.1%) treated and untreated fibropapillomatosis tumors. Profiles for eight tumors across three individual *C. mydas* patients are shown. Duration of treatment was under veterinary determination. Patient 'name' abbreviations, Ferd., 'Ferdinand' (07-2018-Cm), Eins., 'Einstein' (28-2018-Cm) and 'Lilac' (25-2018-Cm). Tumor growth is relative to the size of each individual tumor before treatment, i.e. Day 1.

Supplemental Figure 4. Additional fibropapillomatosis tumor immunohistochemistry. **A)** Anti-body based immunohistochemistry of established external, regrowth external, regrowth eye external and kidney internal tumor tissue. Tissue sections are stained for β-catenin (anti-β-catenin antibody) and counter stained with Hoechst 33342 to visualize nuclei and Anti β-actin. Selected cells with nuclear (activated) β -catenin staining are indicated by white arrows.

Supplemental Figure 5. Additional fibropapillomatosis inferred transcriptional regulator (ITR) interaction networks. **A, B)** Interaction networks of the top 200 ITRs of established **A)**

external tumors and of **B)** lung tumors. Networks were generated by String[99] (https://string-db.org/).

Supplemental Figure 6. Additional fibropapillomatosis inferred transcriptional regulator (ITR) network-based functional module discovery. **A-C)** Network-based functional module discovery of the top 200 ranked ITRs (called by IPA) of **A)** kidney tumors **B)** new growth external tumors and **C)** regrowth external tumors. Networks were generated using HumanBase (https://hb.flatironinstitute.org/). **D)** Activation z-scores of the 'Quantity of Metal' gene ontology (GO) term associated with transcripts differentially expressed in different types of fibropapillomatosis tumors (kidney FP, lung FP, external FP) when compared to their respective non-tumored tissue sources, as detected by IPA. Note: the kidney FP disease-associated GO analysis did not have the 'Quantity of Metal' GO term called when compared to healthy kidney tissue.

Supplemental Figure 7. Additional copy number variation (CNV) data. **A)** Genome-wide copy number variation (CNV) analysis of whole genome sequencing of patient-matched tumor and non-tumor tissue (top: kidney tumor compared to healthy kidney from patient 27-2017-Cm; bottom: early external tumor compared to non-tumor skin from 25-2018-Cm, 'Lilac'). Copy number variation analysis was carried out using the cn.MOPS package[100] in R. Tumored tissue was compared to non-tumored tissue from the same tissue source and same patient for each comparison. Each point shows a genomic segment 5000 bp long and its associated copy number within the tumor relative to its patient-matched non-tumor tissue, with a copy number of 2 being the null hypothesis for a diploid organism. Red data points indicate where the segmentation algorithm called a potential significant CNV. Both samples shown above are examples of a copy number neutral state. **B**) Network-based functional module discovery of the genes harboring CNV (gains) in 27-2017-Cm's lung fibropapillomatosis tumor. Network was generated using HumanBase (https://hb.flatironinstitute.org/).

Supplemental Figure 8. Additional gene level ChHV5 transcriptomics analysis. **A)** Bar graph showing level of ChHV5 viral transcript expression of every gene in the ChHV5 genome, averaged by tissue type. Reads were first aligned from RNA-seq samples to the ChHV5 genome[91] via bowtie2 and read counts generated per viral gene using htseq-count. Reads were normalized by gene length and sequencing depth to transcripts per million (TPM) and averaged across samples for the following groups: non-tumor tissue n = 20, external fibropapillomatosis tumors n = 56, lung fibropapillomatosis tumors n = 7, and kidney fibropapillomatosis tumors n = 6.

Supplemental Figure 9. ChHV5 transcript levels compared with rehabilitation outcome. **A)** Bar graph showing total level of ChHV5 RNA among samples by individual patient, as assessed by RNA-seq. Reads were normalized by sequencing depth to number of viral reads

per 10 million total reads per sample. Bars show individual samples per patient. **B)** Dot plot of the number of ChHV5 reads (RNA-Seq) per 10 million total reads between patients based on outcome (released vs died in care/humanely euthanized). Line denotes mean, error bars denote standard error. Significant difference in averages between the two outcomes was determined by a Mann-Whitney U Test and is denoted by an asterisk (*). Per outcome: released = 7 turtles; died/euthanized = 5 turtles.

Supplemental Figure 10. Additional gene level ChHV5 genomic (DNA-seq) analysis. **A)** Bar graph showing level of ChHV5 viral DNA reads of every gene in the ChHV5 genome. For the three non-tumor samples reads per gene were average. Reads were first aligned from DNA-seq samples to the ChHV5 genome[91] via bowtie2 and read counts generated using htseq-count. Reads were normalized by gene length and sequencing depth to transcripts per million (TPM). Per sample type: nontumor tissue n = 3, external FP tumor n = 1, lung FP tumor n = 1, and kidney FP tumor n = 1.

References

1   Quesada, V. *et al.* Giant tortoise genomes provide insights into longevity and age-related disease. *Nature Ecology & Evolution* **3**, 87-95, doi:10.1038/s41559-018-0733-x (2019).

2   Whilde, J., Martindale, M. Q. & Duffy, D. J. Precision wildlife medicine: applications of the human-centred precision medicine revolution to species conservation. *Global Change Biol.* **23**, 1792-1805, doi:10.1111/gcb.13548 (2017).

3   Schulz, T. F. & Cesarman, E. Kaposi Sarcoma-associated Herpesvirus: mechanisms of oncogenesis. *Current Opinion in Virology* **14**, 116-128, doi:http://dx.doi.org/10.1016/j.coviro.2015.08.016 (2015).

4   Davidson, N. E. *et al.* AACR Cancer Progress Report 2016. *Clin. Cancer. Res.* **22**, S1-S137, doi:10.1158/1078-0432.ccr-16-1993 (2016).

5   McAloose, D. & Newton, A. L. Wildlife cancer: a conservation perspective. *Nat. Rev. Cancer* **9**, 517-526 (2009).

6   Williams Jr, E. H. *et al.* An epizootic of cutaneous fibropapillomas in green turtles Chelonia mydas of the Caribbean: part of a panzootic? *J. Aquat. Anim. Health* **6**, 70-78 (1994).

7   Jones, K., Ariel, E., Burgess, G. & Read, M. A review of fibropapillomatosis in Green turtles (Chelonia mydas). *The Veterinary Journal* **212**, 48-57, doi:http://dx.doi.org/10.1016/j.tvjl.2015.10.041 (2016).

8   Mascarenhas, R. & Iverson, P. J. Fibropapillomatosis in stranded green turtles (Chelonia mydas) in Paraiba State, Northeastern Brazil: evidence of a Brazilian epizootic. *Mar. Turtle Newsl.* **120**, 3-6 (2008).

9   Williams, E. H. *et al.* An Epizootic of Cutaneous Fibropapillomas in Green Turtles Chelonia mydas of the Caribbean: Part of a Panzootic? *J. Aquat. Anim. Health* **6**, 70-78, doi:10.1577/1548-8667(1994)006<0070:AEOCFI>2.3.CO;2 (1994).

10 Hargrove, S., Work, T., Brunson, S., Foley, A. M. & Balazs, G. Proceedings of the 2015 International Summit on Fibropapillomatosis: Global Status, Trends, and Population Impacts. 85 (2016).

11 Page-Karjian, A. *et al.* Factors influencing survivorship of rehabilitating green sea turtles (Chelonia mydas) with fibropapillomatosis. *J. Zoo Wildl. Med.* **45**, 507-519, doi:doi:10.1638/2013-0132R1.1 (2014).

12   Herbst, L. H. Fibropapillomatosis of marine turtles. *Annu. Rev. Fish Dis.* **4**, 389-425, doi:http://dx.doi.org/10.1016/0959-8030(94)90037-X (1994).

13   Aguirre, A. A. & Lutz, P. Marine Turtles as Sentinels of Ecosystem Health: Is Fibropapillomatosis an Indicator? *EcoHealth* **1**, 275-283, doi:10.1007/s10393-004-0097-3 (2004).

14   Williams, E. H. & Bunkley-Williams, L. Early Fibropapillomas in Hawaii and Occurrences in All Sea Turtle Species: the Panzootic, Associated Leeches Wide-Ranging on Sea Turtles, and Species of Study Leeches Should Be Identified. *J. Virol.* **80**, 4643-4644, doi:10.1128/JVI.80.9.4643-4644.2006 (2006).

15   Duarte, A. *et al.* First histological and virological report of fibropapilloma associated with herpesvirus in Chelonia mydas at Príncipe Island, West Africa. *Arch. Virol.* **157**, 1155-1159, doi:10.1007/s00705-012-1285-z (2012).

16   Foley, A. M., Singel, K., Hardy, R., Bailey, R. & Schaf, S. Distributions, relative abundances, and mortality factors for sea turtles in Florida from 1980 through 2007 as determined from strandings. *Florida Fish and Wildlife Conservation Commission* (2007).

17   International, C. *In: Invasive Species Compendium: fibropapillomatosis of sea turtles* 2019).

18   Stacy, B. A. *et al.* Report of the Technical Expert Workshop: Developing Recommendations for Field Response, Captive Management, and Rehabilitation of Sea Turtles with Fibropapillomatosis. doi:https://doi.org/10.25923/adg2-g771 (2019).

19   Shaver, D. J., Walker, J. S. & Backof, T. F. Fibropapillomatosis prevalence and distribution in green turtles Chelonia mydas in Texas (USA). *Dis. Aquat. Org.* **136**, 175-182 (2019).

20   Foley, A. M. *et al.* Distributions, relative abundances, and mortality factors of sea turtles in Florida during 1980–2014 as determined from strandings. *Fish and Wildlife Research Institute, Jacksonville Field Laboratory, Jacksonville, Florida, USA* (2015).

21   da Silva-Júnior, E. S. *et al.* Stranded Marine Turtles in Northeastern Brazil: Incidence and Spatial–Temporal Distribution of Fibropapillomatosis. *Chelonian Conserv. Biol.* **In Press**, doi:10.2744/ccb-1359.1 (2019).

22   Loureiro, N. d. S. & Matos, D. Presence of fibropapillomatosis in green turtles Chelonia mydas at Príncipe Island in the Gulf of Guinea. *Arquipélago-Life and Marine Sciences* **26**, 79-83 (2009).

23   Herbst, L. *et al.* Experimental transmission of green turtle fibropapillomatosis using cell-free tumor extracts. *Dis. Aquat. Org.* **22**, 1-12, doi:10.3354/dao022001 (1995).

24   Alfaro-Núñez, A. *et al.* Further evidence of Chelonid herpesvirus 5 (ChHV5) latency: high levels of ChHV5 DNA detected in clinically healthy marine turtles. *PeerJ* **4:e2274** (2016).

25   Page-Karjian, A. *et al.* Quantifying chelonid herpesvirus 5 in symptomatic and asymptomatic rehabilitating green sea turtles. *Endangered Species Research* **28**, 135-146 (2015).

26   Duffy, D. J. *et al.* Sea turtle fibropapilloma tumors share genomic drivers and therapeutic vulnerabilities with human cancers. *Communications Biology* **1**, 63, doi:10.1038/s42003-018-0059-x (2018).

27   Alfaro-Núñez, A. *et al.* Global distribution of Chelonid fibropapilloma-associated herpesvirus among clinically healthy sea turtles. *BMC Evol. Biol.* **14**, 1-12, doi:10.1186/s12862-014-0206-z (2014).

28   Herbst, L. H. & Klein, P. A. Green turtle fibropapillomatosis: challenges to assessing the role of environmental cofactors. *Environ. Health Perspect.* **103**, 27-30 (1995).

29   Van Houtan, K. S., Hargrove, S. K. & Balazs, G. H. Land Use, Macroalgae, and a Tumor-Forming Disease in Marine Turtles. *PLoS ONE* **5**, e12900, doi:10.1371/journal.pone.0012900 (2010).

30   Duffy, D. J. & Martindale, M. Q. Perspectives on the expansion of human precision oncology and genomic approaches to sea turtle fibropapillomatosis. *Communications Biology* **2**, 54, doi:10.1038/s42003-019-0301-1 (2019).

31   Work, T. M., Dagenais, J., Balazs, G. H., Schettle, N. & Ackermann, M. Dynamics of Virus Shedding and In Situ Confirmation of Chelonid Herpesvirus 5 in Hawaiian Green Turtles With Fibropapillomatosis. *Vet. Pathol.* **52**, 1195-1201, doi:10.1177/0300985814560236 (2015).

32   Page-Karjian, A. *et al.* Potential Non-Cutaneous Sites of Chelonid Herpesvirus 5 Persistence and Shedding in Green Sea Turtles (Chelonia mydas). *J. Aquat. Anim. Health* **29**, 136-142, doi:10.1080/08997659.2017.1321590 (2017).

33   Chaves, A. *et al.* Examining the Role of Transmission of Chelonid Alphaherpesvirus 5. *EcoHealth* **14**, 530-541, doi:10.1007/s10393-017-1248-7 (2017).

34   Greenblatt, R. J. *et al.* The Ozobranchus leech is a candidate mechanical vector for the fibropapilloma-associated turtle herpesvirus found latently infecting skin tumors on Hawaiian green turtles (Chelonia mydas). *Virology* **321**, 101-110, doi:http://dx.doi.org/10.1016/j.virol.2003.12.026 (2004).

35   Adams, C. I. *et al.* Beyond Biodiversity: Can Environmental DNA (eDNA) Cut It as a Population Genetics Tool? *Genes* **10**, 192 (2019).

36   Boussarie, G. *et al.* Environmental DNA illuminates the dark diversity of sharks. *Science Advances* **4**, doi:10.1126/sciadv.aap9661 (2018).

37   Evans, N. T. & Lamberti, G. A. Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool. *Fisheries Research* **197**, 60-66 (2018).

38   Qu, C. & Stewart, K. A. Evaluating monitoring options for conservation: comparing traditional and environmental DNA tools for a critically endangered mammal. *The Science of Nature* **106**, 9 (2019).

39   Roussel, J. M., Paillisson, J. M., Treguier, A. & Petit, E. The downside of eDNA as a survey tool in water bodies. *J. Appl. Ecol.* **52**, 823-826 (2015).

40   Strand, D. A. *et al.* Monitoring a Norwegian freshwater crayfish tragedy: eDNA snapshots of invasion, infection and extinction. *J. Appl. Ecol.* **56**, 1661-1673, doi:10.1111/1365-2664.13404 (2019).

41   Seymour, M. *et al.* Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology* **1**, 4, doi:10.1038/s42003-017-0005-3 (2018).

42   Creer, S. *et al.* The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution* **7**, 1008-1018, doi:10.1111/2041-210X.12574 (2016).

43   Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **29**, 358-367, doi:10.1016/j.tree.2014.04.003 (2014).

44   Erickson, R. A., Merkes, C. M. & Mize, E. L. Sampling designs for landscape-level eDNA monitoring programs. *Integr. Environ. Assess. Manage.* (2019).

45   Miaud, C., Arnal, V., Poulain, M., Valentini, A. & Dejean, T. eDNA Increases the Detectability of Ranavirus Infection in an Alpine Amphibian Population. *Viruses* **11**, 526 (2019).

46   Farrell, J., Thomas, R., Martindale, M. Q. & Duffy, D. J. Characterisation of fibropapillomatosis tumour growth profiles in green sea turtles (Chelonia mydas). *Testudo* **8**, 12-29 (2018).

47   Whilde, J. *et al.* Behaviour of juvenile green turtles (*Chelonia mydas*) before and after fibropapillomatosis tumour removal. *Testudo* **9**, 22-35 (2019).

48   Work, T. M., Dagenais, J., Weatherby, T. M., Balazs, G. H. & Ackermann, M. In-vitro replication of Chelonid herpesvirus 5 in organotypic skin cultures from Hawaiian green turtles (Chelonia mydas). *J. Virol.* **91**, e00404-00417 doi:10.1128/jvi.00404-17 (2017).

49   Morrison, C. L. *et al.* Genomic evolution, recombination, and inter-strain diversity of chelonid alphaherpesvirus 5 from Florida and Hawaii green sea turtles with fibropapillomatosis. *PeerJ* **6**, e4386, doi:10.7717/peerj.4386 (2018).

50   Sellera, F. P. *et al.* Investigation of green turtle (Chelonia mydas) cutaneous fibropapillomatosis recurrence rates following diode laser surgery. *Journal of Exotic Pet Medicine* **28**, 180-184, doi:https://doi.org/10.1053/j.jepm.2018.10.010 (2019).

51   Domiciano, I. G. *et al.* Chelonid Alphaherpesvirus 5 DNA in Fibropapillomatosis-Affected Chelonia mydas. *EcoHealth*, doi:10.1007/s10393-019-01412-8 (2019).

52   Rossi, S. *et al.* Monitoring green sea turtles in Brazilian feeding areas: relating body condition index to fibropapillomatosis prevalence. *J. Mar. Biol. Assoc. U.K.*, 1-9, doi:10.1017/S0025315419000730 (2019).

53   Jones, K. *et al.* Molecular evidence for horizontal transmission of chelonid alphaherpesvirus 5 at green turtle (Chelonia mydas) foraging grounds in Queensland, Australia. *PLOS ONE* **15**, e0227268, doi:10.1371/journal.pone.0227268 (2020).

54   Duffy, D. J. Problems, challenges and promises: perspectives on precision medicine. *Briefings in Bioinformatics* **17**, 494-504, doi:10.1093/bib/bbv060 (2016).

55   Stammnitz, M. R. *et al.* The Origins and Vulnerabilities of Two Transmissible Cancers in Tasmanian Devils. *Cancer Cell* **33**, 607-619.e615, doi:10.1016/j.ccell.2018.03.013 (2018).

56   Muñoz, F. A. *et al.* Identification of CD3+ T lymphocytes in the green turtle Chelonia mydas. *Vet. Immunol. Immunopathol.* **131**, 211-217, doi:https://doi.org/10.1016/j.vetimm.2009.04.015 (2009).

57   Gouaillard, C., Huchenq-Champagne, A., Arnaud, J., Chen, C.-l. H. & Rubin, B. Evolution of T cell receptor (TCR) α β heterodimer assembly with the CD3 complex. *Eur. J. Immunol.* **31**, 3798-3805, doi:10.1002/1521-4141(200112)31:12<3798::aid-immu3798>3.0.co;2-z (2001).

58   Duffy, D. J. *et al.* Retinoic acid and TGF-β signalling cooperate to overcome MYCN-induced retinoid resistance. *Genome Medicine* **9**, 15, doi:10.1186/s13073-017-0407-3 (2017).

59   Hanahan, D. & Weinberg, Robert A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013.

60   Perrault, J. R. & Coppenrath, C. M. Albinism In Florida Green Turtle (Chelonia mydas) Hatchlings: Ratio-Based Evidence Of Basic Mendelian Recessiveness. *Mar. Turtle Newsl.*, 38-40 (2019).

61   Monezi, T. A. *et al.* Chelonid herpesvirus 5 in secretions and tumor tissues from green turtles (Chelonia mydas) from Southeastern Brazil: A ten-year study. *Vet. Microbiol.* **186**, 150-156, doi:http://dx.doi.org/10.1016/j.vetmic.2016.02.020 (2016).

62   Smith, K., Acevedo-Whitehouse, K. & Pedersen, A. The role of infectious diseases in biological conservation. *Anim. Conserv.* **12**, 1-12 (2009).

63   Brearley, G. *et al.* Wildlife disease prevalence in human-modified landscapes. *Biological Reviews* **88**, 427-442, doi:10.1111/brv.12009 (2013).

64   Burge, C. A. *et al.* Climate Change Influences on Marine Infectious Diseases: Implications for Management and Society. *Annual Review of Marine Science* **6**, 249-277, doi:doi:10.1146/annurev-marine-010213-135029 (2014).

65   Maxwell, S. L., Fuller, R. A., Brooks, T. M. & Watson, J. E. Biodiversity: The ravages of guns, nets and bulldozers. *Nature* **536**, 143-145 (2016).

66   Heard, M. J. *et al.* The Threat of Disease Increases as Species Move Toward Extinction. *Conserv. Biol.* **27**, 1378-1388, doi:10.1111/cobi.12143 (2013).

67   Giraudeau, M., Sepp, T., Ujvari, B., Ewald, P. W. & Thomas, F. Human activities might influence oncogenic processes in wild animal populations. *Nature Ecology & Evolution*, doi:10.1038/s41559-018-0558-7 (2018).

68   Aguirre, A. A., Balazs, G. H., Spraker, T. R. & Gross, T. S. Adrenal and hematological responses to stress in juvenile green turtles (Chelonia mydas) with and without fibropapillomas. *Physiol. Zool.* **68**, 831-854 (1995).

69   dos Santos, R. G. *et al.* Relationship between fibropapillomatosis and environmental quality: a case study with Chelonia mydas off Brazil. *Dis Aquat Organ* **89**, 87-95, doi:10.3354/dao02178 (2010).

70   Rossi, S., Hazarbassanov, N. G. T. d. Q., Sánchez-Sarmiento, A. M., Prioste, F. E. S. & Matushima, E. R. Immune Response of Green Sea Turtles with and without

Fibropapillomatosis: Evaluating Oxidative Burst and Phagocytosis via Flow Cytometry. *Chelonian Conserv. Biol.* **15**, 273-278, doi:10.2744/ccb-1202.1 (2016).

71 Work, T. M., Rameyer, R. A., Balazs, G. H., Cray, C. & Chang, S. P. Immune status of free-ranging green turtles with fibropapillomatosis from Hawaii. *J. Wildl. Dis.* **37**, 574-581 (2001).

72 Guimarães, S. M., Gitirana, H. M., Wanderley, A. V., Monteiro-Neto, C. & Lobo-Hajdu, G. Evidence of regression of fibropapillomas in juvenile green turtles Chelonia mydas caught in Niterói, southeast Brazil. *Dis. Aquat. Org.* **102**, 243-247 (2013).

73 London, C. A. Tyrosine kinase inhibitors in veterinary medicine. *Topics in companion animal medicine* **24**, 106-112 (2009).

74 Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985-999, doi:10.1016/j.cell.2017.05.016 (2017).

75 Takebe, N. *et al.* Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nat Rev Clin Oncol* **12**, 445-464, doi:10.1038/nrclinonc.2015.61 (2015).

76 Neuzillet, C. *et al.* Targeting the TGFβ pathway for cancer therapy. *Pharmacol. Ther.* **147**, 22-31, doi:https://doi.org/10.1016/j.pharmthera.2014.11.001 (2015).

77 Kolch, W., Calder, M. & Gilbert, D. When kinases meet mathematics: the systems biology of MAPK signalling. *FEBS Lett.* **579**, 1891-1895 (2005).

78 Welsh, S. J. & Corrie, P. G. Management of BRAF and MEK inhibitor toxicities in patients with metastatic melanoma. *Therapeutic advances in medical oncology* **7**, 122-136 (2015).

79 Sinha, R. *et al.* Cutaneous adverse events associated with vemurafenib in patients with metastatic melanoma: practical advice on diagnosis, prevention and management of the main treatment-related skin toxicities. *Br. J. Dermatol.* **167**, 987-994 (2012).

80 Cohen, D. N. *et al.* Contribution of Beta-HPV Infection and UV Damage to Rapid-Onset Cutaneous Squamous Cell Carcinoma during BRAF-Inhibition Therapy. *Clin. Cancer. Res.* **21**, 2624-2634, doi:10.1158/1078-0432.ccr-14-2667 (2015).

81 Holderfield, M. *et al.* Vemurafenib Cooperates with HPV to Promote Initiation of Cutaneous Tumors. *Cancer Res.* **74**, 2238-2245, doi:10.1158/0008-5472.can-13-1065-t (2014).

82 Niller, H. H., Wolf, H. & Minarovits, J. Viral hit and run-oncogenesis: Genetic and epigenetic scenarios. *Cancer Lett.* **305**, 200-217, doi:http://dx.doi.org/10.1016/j.canlet.2010.08.007 (2011).

83 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415 (2013).

84 Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836.e816, doi:https://doi.org/10.1016/j.cell.2019.03.001 (2019).

85 Ene, A. *et al.* DISTRIBUTION OF CHELONID FIBROPAPILLOMATOSIS-ASSOCIATED HERPESVIRUS VARIANTS IN FLORIDA: MOLECULAR GENETIC EVIDENCE FOR INFECTION OF TURTLES FOLLOWING RECRUITMENT TO NERITIC DEVELOPMENTAL HABITATS. *J. Wildl. Dis.* **41**, 489-497, doi:10.7589/0090-3558-41.3.489 (2005).

86 Schneider, J. *et al.* Detection of Invasive Mosquito Vectors Using Environmental DNA (eDNA) from Water Samples. *PLOS ONE* **11**, e0162493, doi:10.1371/journal.pone.0162493 (2016).

87 Münz, C. Latency and lytic replication in Epstein–Barr virus-associated oncogenesis. *Nature Reviews Microbiology* **17**, 691-700, doi:10.1038/s41579-019-0249-7 (2019).

88 Ljubojevic, S. & Skerlev, M. HPV-associated diseases. *Clinics in Dermatology* **32**, 227-234, doi:http://dx.doi.org/10.1016/j.clindermatol.2013.08.007 (2014).

89 Sanclemente, G. & Gill, D. Human papillomavirus molecular biology and pathogenesis. *Journal of the European Academy of Dermatology and Venereology* **16**, 231-240, doi:10.1046/j.1473-2165.2002.00419.x (2002).

90 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, doi:10.1093/bioinformatics/btu170 (2014).

91 Ackermann, M. *et al.* The Genome of Chelonid Herpesvirus 5 Harbors Atypical Genes. *PLoS ONE* **7**, e46623, doi:10.1371/journal.pone.0046623 (2012).

92   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, doi:10.1038/nmeth.1923 (2012).

93   Wang, Z. *et al.* The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat. Genet.* **45**, 701-706, doi:10.1038/ng.2615

http://www.nature.com/ng/journal/v45/n6/abs/ng.2615.html#supplementary-information (2013).

94   Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357, doi:10.1038/nmeth.3317

https://www.nature.com/articles/nmeth.3317#supplementary-information (2015).

95   Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

96   Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotech* **32**, 896-902, doi:10.1038/nbt.2931

http://www.nature.com/nbt/journal/v32/n9/abs/nbt.2931.html#supplementary-information (2014).

97   Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488 (2008).

98   Koskinen, P., Törönen, P., Nokso-Koivisto, J. & Holm, L. PANNZER: high-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics* **31**, 1544-1552 (2015).

99   Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607-D613, doi:10.1093/nar/gky1131 (2018).

100   Klambauer, G. *et al.* cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* **40**, e69-e69, doi:10.1093/nar/gks003 (2012).

101   Powell., D. R., Milton, M., Perry, A. & Santos, K. Vol. Degust: interactive RNA-seq analysis. DOI: 10.5281/zenodo.3258932   (2019).
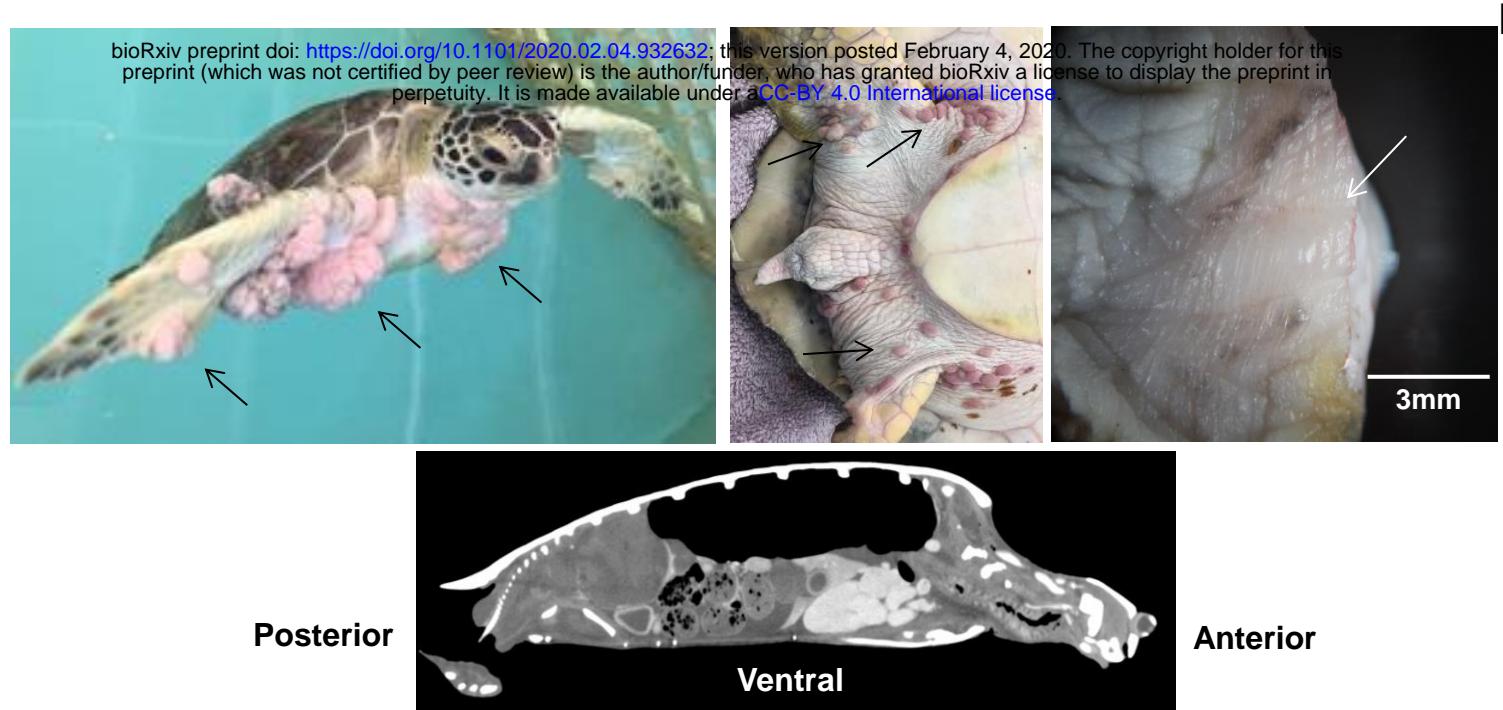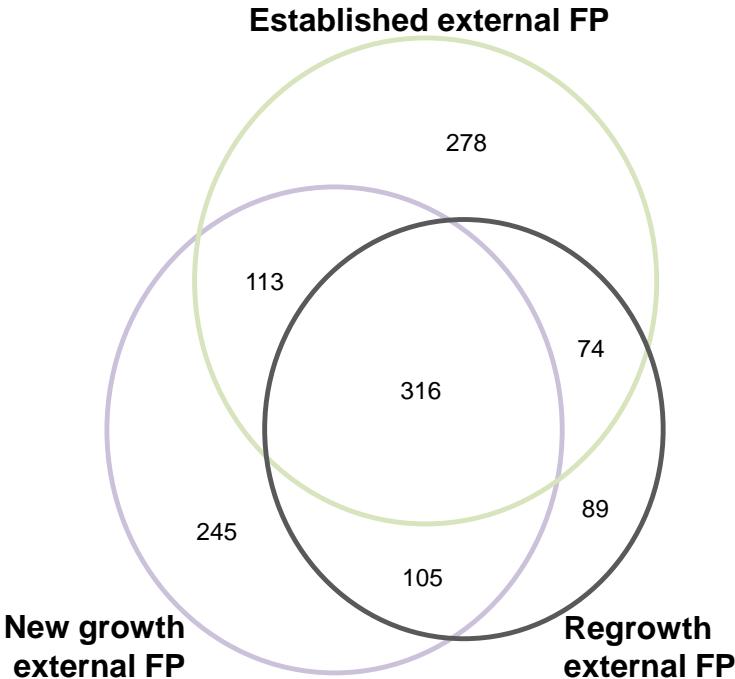
Fig.1

A

Posterior

Ventral

Anterior

3mm

B **Differentially expressed transcript overlap** (RNA-seq)

Established external FP

278

113

74

316

245

89

105

New growth external FP

Regrowth external FP

C **Differentially expressed transcript overlap** (RNA-seq)

Established external FP

605

68

65

43

429

109

494

Kidney FP

Lung FP
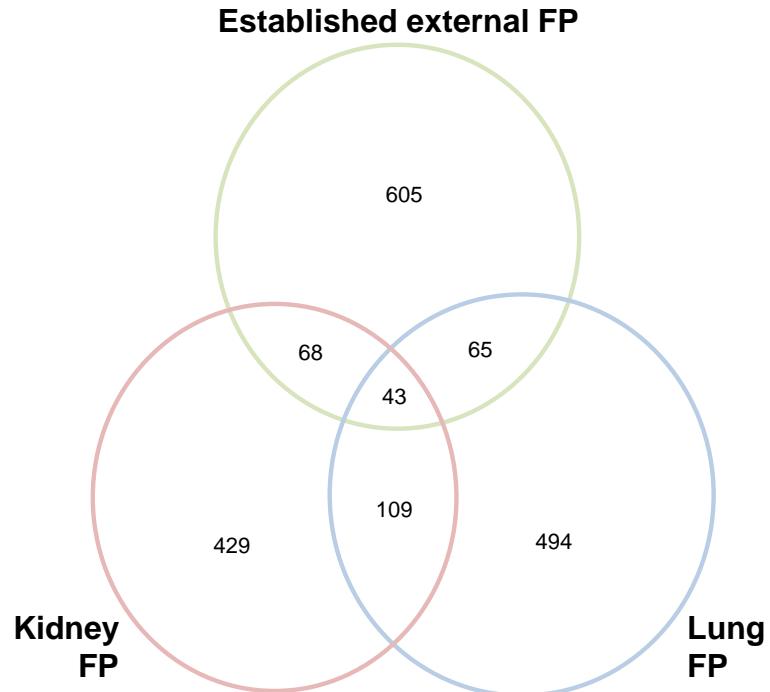
D **Principle component analysis of external and internal tumor RNA-seq samples**

PC1 (96.66%)

PC2 (1.48%)

△ External tumor   ○ Internal tumor
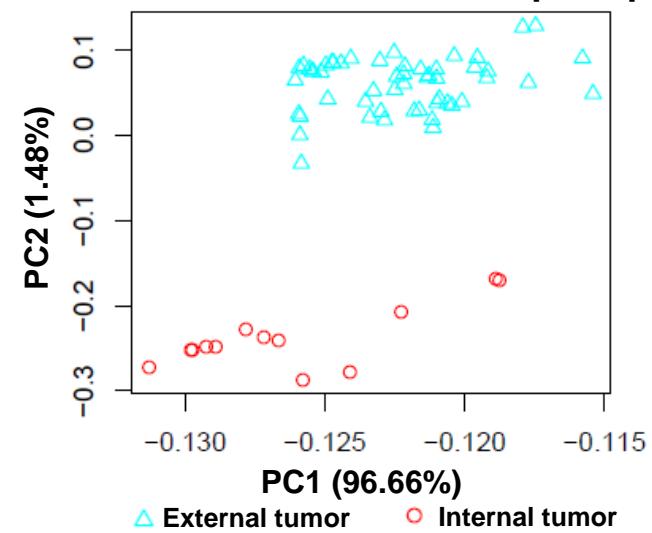
E **Principle component analysis of all RNA-seq samples**
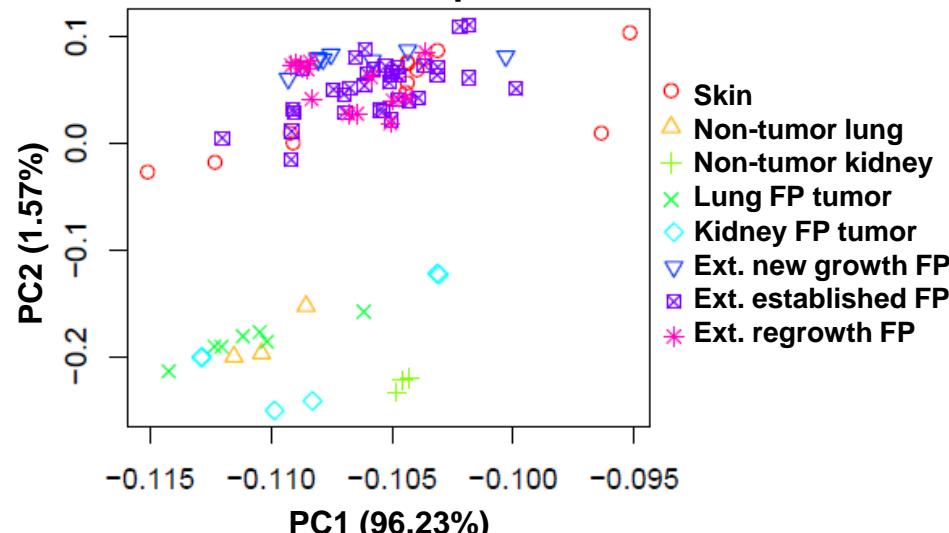
PC1 (96.23%)

PC2 (1.57%)

○ Skin
△ Non-tumor lung
+ Non-tumor kidney
✕ Lung FP tumor
◇ Kidney FP tumor
▽ Ext. new growth FP
⊠ Ext. established FP
✳ Ext. regrowth FP

Figure 2

A

Genes differentially expressed between all FP tumor and all non-tumor tissue samples (RNA-seq)

A

**Established external FP versus non-tumored skin (RNA-seq) Top 20 GO terms**

p-value range: 1.93E-32 to 3.59E-21

B

Kidney FP versus non-tumored kidney (RNA-seq) Top 20 GO terms

p-value range: 6.09E-35 to 4.51E-25
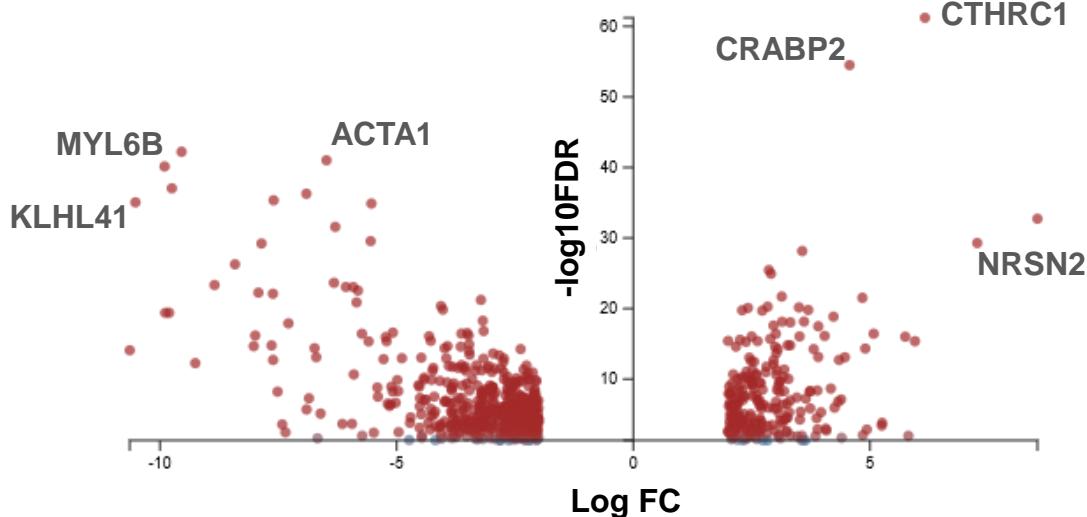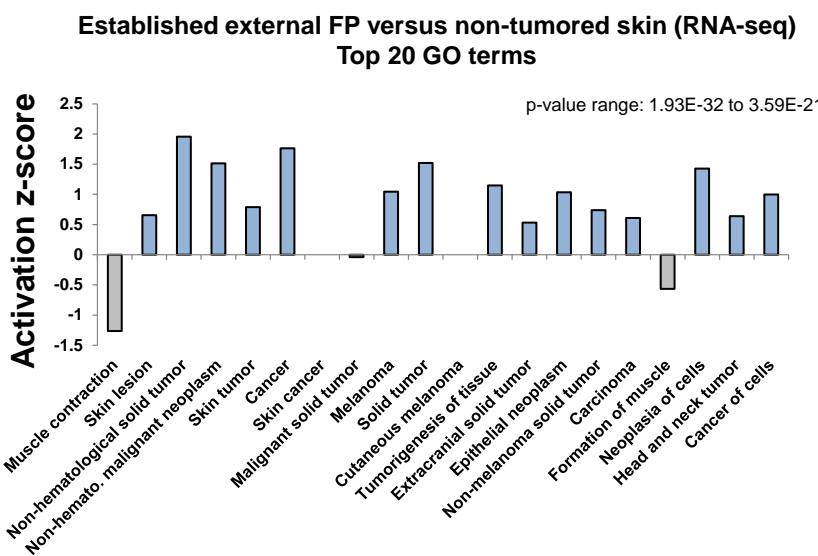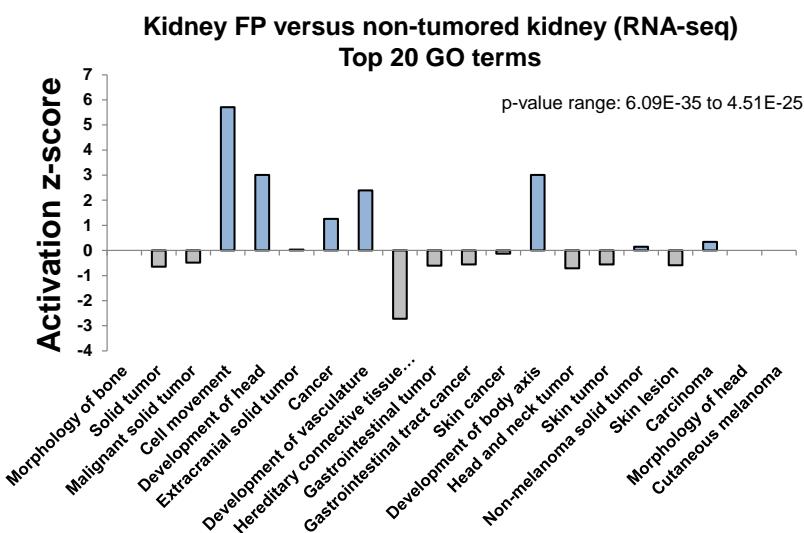
C

Six GO terms common to top 20 of all, plus 'Cancer' and 'Melanoma' (RNA-seq)
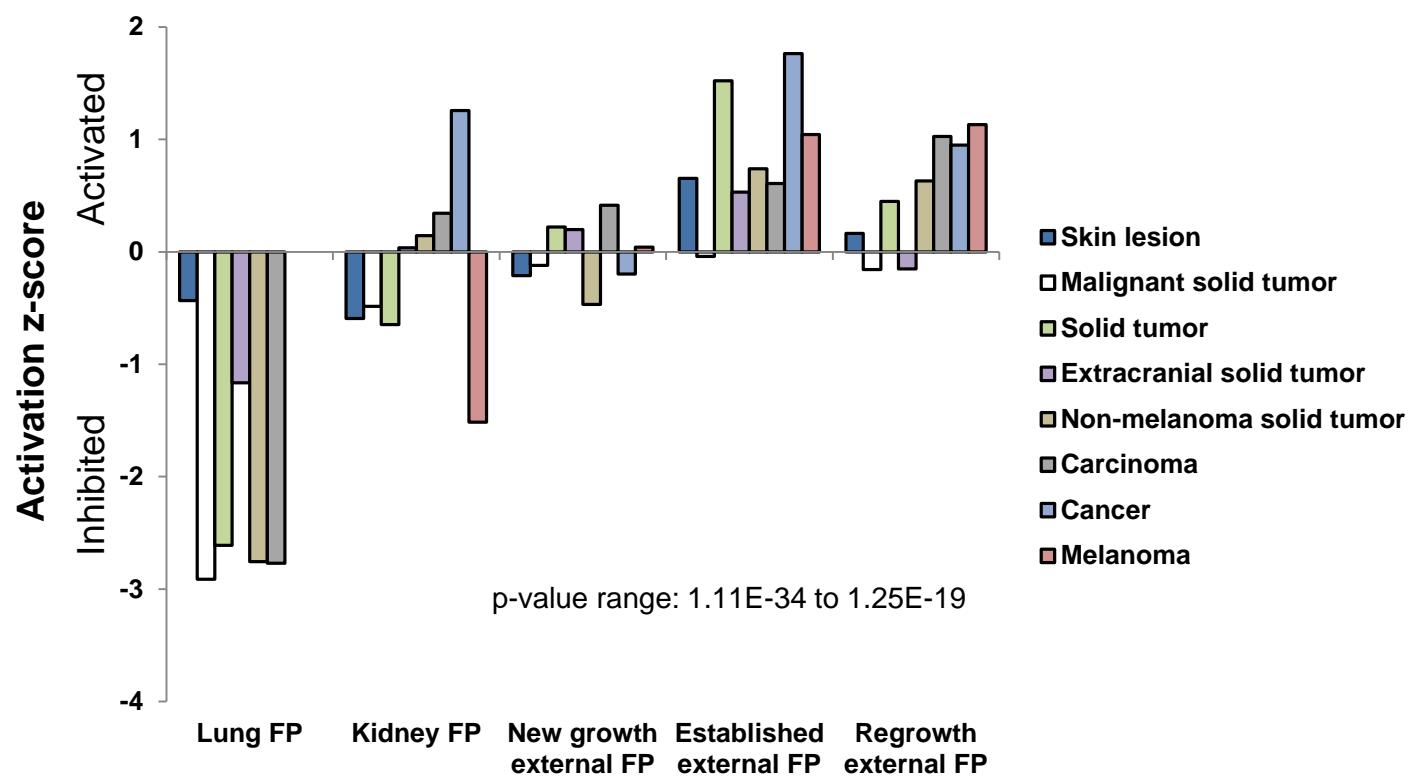
p-value range: 1.11E-34 to 1.25E-19

A

Figure 3

Overlap of top 200 GO terms (ranked by p value) of differentially expressed transcripts (Tumor Vs Non-tumor)

New growth external FP

Uniquely common to early FP (new and re- growth tumors)

28

13

22

137

25

34

16

**Regrowth external FP**

Established external FP
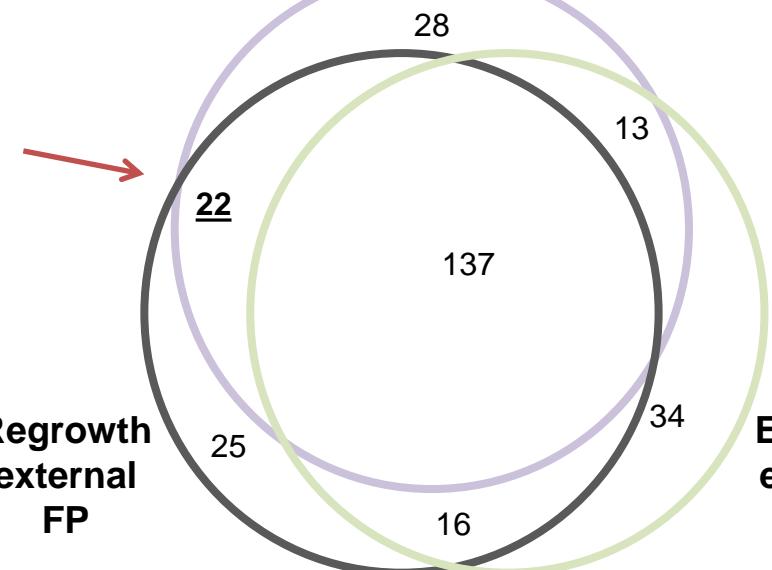
B

Leukocyte/lymphatic GO terms uniquely common to the top 200 new growth and regrowth FP GO terms (8 terms of the 22 common terms above)

p-value range: 6.84E-18 to 8.24E-08

Activation z-score — Activated / Inhibited

Legend:
- Lung FP
- Kidney FP
- Established external FP
- New growth external FP
- Regrowth external FP

X-axis categories: Quantity of lymphocytes, Quantity of T lymphocytes, Quantity of mononuclear leukocytes, Quantity of leukocytes, Quantity of lymphatic system cells, Cell movement of leukocytes, Function of leukocytes, Leukocyte migration

C

New growth tumor CD3+

New growth tumor secondary antibody control
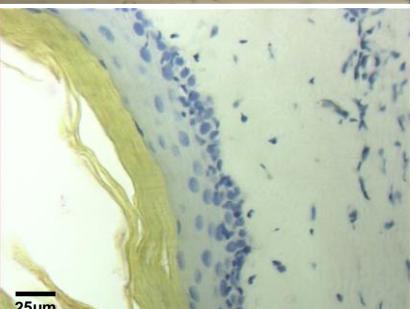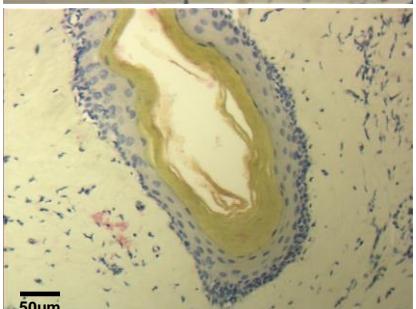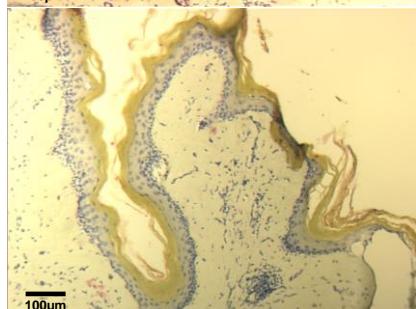
100μm 25μm 25μm

100μm 50μm 25μm

Fig. 4

**A** Top 100 ITRs Overlap (from of differentially expressed transcripts)

(Tumor Vs Non-tumor)
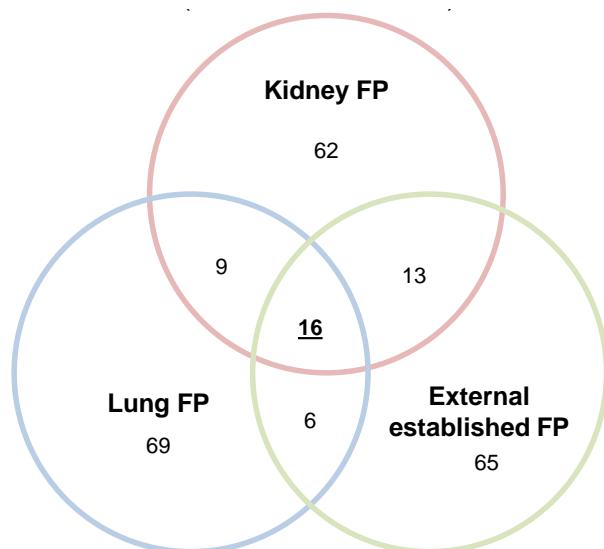
B Interaction networks of the 16 top ITRs (9 genes) common between lung, kidney and established external tumors

Core network

Expanded network
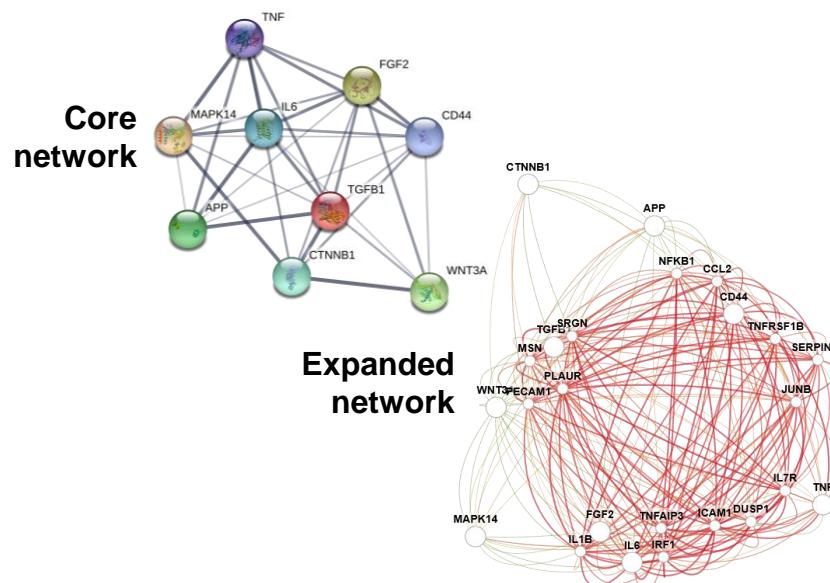
C Activation scores of the 16 ITRs (out of the top 100) common between lung, kidney and established external tumors

Genes

Drugs

Activation z-score

Activated

Inhibited

TGFB1, TNF, CD44, IL6, APP, WNT3A, CTNNB1 (β-catenin), P38 MAPK, FGF2

Immune-related

tretinoin (all-trans retinoic acid), U0126 (MEK [MAPK] inhibitor), SB203580 (p38 MAPK inhibitor), diethylstilbestrol (synthetic estrogen), lipopolysaccharide, dexamethasone (Anti-inflammatory), phorbol myristate acetate

p-value range: 2.64E-28 to 2.12E-03

□ Lung FP □ Kidney FP □ Established external FP □ New growth external FP ■ Regrowth external FP

D β-catenin (CTNNB1) staining in FP tumors
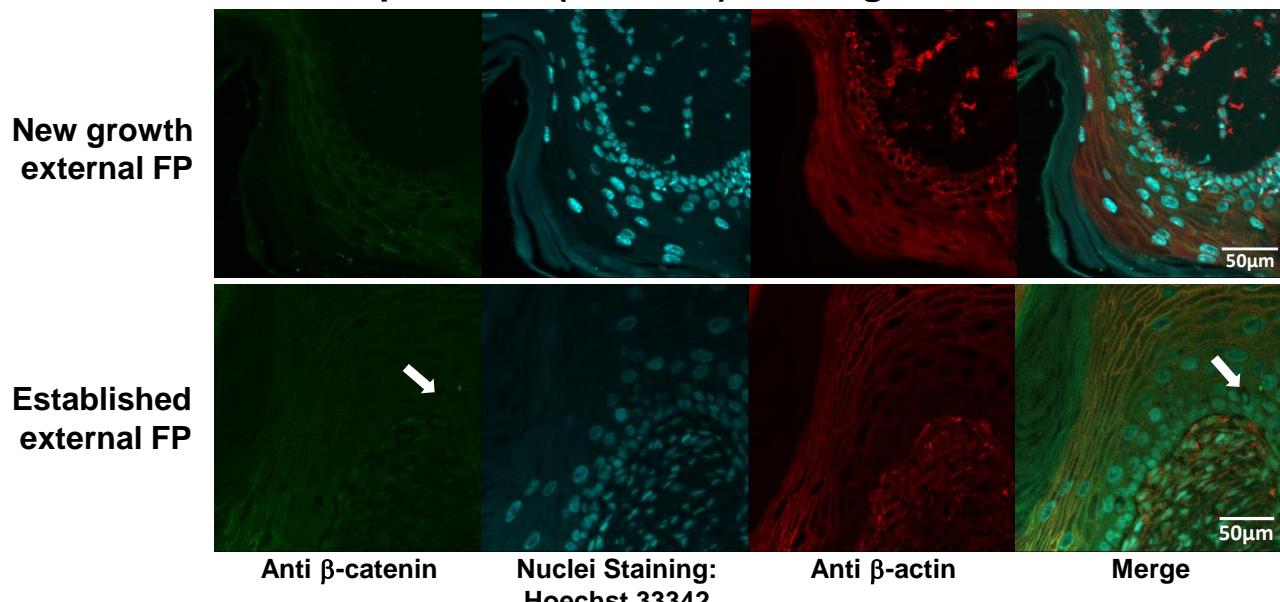
New growth external FP

Established external FP

Anti β-catenin Nuclei Staining: Hoechst 33342 Anti β-actin Merge

Figure 5

A Established external FP versus non-tumored skin top 200 ITR network (RNA-seq)

**B** Lung FP versus non-tumored lung tissue top 200 ITR network (RNA-seq)
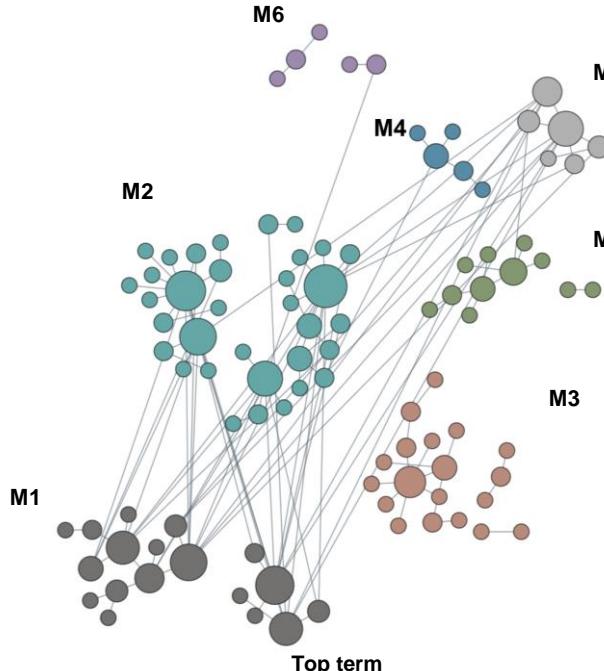
|	Top term	Q value
M1	Response to virus	<1e-04
M2	Inflammatory response	<1e-04
M3	T cell differentiation	<1e-04
M4	Negative regulation of anoikis	<1e-04
M5	Cellular senescence	<1e-04
M6	Positive regulation of interleukin-12	<1e-04
M7	Positive regulation of pre-miRNA transcription	0.0001

	Top term	Q value
M1	Leukocyte and neutrophil aggregation and migration	<1e-04
M2	Response to inorganic substance and metal ion	<1e-04
M3	Myeloid cell differentiation	0.0002
M4	Leukocyte activation	0.0003
M5	Cellular response to radiation	0.0004
M6	Response to virus	0.0008
M7	Activation of immune response	0.0008
M8	Positive regulation of angiogenesis	0.0009
M9	Positive regulation of blood vessel endothelia	0.0010

C

Colorectal Cancer Metastasis Signalling Pathway (RNA-seq)
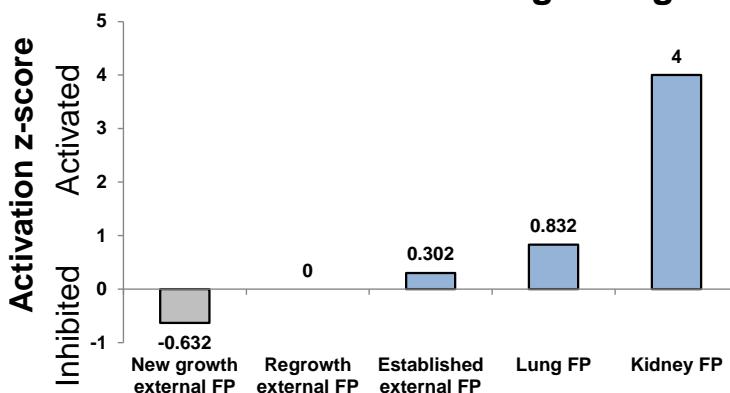
D

Invasion and Metastasis GO terms (RNA-seq)

p-value range: 2.52E-18 to 1.68E-04

Genome-wide copy number variation (CNV)

Figure 6

**A** Patient 27-2017-Cm, lung FP vs. non-tumor tissue (WGS, DNA-seq)
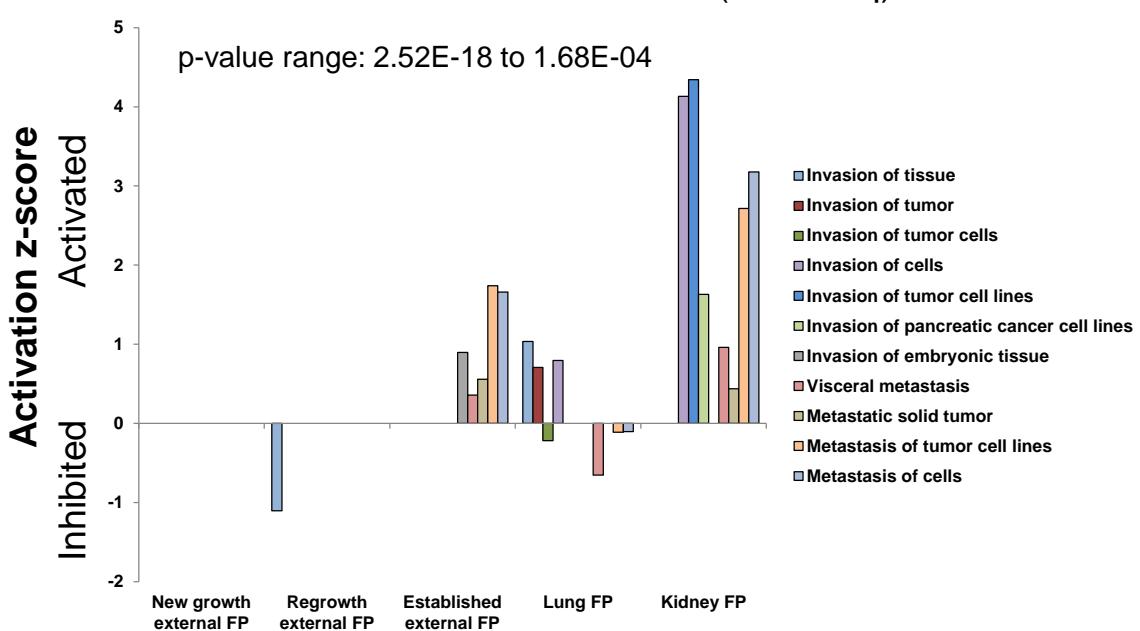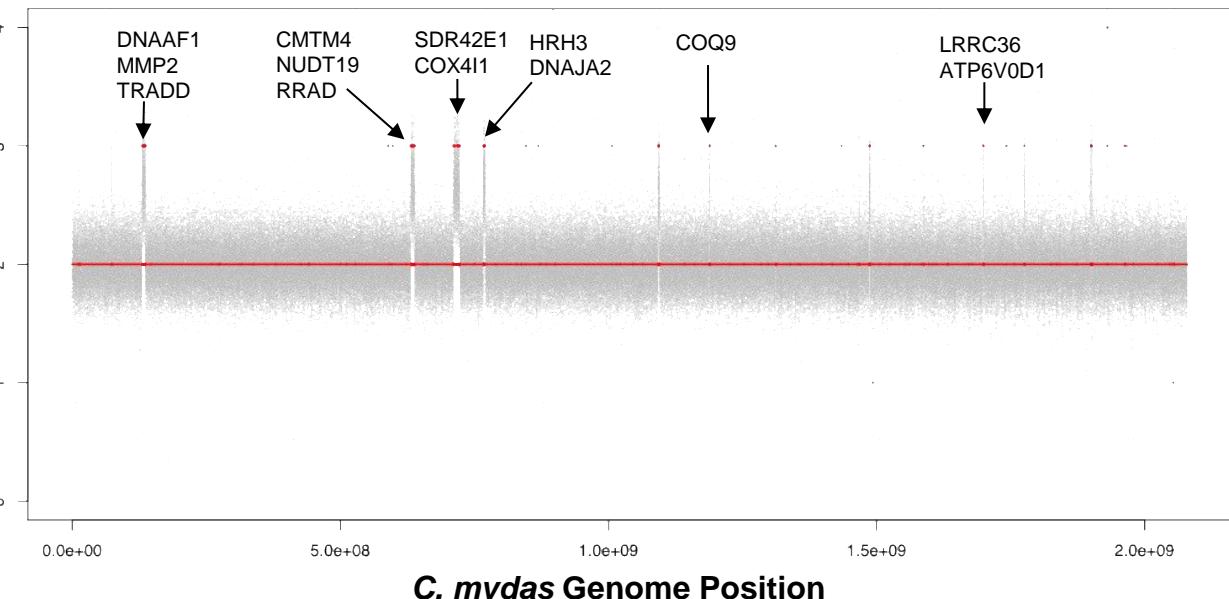

B Top 25 GO terms (ranked by p-value) of genes harboring CNVs in lung FP (DNA-seq of 27L1F)
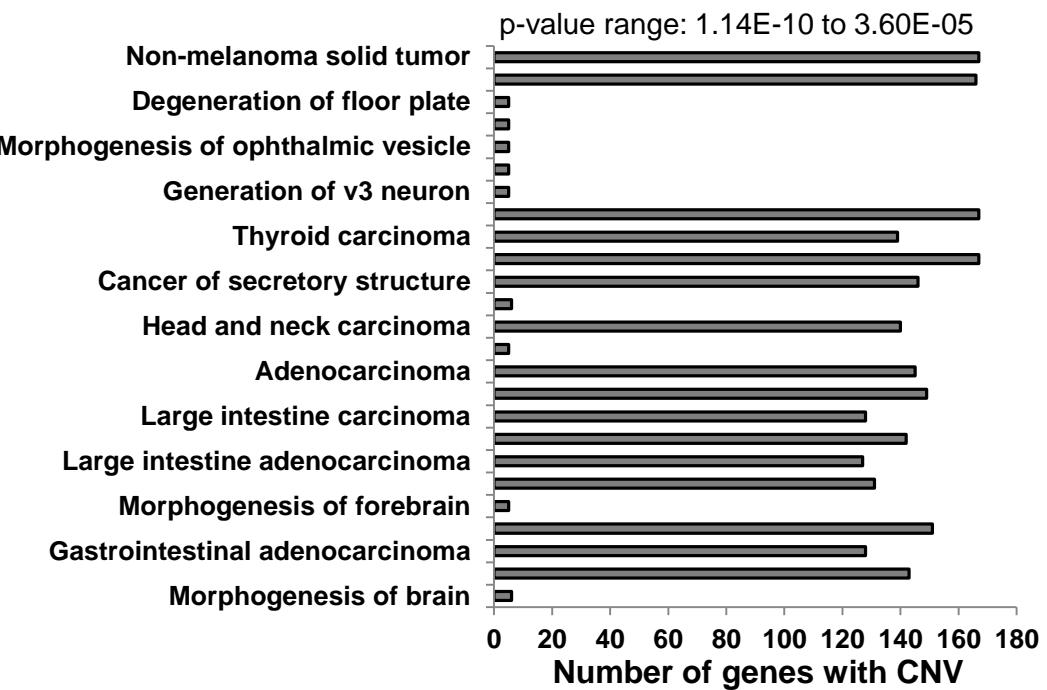
C MAPK-associated genes harboring CNVs in lung FP (DNA-seq of 27L1F)
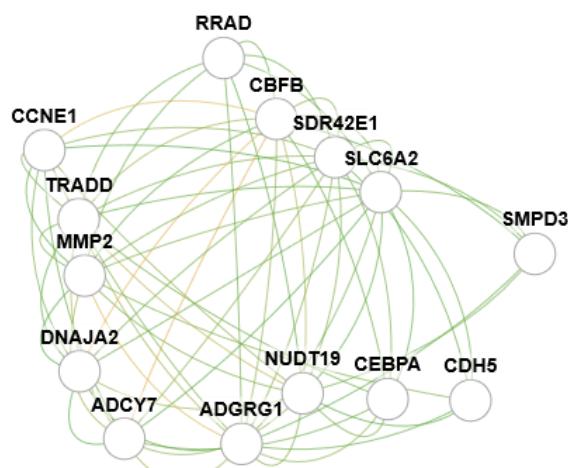
D Potential high dose sensitive genes, Patient 27-2017-Cm internal tumors (Expression change of genes harboring CNV, RNA-seq)
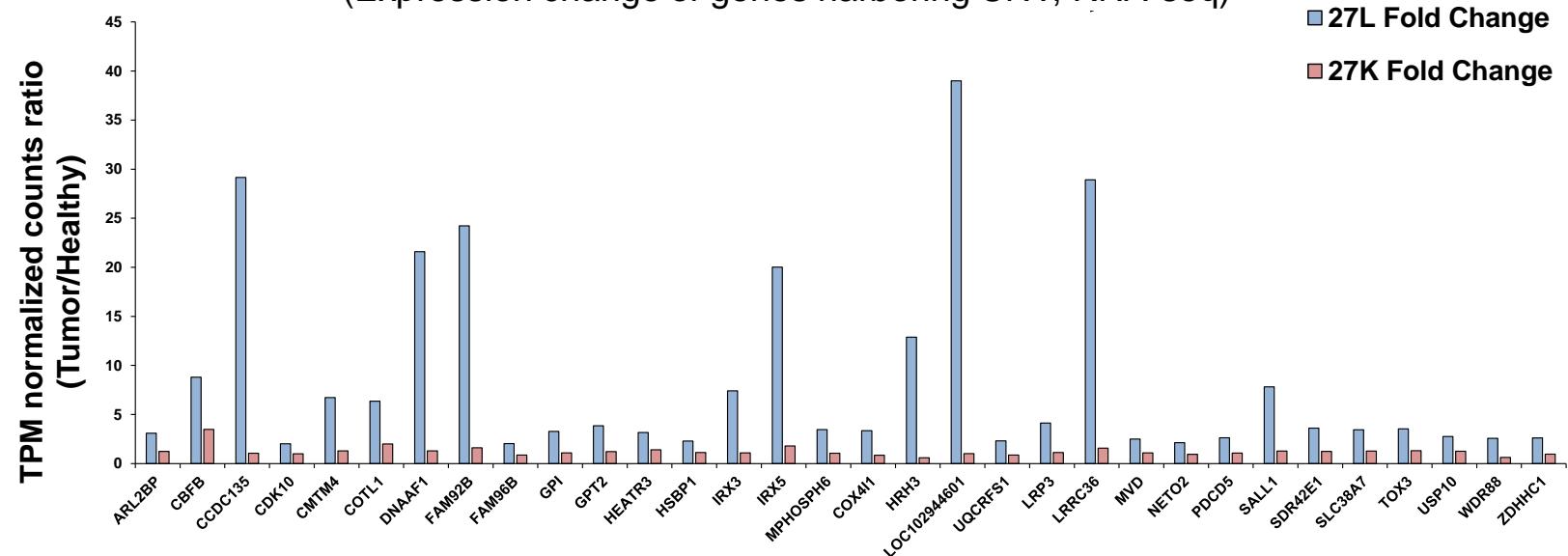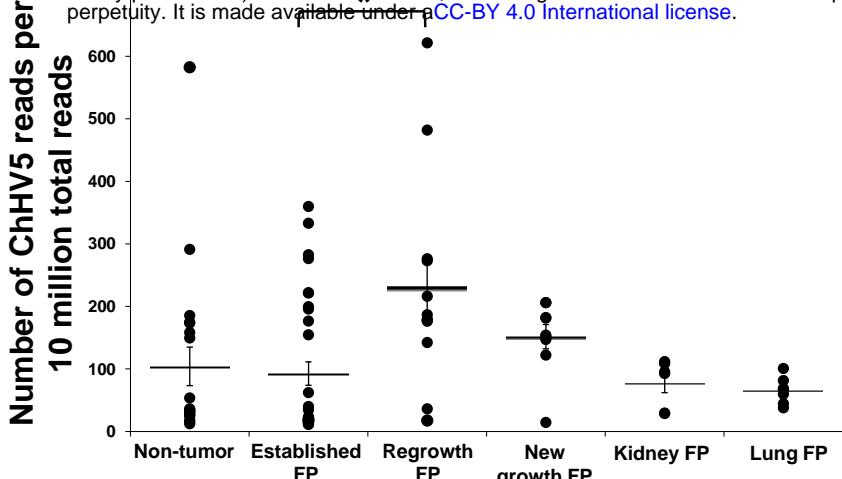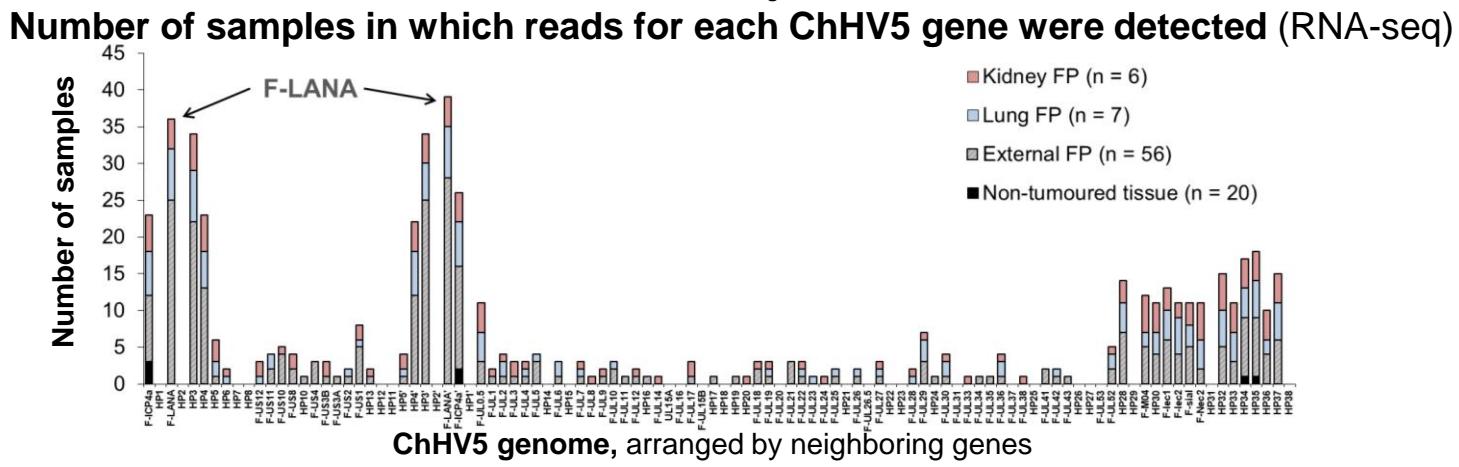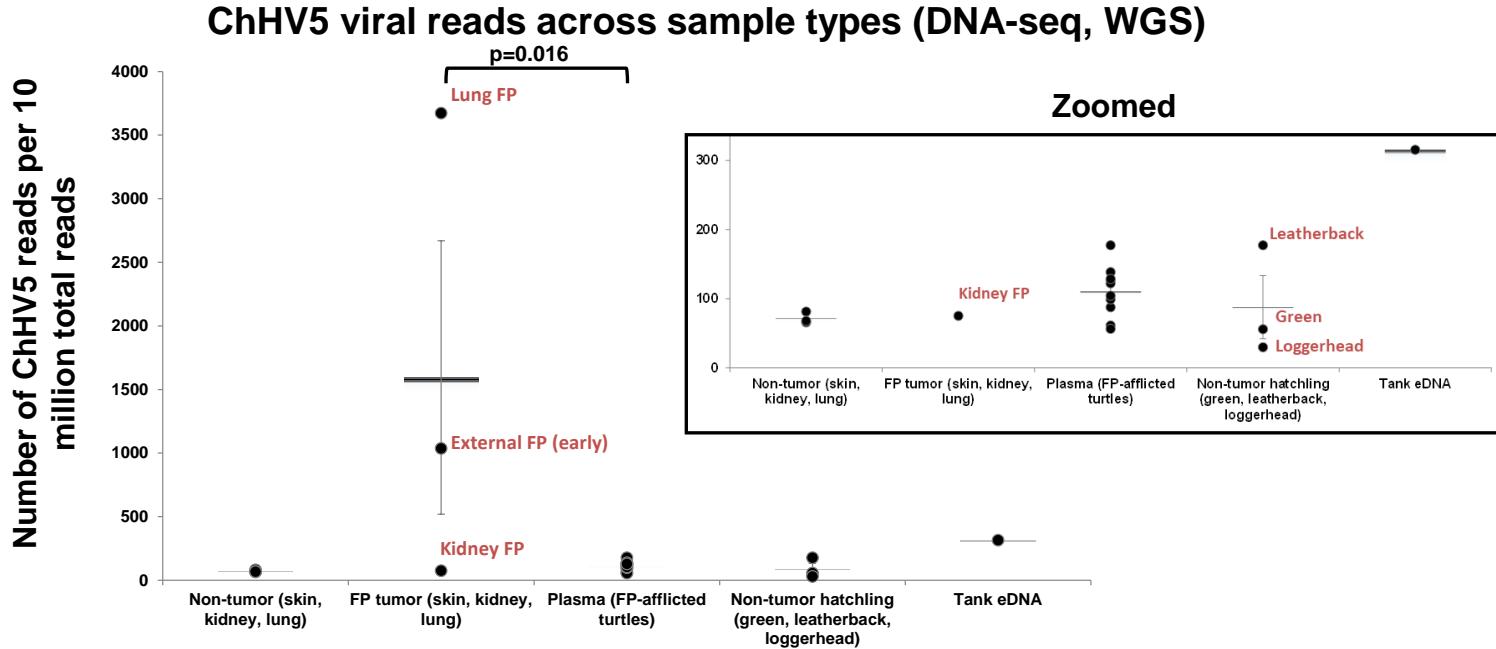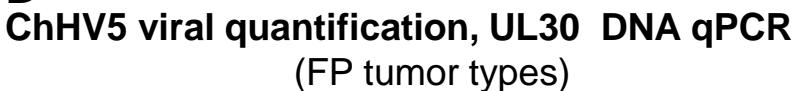
ChHV5 viral reads across sample types (RNA-seq)

A

Figure 7

B Number of samples in which reads for each ChHV5 gene were detected (RNA-seq)

C ChHV5 viral reads across sample types (DNA-seq, WGS)

D ChHV5 viral quantification, UL30 DNA qPCR (FP tumor types)

ChHV5 viral quantification, UL30 DNA qPCR (Non-tumor tissue types)

A

Figure 8

B

Viral shedding increases inline with tumor burden
(UL30 qPCR tank eDNA, versus tumor surface area)

$R^2 = 0.5431$

X-axis: Tumor Surface Area (mm^2)
Y-axis: ChHV5 Concentration (pg/µl)
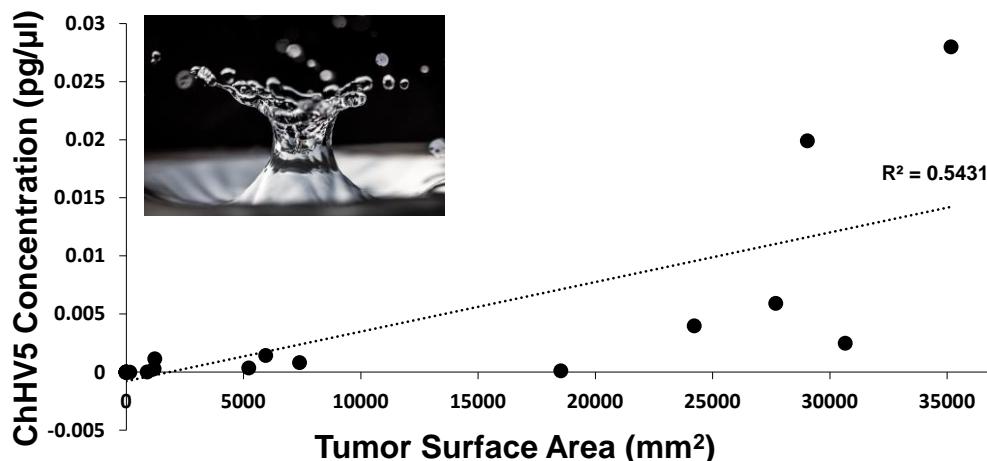
ChHV5 virus detectible from leeches
(UL30 qPCR, leech eDNA samples)

Y-axis: ChHV5 Concentration (pg/µl)

Categories: Leech, from no tumors C. caretta | Leech, from FP tumor C. mydas (Cm) | Cm FP tumor tissue, tissue that leech was removed from
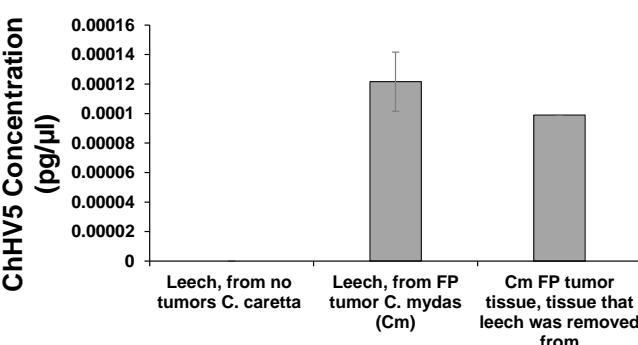
C

Viral shedding correlates to tumor removal
(ChHV5 UL30 qPCR, tank eDNA samples)

* Tumor removal surgery occurred

Left Y-axis: Tumor Surface Area (mm^2)
Right Y-axis: ChHV5 Concentration (pg/µl)
X-axis: 16-Oct-18, 01-Nov-18, 13-Nov-18, 27-Nov-18, 13-Dec-18

-●- Tumor surface area   -●- ChHV5 concentration
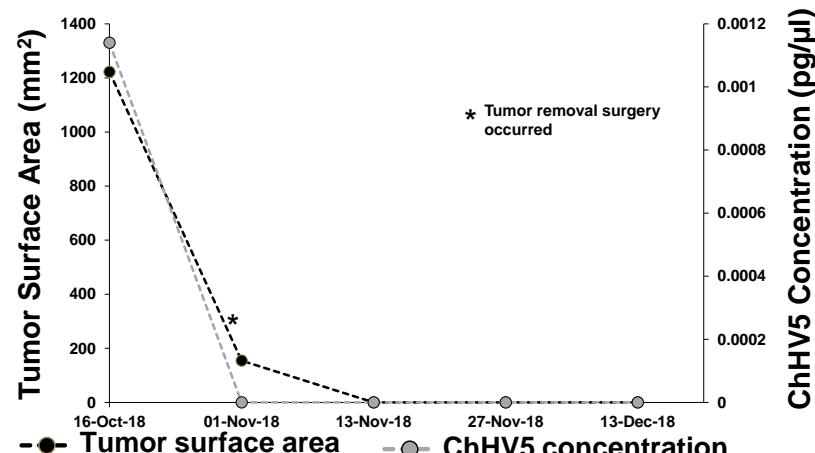
D

Viral shedding time-course
(ChHV5 UL30 qPCR, tank eDNA samples)

High tumor surface area individual

* Tumor removal surgery occurred

□ Week 1
□ Week 2
□ Week 3
□ Week 4
■ Week 5

FP-Free

Lower tumor surface area individual

Low tumor surface area to FP-free individual

Y-axis: ChHV5 Concentration (pg/µl)
Categories: Control (Milli-Q Water) | Tank 1 | Tank 2 | Tank 3 | Tank 4
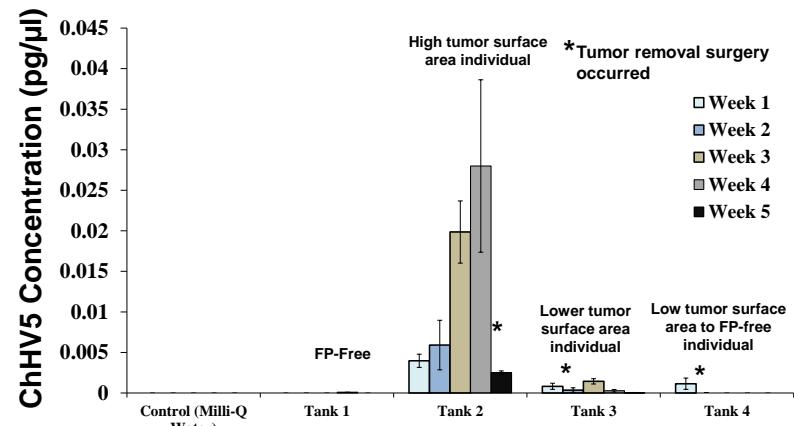
E

Correlation between ChHV5 and *C. mydas* eDNA concentrations in tank water
(ChHV5 UL30 eDNA versus *C. mydas* 16sRNA eDNA, qPCR)

$R^2 = 0.66$

Y-axis: ChHV5 Concentration (pg/µl)
X-axis: *C. mydas* concentration (pg/µl)