

RESEARCH

Open Access



# Spatial variation in gene expression of Tasmanian devil facial tumors despite minimal host transcriptomic response to infection

Christopher P. Kozakiewicz<sup>1\*</sup>, Alexandra K. Fraik<sup>1</sup>, Austin H. Patton<sup>1,2</sup>, Manuel Ruiz-Aravena<sup>3,4</sup>, David G. Hamilton<sup>4</sup>, Rodrigo Hamede<sup>4,5</sup>, Hamish McCallum<sup>6</sup>, Paul A. Hohenlohe<sup>7</sup>, Mark J. Margres<sup>1,8</sup>, Menna E. Jones<sup>4</sup> and Andrew Storfer<sup>1</sup>

## Abstract

**Background:** Transmissible cancers lie at the intersection of oncology and infectious disease, two traditionally divergent fields for which gene expression studies are particularly useful for identifying the molecular basis of phenotypic variation. In oncology, transcriptomics studies, which characterize the expression of thousands of genes, have identified processes leading to heterogeneity in cancer phenotypes and individual prognoses. More generally, transcriptomics studies of infectious diseases characterize interactions between host, pathogen, and environment to better predict population-level outcomes. Tasmanian devils have been impacted dramatically by a transmissible cancer (devil facial tumor disease; DFTD) that has led to widespread population declines. Despite initial predictions of extinction, populations have persisted at low levels, due in part to heterogeneity in host responses, particularly between sexes. However, the processes underlying this variation remain unknown.

**Results:** We sequenced transcriptomes from healthy and DFTD-infected devils, as well as DFTD tumors, to characterize host responses to DFTD infection, identify differing host-tumor molecular interactions between sexes, and investigate the extent to which tumor gene expression varies among host populations. We found minimal variation in gene expression of devil lip tissues, either with respect to DFTD infection status or sex. However, 4088 genes were differentially expressed in tumors among our sampling localities. Pathways that were up- or downregulated in DFTD tumors relative to normal tissues exhibited the same patterns of expression with greater intensity in tumors from localities that experienced DFTD for longer. No mRNA sequence variants were associated with expression variation.

**Conclusions:** Expression variation among localities may reflect morphological differences in tumors that alter ratios of normal-to-tumor cells within biopsies. Phenotypic variation in tumors may arise from environmental variation or differences in host immune response that were undetectable in lip biopsies, potentially reflecting variation in host-tumor coevolutionary relationships among sites that differ in the time since DFTD arrival.

**Keywords:** Transmissible cancer, Host-pathogen coevolution, Wildlife disease, Population transcriptomics

\* Correspondence: [Chris.Kozakiewicz@wsu.edu](mailto:Chris.Kozakiewicz@wsu.edu)

<sup>1</sup>School of Biological Sciences, Washington State University, Pullman, Washington, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Identification of the processes underlying cancer development as well as those associated with heterogeneity in its progression is critical for predicting both individual and population-level outcomes [1, 2]. Measuring relative levels of gene expression has been key for identifying genes and biological pathways associated with cancers and heterogeneity in cancer phenotypes [3–5]. For example, gene expression studies have shown regulatory dysfunctions associated with tumorigenesis [6], identified expression profiles associated with therapy resistance and poor prognoses [7–9], and provided insights into interactions between tumor cells and the immune system [10]. Similarly, gene expression analyses are frequently used to understand how organisms respond to environmental stressors, such as thermal stress [11], pollutants [12], and infections [13]. Variation in gene expression among individuals can also indicate how biotic and abiotic pressures underlie population-level responses. For example, temperature-dependent susceptibility of salamanders to *Batrachochytrium dendrobatidis* (Bd) is driven by temperature-mediated shifts in innate versus adaptive immune gene expression, enabling predictions of Bd effects on amphibian communities under different climate change scenarios [14]. The fields of oncology and wildlife disease intersect in the case of transmissible cancers, which are clonal, transmissible tumors that are spread among individuals by the transfer of cancerous cells [15].

Transmissible cancers are rare; the two most well-known examples are canine transmissible venereal tumor (CTVT) and Tasmanian devil facial tumor disease (DFTD), with other examples found in Syrian hamsters and mollusks [15, 16]. Whereas CTVT has relatively benign effects on its hosts both at the individual and population levels, DFTD is an aggressive, highly virulent pathogen with nearly 100% case fatality rate and has had devastating effects on Tasmanian devil populations [17–19]. Since its discovery in 1996, DFTD has spread throughout the entire geographic range of the Tasmanian devil, leading to rapid population declines exceeding 80% [20, 21]. Likely derived from a Schwann (peripheral nerve) cell cancer in a female devil [22], mutations leading to downregulation of MHC class I expression in the tumor [23, 24], coupled with potential natural killer cell dysfunction in devils [25], enabled DFTD to evade host allograft rejection and become transmissible [26, 27]. DFTD is transmitted via biting, a fundamental behavior in devil social interactions [28, 29], with DFTD tumors manifesting externally and predominantly around the mouth and face. Following visible presentation of tumor growth, progression is rapid, leading to death within 12 months [30].

Despite observations of frequency-dependent transmission leading to early predictions of devil extinction [18, 31], recent epidemiological models incorporating individual variation in pathogen load suggest that extinction is an unlikely outcome [32, 33]. Inter-individual variation in tumor growth rates and latency periods, together with a lack of vertical transmission, often enables females to survive through their first breeding season, allowing populations to persist at low densities despite high DFTD prevalence [32]. Female responses contributing to this success include increased precocial breeding and fecundity in the first breeding season in DFTD-affected populations, as well as tolerance of DFTD that manifests in a slower loss of body condition in DFTD-infected females relative to males [34–36].

Recent genomic evidence suggests rapid evolutionary responses of devils to DFTD. Genomic regions putatively associated with immune response, cancer resistance, and behavior appear to be evolving within as few as four generations (approx. 8 years) under positive selection in DFTD-affected populations [37, 38]. Females may also be adapting to DFTD through greater tolerance, with several associated genes of large effect detected in a genome-wide association study [39]. Although putative functions of candidate genes identified in these studies suggest biological functions that may underlie variation in host fitness, mechanistic differences in response to infection between sexes remain unclear. Additionally, isolated cases of spontaneous tumor regression have been observed in the field, the molecular underpinnings of which appear to be in regulatory regions because no non-synonymous substitutions have been found in either the devils or tumors [40, 41]. Further, phenotypic responses in devil populations following DFTD arrival have been observed within one or two generations [36], suggesting existing plasticity and not a purely adaptive response.

Studies comparing transcriptomic and genomic variation are ideally suited for elucidating the molecular basis for variation in devil responses to DFTD. However, previous gene expression studies of DFTD have aimed to identify cell type of origin [22] or targeted specific sets of immune-related genes, primarily in laboratory-cultured DFTD cell lines [e.g., 41–44]. Thus, there is a need to understand variation in both host and DFTD tumor gene expression in natural populations, particularly with respect to observed variation in DFTD-associated impacts on hosts.

To investigate the role of gene expression variation in manifestation of DFTD, we performed a population transcriptomics study of wild Tasmanian devils from DFTD-affected populations. We sequenced mRNA from normal tissues in both healthy and DFTD-infected devils, as well as tissues from DFTD tumors, and tested

predictions of differential gene expression with respect to sex and population that could explain individual- and population-level responses to disease. We predicted that 1) similar to other types of cancer, DFTD exhibits gene expression that is distinct from normal host tissues, 2) the severe disease associated with DFTD infection induces significant responses in infected hosts as evidenced by expression differentiation between infected and uninfected hosts, 3) previously observed differences in DFTD tolerance between host sexes will be reflected in inter-sex variation in gene expression, and 4) spatial genetic variation in both devils and DFTD produces variation in gene expression across geographic localities.

## Results

### Sequencing, alignment, and transcript assembly

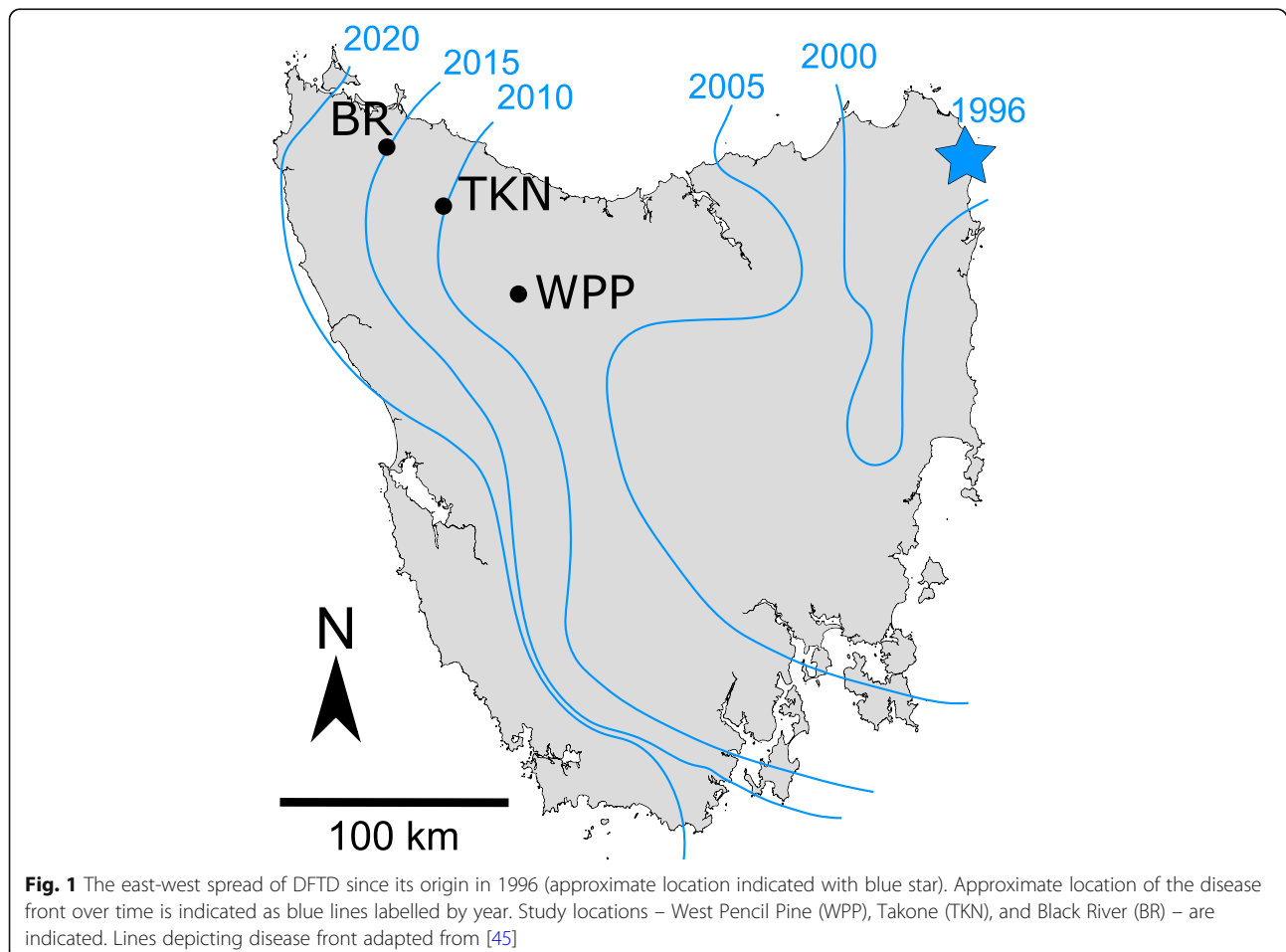
To identify expression differentiation across hosts and tumor and test the four predictions above, we sequenced mRNA libraries from 58 tissue samples from 39 devils (20 males, 19 females). Samples were collected from wild devils between 2016 and 2018 from three distinct localities (Fig. 1): Black River (BR; first infected in 2016), Takone (TKN; first infected in 2011) and West Pencil

Pine (WPP; first infected in 2006). Samples comprised lip biopsies from 20 devils putatively uninfected by DFTD, and paired lip and tumor biopsies from 19 devils (38 samples) clinically infected with DFTD. Volumes of sampled tumors did not significantly differ among localities (analysis of variance:  $P = 0.365$ ). We generated a total of 1,168,476,356 sequence reads, which were reduced to 1,167,199,873 following quality trimming and filtering (per sample mean = 20,124,136; SD = 5,905,604). Reads aligned to the reference genome at an average of 91.9% per library (SD = 3.4%), with 63.1% (SD = 3.9%) of bases mapping to annotated mRNA regions.

### Differential gene expression

#### Between devil lip and DFTD tissues

For the comparative analysis of differential gene expression between lip and tumor tissues, we retained a total of 14,807 expressed genes after filtering (Additional file 8: File S8). We found a dramatic difference in gene expression between lips and tumors (11,149 significant differentially expressed genes at a false discovery rate [FDR] of 0.05), which were clearly delineated as multidimensional scaling clusters (Fig. 2a; Additional file 5: Fig. S5).



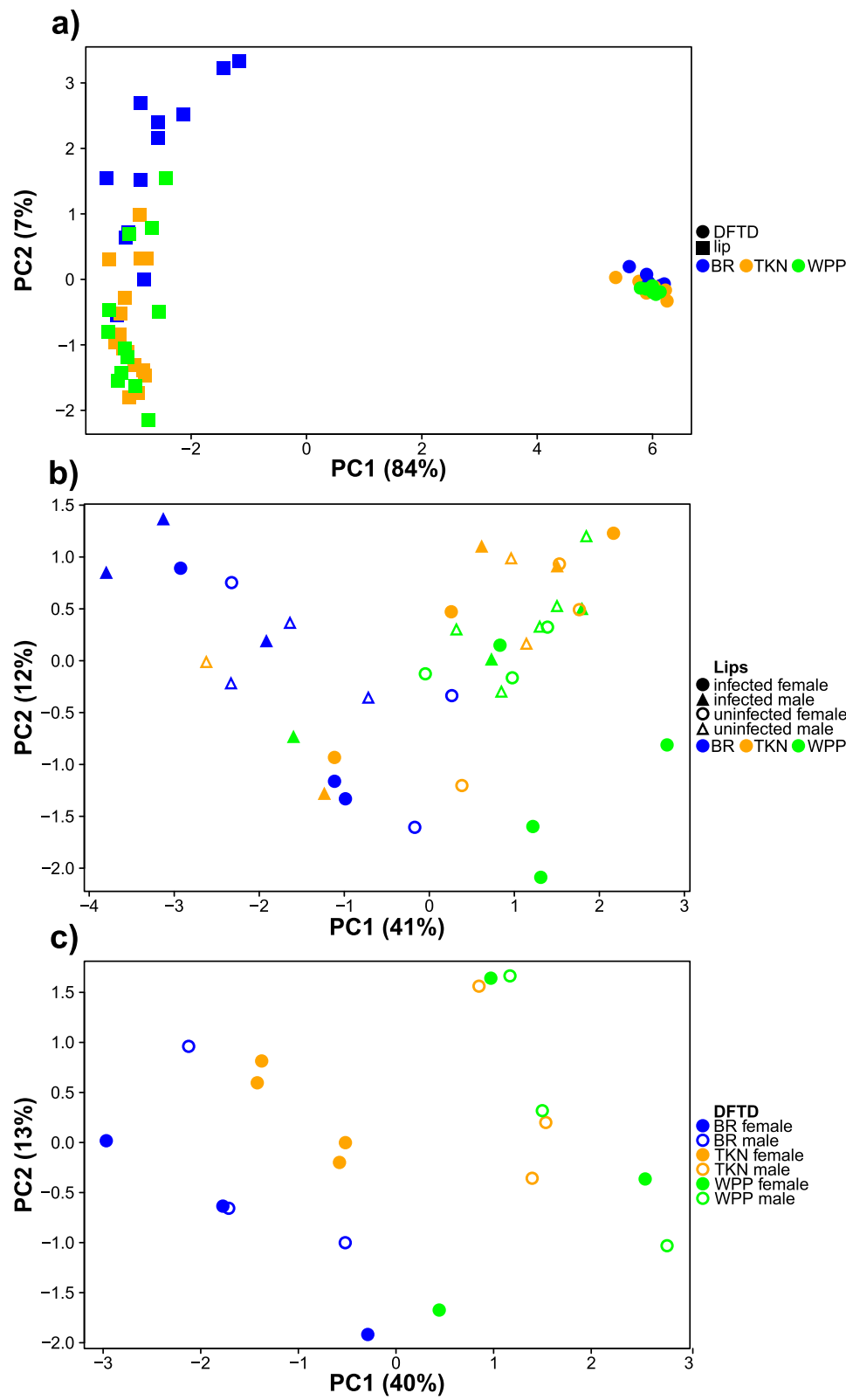


Fig. 2 (See legend on next page.)

(See figure on previous page.)

**Fig. 2** Variation in gene expression among a) DFTD tumor and Tasmanian devil lip biopsies, b) lip biopsies sampled from devils of different sex and DFTD infection status, and c) DFTD tumors sampled from devils of different sex and from different localities. Each plot was generated using the 500 genes exhibiting the most expression variation (i.e., the highest standard deviation across all samples) from each differential expression analysis

Because the number of differentially expressed genes for this contrast was so large, we applied a log<sub>2</sub> fold change (log<sub>2</sub>FC) minimum threshold of  $\pm 2$  to focus our efforts on only the most highly differentially expressed genes. With this more stringent filter, we identified 4234 genes that were still differentially expressed between tissue types, with 2286 upregulated and 1948 downregulated in DFTD tumors.

To evaluate variation among localities in genes differentially expressed between lip and tumor, we analyzed differential expression between lips and tumors for each locality separately and detected similarly high numbers of differentially expressed genes. We identified 4282 genes differentially expressed at log<sub>2</sub>FC  $\pm > 2$  in BR (646 DE genes unique to BR), 4238 in TKN (268 unique), and 4181 in WPP (350 unique).

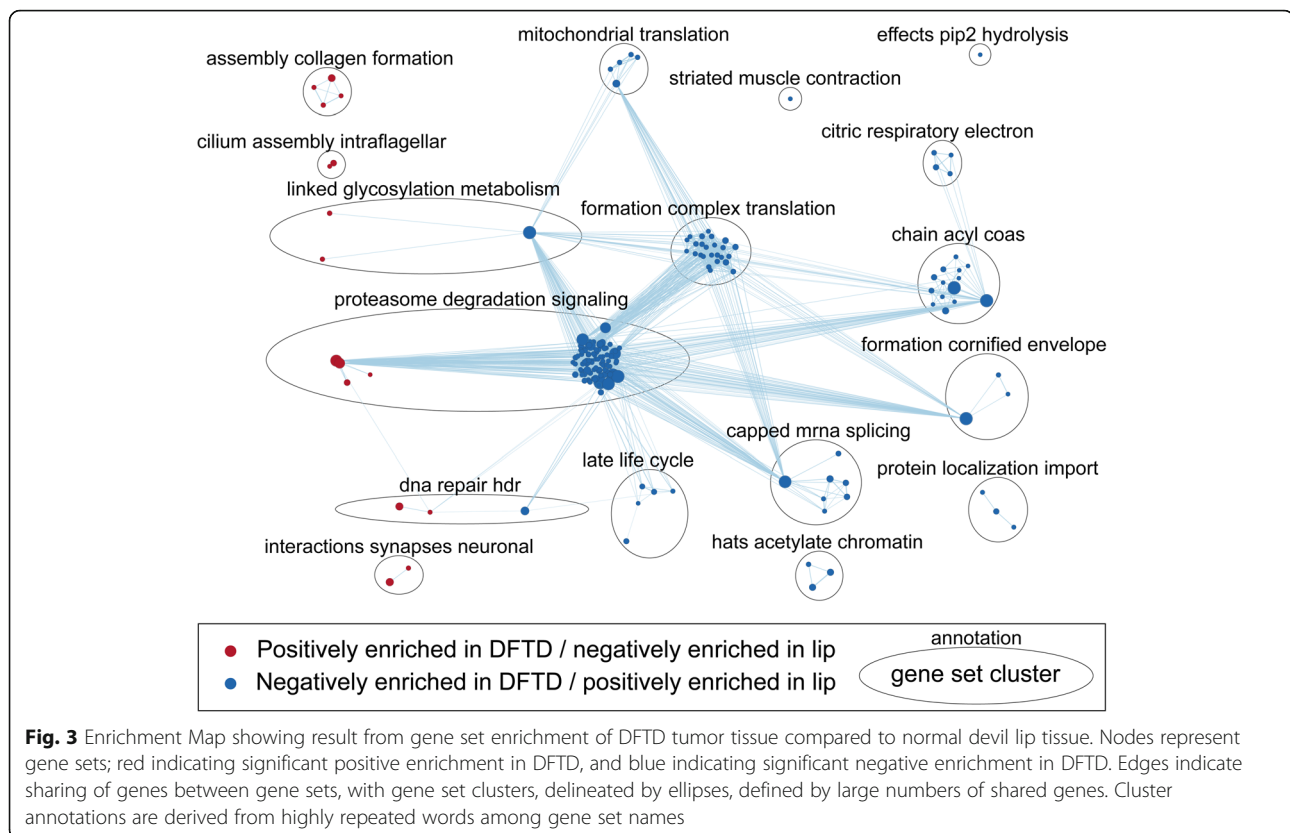
Using the PANTHER Overrepresentation Test [46], we identified numerous biological processes that were significantly enriched among genes that were highly differentially expressed between tumor and lip tissues (Additional file 6: Fig. S6). Many of these gene ontological (GO) terms were clearly associated with cellular differentiation due to the epidermal and skeletal muscle tissue present in lips as opposed to the Schwann cell (a peripheral nerve cell) origin of DFTD. For example, genes upregulated in tumors were enriched for nervous system processes and extracellular matrix organization but depleted in DNA damage response, transcription, and protein metabolism. Genes downregulated in tumors were enriched for muscle cell development and function but depleted in double-strand break repair, negative cell cycle regulation, transcription, and translation. Of genes that were differentially expressed between lips and tumors that were unique to a given locality, only BR yielded significantly enriched biological processes. Overrepresented processes were associated with regulation of cell shape and signaling, whereas underrepresented processes were associated with gene expression and metabolism. Full lists of significantly over- or underrepresented GO-terms are provided in Additional file 9: File S9.

Gene Set Enrichment Analysis revealed 16 positively and 163 negatively enriched Reactome pathways in DFTD at FDR = 0.01 (Fig. 3). Summarization via EnrichmentMap clustering highlighted a variety of generalized differences between tumor and lip tissues. The largest pathway gene set cluster contained 95 pathways and was

annotated “proteasome degradation signaling” (Fig. 3). This cluster contained only four pathways that were positively enriched in DFTD, which were associated with the mitotic cell cycle. The remaining pathways in this cluster were negatively enriched in DFTD, including many that were associated with the immune system and various signaling pathways such as Notch, slits and robos, and DNA damage checkpoints (both dependent and independent of P53). A closely related cluster annotated as “DNA repair HDR”, contained pathways associated with homology-directed DNA repair that were positively enriched in DFTD, as well as pathways negatively enriched in DFTD associated with transcriptional regulation by TP53 (Fig. 3). Clusters entirely positively enriched in DFTD were associated with collagen formation and extracellular matrix organization, cilium formation, and neuron function (Fig. 3; see Additional file 10: File S10 for full list of enriched gene sets and annotated gene set clusters).

#### Among devil lip tissues

For comparison among lip tissues only, we retained a total of 14,165 expressed genes (Additional file 8: File S8). Overall, we found little evidence of differential gene expression between lips from devils of different sex or based on infection status, irrespective of sampling locality. No genes were differentially expressed in any of our contrasts between DFTD-infected devils and those from clinically healthy devils, irrespective of host sex (Fig. 2b). However, when contrasting host sexes directly regardless of infection status, we identified seven differentially expressed genes. Specifically, we observed significant upregulation of *FRMD7*, *HMGB3*, *MECP2*, and three uncharacterized novel genes in females, and upregulation of one uncharacterized novel gene in males (see Additional file 13: Table S13 for full names and descriptions of gene symbols). No additional genes differentially expressed between males and females were identified when considering DFTD-infected and uninfected devils separately. Both groups exhibited significant upregulation in females of *FRMD7*, *HMGB3*, and two uncharacterized novel genes, whereas *MECP2* was upregulated in females only among uninfected individuals. Among localities, we identified 1564 genes differentially expressed between TKN and BR (897 up and 667 downregulated in TKN) but no evidence of differential expression in lips between any other pair of localities. Genes differentially expressed between TKN and BR were associated



with immune system processes and cellular developmental processes (upregulated in BR), as well as lipid metabolism (upregulated in TKN) (Additional file 9: File S9). Six gene sets were positively enriched in BR relative to TKN and were associated with phagocytosis and T-cell activation (Additional file 10: File S10). For all other lip-only contrasts, there were insufficient numbers of differentially expressed genes for GO-term enrichment analysis. Similarly, no significantly enriched pathway gene sets ( $FDR = 0.12-1.0$ ) were identified for any other lip-only statistical contrasts.

#### Among tumor tissues

In DFTD tumors, we retained 14,204 expressed genes following filtering (Additional file 8: File S8). Genes expressed in DFTD but not lip tissues, and vice-versa, were almost entirely linked to functions associated with the cell type of origin (i.e., Schwann vs muscular/epidermal cells). When contrasting tumors infecting devils of opposite sex, we found no genes that were significantly differentially expressed between sexes across all localities, but we observed some locality-specific differences. Although no genes were differentially expressed between tumors from different host sexes in BR, 17 genes were differentially expressed in tumors between sexes in TKN, all of which were upregulated in tumors from male devils relative to females: *SYNE2*, *CLUH*, *PARS2*,

*VPS13A*, *KIAA1586*, *TASOR2*, *SLC12A5*, *TRRAP*, *DIP2A*, *BRPF1*, *PCNX3*, and six uncharacterized novel genes. In WPP, only *MRPL53* was differentially expressed – upregulated in tumors infecting female devils.

Direct contrasts between localities – irrespective of host sex – revealed substantial differences in tumor gene expression. In general, the greatest difference was between BR and WPP, with TKN being somewhat intermediate of the two other localities (Fig. 2c, Additional file 7: Fig. S7). In all, 3825 genes were differentially expressed between BR and WPP, of which 1635 were upregulated and 2190 were downregulated in BR compared to WPP. In BR tumors relative to TKN, 414 genes were upregulated, and 548 genes were downregulated. In WPP tumors relative to TKN, 823 genes were upregulated, and 416 genes were downregulated.

All three pairwise contrasts of tumors between localities were significantly enriched for biological processes. All processes overrepresented among genes upregulated in BR relative to TKN were also overrepresented among genes upregulated in BR relative to WPP – these were broadly associated with translation and protein synthesis. Additional processes overrepresented among genes upregulated in BR relative to WPP were predominantly associated with other metabolic and biosynthetic processes. Genes downregulated in BR relative to WPP were disproportionately associated with chromosome

organization and gene expression. Similarly, in TKN tumors relative to WPP, downregulated genes were also associated with chromosome organization and gene expression regulation, whereas upregulated genes were associated with transcription, translation, and various metabolic and biosynthetic processes.

We identified significantly enriched Reactome pathway gene sets in comparisons between DFTD tumors sampled from different host sexes in TKN, as well as in comparisons among tumors from different localities generally. The most pronounced difference (in terms of both the number of enriched pathways as well as the magnitude of enrichment) was between BR and WPP, with TKN intermediate of the two but generally more similar to BR. Interestingly, the pathway gene set clusters positively enriched among tumors from WPP compared to other localities were characteristic of those positively enriched in DFTD compared to devil lip tissue. We identified five clusters of pathway gene sets that were significantly enriched in all three pairwise contrasts between localities (Fig. 4). The largest of these clusters – “regulation mediated degradation” – contained 119 pathways and was approximately equivalent to the largest cluster identified in the DFTD-lip contrast above. Accordingly, it contained pathways predominantly associated with cell cycle checkpoints and programmed cell death, and the innate and adaptive immune system; these pathways were all significantly negatively enriched in WPP tumors compared to those from TKN and BR. A small number of pathways in this cluster associated with the mitotic cell cycle were positively enriched in WPP compared to other localities (Fig. 4; see Additional file 10: File S10 for full list of enriched gene sets and annotated gene set clusters).

Of the three between-locality tumor contrasts, the TKN-BR contrast appeared to be the most distinct, with four pathway gene set clusters that were significantly enriched only in TKN-BR, while eight clusters were enriched in the other two between-locality contrasts but were not enriched in TKN-BR (Fig. 4). Clusters unique to the TKN-BR contrast included many pathways associated with the mitotic cell cycle (Fig. 4). Clusters that were lacking only from the TKN-BR contrast included pathways associated with protein folding and transport, metabolism, negative regulation of transcription, and autophagy (Fig. 4).

In general, we observed no enrichment of pathways among tumors sampled from devils of different sexes, either across all localities or within localities individually, with the exception of TKN. As reflected in Fig. 2c, gene set enrichment between tumors sampled from different sexes from TKN generally mirrored enrichment between WPP and BR, with TKN males negatively enriched for pathways that were also negatively enriched in WPP

when compared to BR. Overall, at FDR = 0.05, there were four pathways significantly positively enriched and 21 negatively enriched in TKN males compared to females. Pathways positively enriched in TKN males were associated with chromatin organization and mitotic prometaphase. Pathways negatively enriched in TKN males were associated with eukaryotic and mitochondrial translation and protein synthesis, nonsense-mediated decay, and respiratory electron transport.

### Genotypic variation and associations with gene expression

To identify genotypic variation potentially associated with variation in gene expression, we genotyped indels and single nucleotide polymorphisms (SNPs) from the transcriptome sequences. Following filtering (see Additional file 3: Text S3), we retained 9559 SNPs and 6013 indels in devils, and 473 SNPs and 1554 indels in tumors (including 100 alleles private to BR, 121 to TKN, and 112 to WPP) for analysis. For both tissue types, we found evidence for weak genetic structure when specifying sample localities a priori, though this structure was more pronounced in devils than in tumors (Fig. 5). However, when identifying genetic clusters purely from SNP variation, only one cluster each was supported for both devils and tumors.

We used SnpEff [47] to annotate all genotyped tumor variants according to their positions with respect to known genes and each variant's predicted impact on gene function. 4203 variant-gene effects were predicted in tumors, comprising 2264 variants associated with 1248 genes (note: loci may contain multiple variants, and variants may have multiple effects involving different genes). Of these variant-gene effects, 1514 involved variants located in intergenic regions, likely due to sequencing of noncoding mRNA, unannotated exons, or, potentially, contamination by genomic DNA. By comparison, similar proportions of intergenic variants have been detected in other RNA-seq studies [e.g., 48, 49]. Most other variants were also in non-coding regions: 159 were < 5 kb upstream, 20 were located in the 5' UTR, 211 were intronic, 392 were located in the 3' UTR, and 1617 were < 5 kb downstream. Only 290 variants were in exons and were not significantly enriched for any biological process. Of these exonic variants, 77 were identified as missense variants, seven were nonsense variants, and none were nonstop variants.

For 4088 tumor genes that were differentially expressed between any pair of localities, 444 contained variants with predicted effects on gene function. By comparison, 920 of 10,116 non-differentially expressed genes also contained variants with predicted effects, suggesting that variant-gene effects were equally likely to be detected within differentially expressed compared to

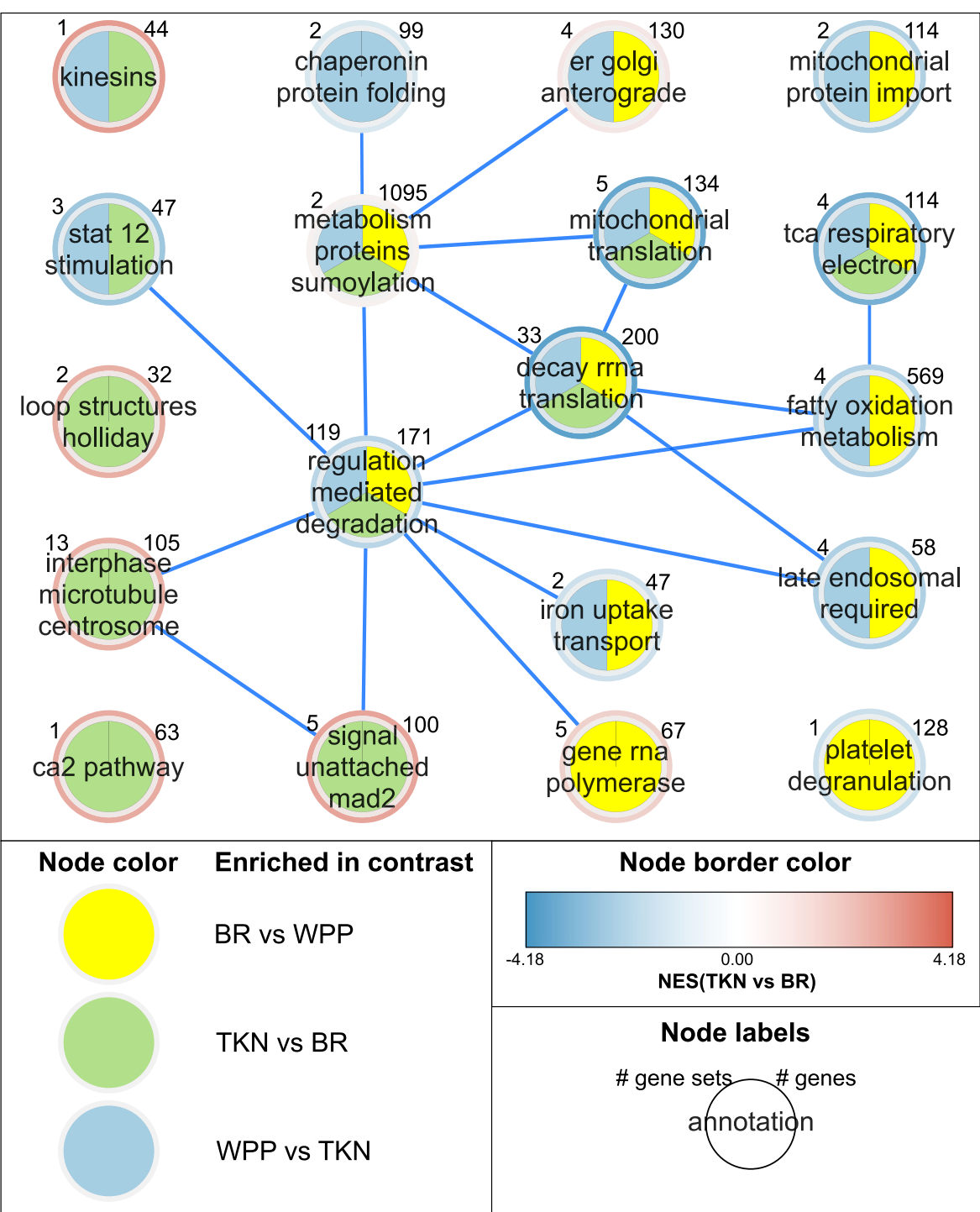


Fig. 4 (See legend on next page.)

(See figure on previous page.)

**Fig. 4** Three-way enrichment Map showing gene set clusters (nodes) that were significantly enriched in DFTD tumors between pairs of localities. Node colors indicate which between-locality contrasts were enriched for gene sets within each cluster (e.g., clusters showing all three colors contain gene sets that were enriched in all three pairwise contrasts between localities). Node border colors indicate average normalized enrichment scores (NES) for gene sets in TKN tumors relative to BR tumors. Edges represent genes shared between gene sets from different clusters. Clusters are annotated and labelled according to the number of genes and gene sets they contain. Cluster annotations are derived from highly repeated words among gene set names

non-differentially expressed genes. Following filtering for biallelic variants, we retained a total of 413 gene-expression-locus combinations for Kruskal-Wallis tests. Including loci with multiple predicted effects, these comprised 238 intergenic, 27 upstream, three 5' UTR, 47 intronic, 88 3' UTR, 357 downstream, and 89 exonic variants. However, after FDR adjustment, we found no significant associations between any variant on the expression of the genes it was predicted by SnpEff to effect.

## Discussion

As predicted, substantial differentiation exists in gene expression between lip and tumor tissues. We observed differential expression patterns consistent with DFTD's origin as a Schwann cell tumor [22], including genes involved in nervous system functions such as synaptic signaling and ion transport, as well as characteristics common to cancers [5]. Lips, in contrast, had differentially expressed genes characteristic of muscle and epidermal tissue function. Overall transcriptional variation appeared to be higher among lip tissues compared to that observed among tumor tissues, potentially due to greater genetic variation among devils or exposure of lip tissues to greater environmental variation. However, in lip tissues, we found surprisingly little transcriptional variation associated with sex or infection status, despite clear sex-biased differences in resistance and tolerance of DFTD [34, 39]. Perhaps most interestingly, we found significant variation in gene expression among DFTD tumors collected from different localities, potentially associated with different tumor strains, environmental effects (either within-host or extrahost) or host-pathogen co-evolutionary relationships. However, we found no association between detected SNP variants in transcribed genomic regions and gene expression, consistent with results from previous studies that showed variants in regulatory regions [40], which are not captured in RNA-seq approaches, likely affect observed phenotypic variation.

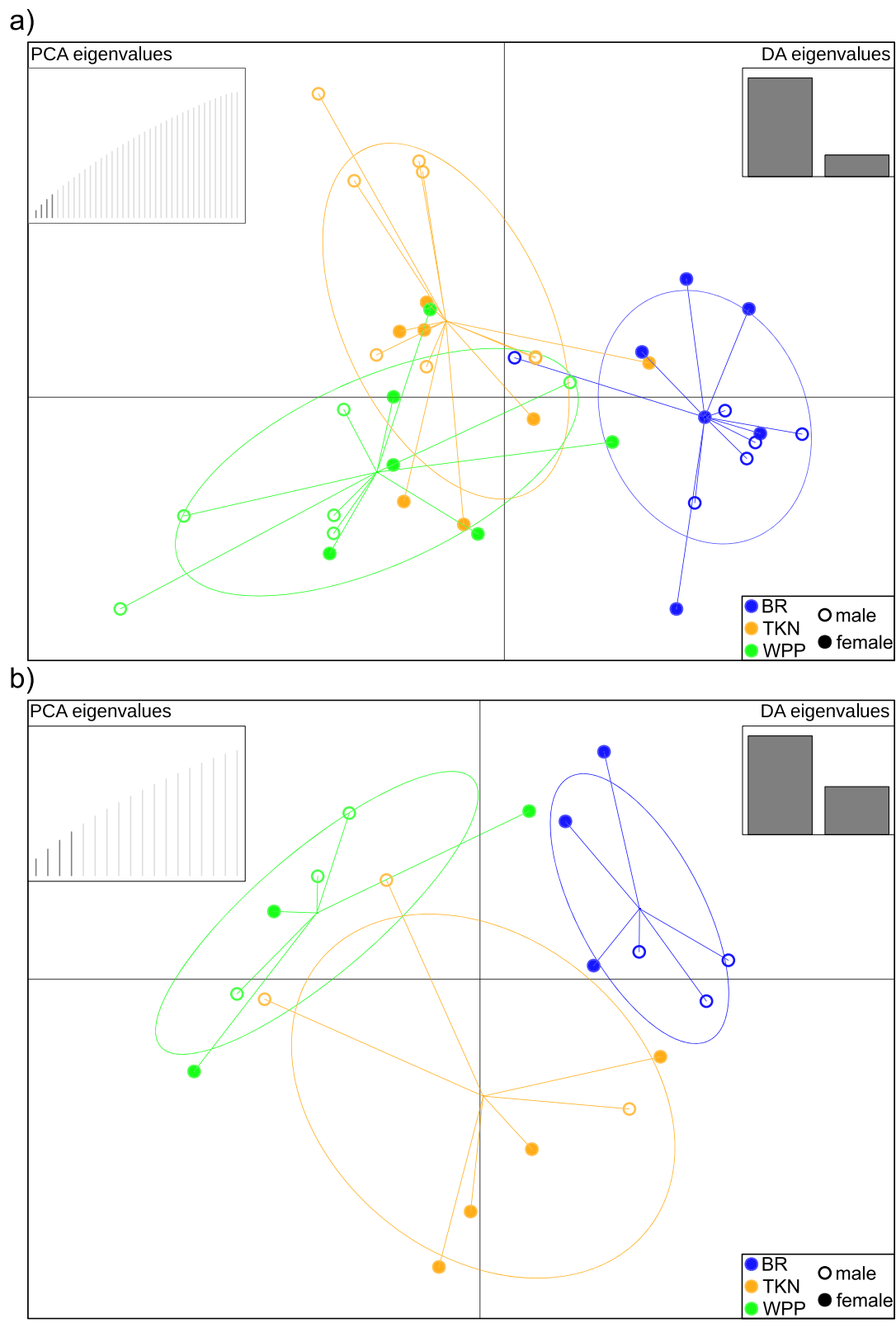
### Gene expression characteristics of DFTD

Despite considerable genotypic and phenotypic diversity among cancers [50], there are relatively common mutational drivers that are necessary for the genesis and proliferation of cancerous cells. Functions of these drivers

include: the sustenance of proliferative signaling, evasion of growth suppressors, resistance to apoptosis, induction of angiogenesis, activation of invasion and metastasis, reprogramming of energy metabolism, and evasion of immune response [51, 52]. Similarly, we observed differential expression of genes and enrichment of pathways in DFTD associated with the extracellular matrix, the cell cycle, DNA replication and repair, and immune function. The extracellular matrix is fundamental to maintaining tissue homeostasis. Dysregulation of the extracellular matrix can both cause – and occur as – a response to the development and proliferation of cancerous cells, facilitating uncontrolled proliferation, angiogenesis, tissue migration and invasion, and metastasis [53–55]. Cell cycle ‘checkpoints’ detect DNA damage or errors in DNA replication and chromosome organization and are regulated by critical tumor suppressor genes that prevent proliferation of defective and potentially cancerous cells [56]. A key example of such a gene is *TP53*, which is activated in response to DNA damage, initiates the inhibition of cell proliferation, and regulates apoptosis [57]. Loss of *TP53* function is common to many cancers [57, 58]. We found various pathways associated with *TP53*, such as G1 and G2 cell cycle checkpoints, to be downregulated in DFTD relative to normal lip tissue. In turn, upregulation of pathways associated with mitosis and the cell cycle that likely lead to increased cell proliferation were also found in DFTD.

Despite downregulation of cell cycle checkpoint and apoptosis pathways, we observed upregulation of homology-directed DNA repair (HDR) in DFTD. Dysfunctional HDR often results in genomic instability and can lead to carcinogenesis [59, 60]; however, cancers can become resistant to standard cancer treatments aimed at inducing DNA damage through reactivation of HDR, emphasizing that cancer cell proliferation is not dependent on HDR suppression [61]. Active yet dysfunctional HDR can lead to errors in double-stranded break repair that produce gross chromosomal rearrangements [62], such as those that are evident in DFTD [21, 24, 63].

Although all cancers exhibit some form of immune avoidance, DFTD and other transmissible tumors are unique in that they can evade host MHC recognition of non-self-cells [27]. In DFTD, MHC avoidance is achieved through the ERBB-STAT3 axis [44]. Specifically, DFTD likely became transmissible when a normal



**Fig. 5** Discriminant analyses of principal components, showing weak genetic structure in a) Tasmanian devils and b) devil facial tumor disease from West Pencil Pine (WPP), Takone (TKN), and Black River (BR)

Schwann cell tumor gained a variant exhibiting overexpression of *ERBB3*, which overactivates the transcription factor *STAT3* and blocks the production of MHC I [26, 27, 44]. Without MHC I expression, DFTD cells are unable to be recognized as foreign by the host. Similar to previous work, we found the differential expression of other genes associated with *STAT3*, including upregulation of *MMP2*, *HDAC5*, and downregulation of *PTGIS* [44]. However, in contrast to this previous study that found upregulation of the *TRIM28* protein, which is typically activated in response to *STAT3* signaling, we detected slight downregulation in expression of this gene. *TRIM28* is often overexpressed in cancer [64]; however, its expression has been shown to be predictive of tumor class in human glioblastomas [65], and positively correlated with tumor size and development stage, whilst being negatively correlated with patient survival in human hepatocellular carcinoma [66]. Lack of detectable *TRIM28* overexpression may thus reflect size or developmental stage variation among the sampled DFTD tumors.

We also observed downregulation of Notch signaling, which is critical to Schwann cell development and required for differentiation of Schwann cell precursors (SCP). Notch activation upregulates *ERBB2*, which acts as a receptor for *NRG1*, which in turn is critical for transitioning SCPs to immature Schwann cells [67, 68]. In the absence of Notch signaling, a scarcity of *ERBB2* receptors would lead to reduced sensitivity of SCPs to *NRG1*, despite its overexpression. Dysfunction of these pathways suggests a decoupling of DFTD cell proliferation from typical Schwann cell developmental controls.

#### Gene expression in devil lip tissue is not associated with infection status or sex

DFTD infection has no effect on gene expression in devil lip tissue. This is surprising as DFTD has significant effects on host physiology and can elicit host immune response [69]. Tumor growth leads to increasing metabolic demands concurrent with difficulty feeding, ulcerations, metastases, and secondary infections that produce a progressive loss of body condition and almost universal mortality within 12 months following visible tumor development [18, 30, 34]. In humans, gene expression in normal tissues adjacent to tumors reflects a state that is intermediate between healthy and cancerous, with commonly expressed pathways including pro-inflammatory responses induced by the tumor [70]. We chose lip tissue due to its proximity to the mouth, where DFTD allografts typically implant, believing that changes in gene expression would be greatest in tissues local to tumors. However, DFTD tumors do not always occur on the lips or in the mouth, and there was likely variation in the proximity of lip biopsies to the site of tumor

growth. Ethical and experimental concerns necessitated a consistent sampling strategy for healthy tissues, preventing individual adjustments of biopsy locations to accommodate variation in tumor location. In addition, lip tissues may have been inappropriate for detecting systemic immunological or metabolic changes. For example, systemic immune responses associated with tumor growth tend to involve the accumulation of immune cells in the peripheral blood or lymphoid tissues, rather than in the epidermis [71]. Biological functions associated with immune response were not underrepresented in lip tissues, suggesting that the lack of differential expression was not specifically due to a lack of immune expression in these tissues.

Despite a lack of a direct immune response in devil lips to DFTD infection, we observed upregulation of immune-associated genes in BR relative to TKN devils, irrespective of infection status. This difference is not necessarily associated with DFTD and may be due to differential exposure to other infectious agents or innate differences in immune function between genetically distinct populations. However, DFTD is an overwhelmingly strong selective force in devils [37, 39, 72] and likely drives immune adaptation in affected populations. Further, documented immune responses to DFTD [69] suggest that differential host immune expression between localities may alter the microenvironment faced by invading tumor cells. We thus recommend further gene expression studies targeting immunologically active host cells; specifically, to investigate systemic host responses to DFTD using blood samples as well as a refined approach for detecting localized responses by targeting healthy tissues < 1 cm from DFTD tumors.

Male and female devils exhibit different responses to DFTD, with females losing body condition at a significantly slower rate when infected and genetic evidence of adaptations that result in greater survival rates among females [34, 39]. Although DFTD infection produced no transcriptomic response in lip tissues for either sex, we found several genes that were differentially expressed between males and females generally, regardless of infection status. Consistent with previous work comparing sexes in uninfected devils [73], the X-linked gene *FRMD7* was downregulated in males ( $\log_2FC = -4.01$ ). *FRMD7* is putatively involved in fatty acid metabolism and has been associated with skin disorders, serving as a potential factor in differential susceptibility between sexes to DFTD [73]. Additionally, we found six other genes that were differentially expressed between sexes. Four of these were uncharacterized, while *HMGB3* and *MECP2* – both also X-linked – were downregulated in males. *HMGB3* is a DNA-binding protein that helps maintain stem cell populations and is overexpressed in some human cancers via the Wnt signaling pathway [74,

75]. Further, HMGB3 affects nucleic acid recognition and innate immune system activation, and its upregulation has been associated with allograft rejection [76–79]. Therefore, higher baseline expression of *HMGB3* in female devils may enhance the innate immune response to DFTD relative to males. Recently, *MECP2* was identified as an oncogene through induction of the MAPK and PI3K growth factor signaling pathways and is overexpressed in many human cancers [80]. However, it is unclear how *MECP2* expression in normal host tissues could affect DFTD progression.

#### DFTD tumor gene expression varies geographically

We observed considerable variation in tumor gene expression among the sampled localities, which varied in the length of time since DFTD arrival. Interestingly, gene expression patterns that we observed in DFTD relative to host lips were more intense (i.e., more differentially expressed genes and greater log2 fold-changes) in WPP tumors – where DFTD has been present longest among our study sites – than in other populations. Specifically, WPP tumors exhibited upregulation of mitosis and downregulation of translation, DNA damage checkpoints, and immune function relative to tumors from BR, which had the shortest time since DFTD arrival. Varying intensity of DFTD-characteristic gene expression may be due to differences in the ratios of normal-to-tumor cells within tumor biopsies, potentially driven by subtle differences in tumor morphology among localities (e.g., differences in the extent to which tumor tissue is delineated from the surrounding host tissue, or the distinctness of tumor margins). On the other hand, expression changes in cancer-associated genes are not only linked to tumorigenesis but can be directly correlated with tumor aggressiveness and overall prognoses in human cancer patients [81, 82]. Meanwhile, incrementally greater gene expression changes through time can produce progressively more severe phenotypes or reflect more advanced stages of tumor development [e.g., 83, 84]. The strength of expression in DFTD-associated pathways therefore may be associated with DFTD phenotypic variation, such as growth rates, although no such differences (nor differences in tumor morphology) between the studied localities have been documented.

Gene expression in tumors from TKN reflected an intermediate state between WPP and BR but were differentiated by sex (Fig. 2c; Additional file 7: Fig. S7). That is, tumors from TKN males exhibited patterns of gene expression that were characteristic of WPP tumors from both sexes, and tumors from TKN females resembled male and female tumors from BR (see Fig. 2c). Perhaps coincidentally, previous work demonstrating more rapid body condition loss in DFTD-infected males than in females was conducted in TKN [34], while no similar

analysis of host body condition has been performed for either of our other study sites. We do not have sufficient serial volume measurements for the tumors in our study to directly associate gene expression with tumor growth rates. In the absence of data comparing tumor growth rates or host body condition between TKN, WPP, and BR, it is difficult to establish a link between relative levels of DFTD-characteristic gene expression and the severity of disease.

Although it remains unclear whether transcriptional variation in DFTD occurs among localities as a result of neutral or adaptive differences among strains, there is an interesting temporal trend that correlates with expression changes. DFTD first arrived in WPP in 2006 and has thus had the longest amount of time to adapt to devils in that locality. DFTD then emerged in TKN in 2011, and tumors there showed intermediate expression values between WPP and BR, where DFTD has been for the shortest time (since 2016). Thus, expression variation among study sites may reflect ongoing DFTD adaptation to devils within populations, which themselves have demonstrated rapid evolutionary responses to DFTD [37, 39]. Additionally, multiple DFTD lineages with differing degrees of pathogenicity have been observed. Specifically, in WPP, DFTD arrival in 2006 was characterized by initially low mortality rates compared to localities that had experienced DFTD for longer [85]. Karyotype analysis confirmed the existence of a distinct tetraploid strain at WPP that was replaced by a more virulent diploid strain in 2012, causing an immediate increase in disease prevalence and population decline [86]. In addition, recent phylodynamic analysis indicates multiple tumor lineages exhibiting differences in transmission rates, demonstrating epidemiological variation among distinct tumor strains [87]. Given its recent emergence, our BR tumor samples may represent a novel DFTD lineage present near the advancing disease front. Such a lineage was not observed at a broad spatio-temporal scale [87] but more intensive sampling of the most recently emerged tumors near and on the west coast of Tasmania provides evidence of spatially structured tumor lineages in this region [88]. Specifically, WPP contained tumors from multiple distinct lineages, whereas tumors sampled from further west (closer to the vicinity of our TKN and BR sites) almost entirely comprised a single lineage [88]. Spatial structuring of tumor lineages may therefore explain the observed transcriptional variation among localities.

We detected weak genetic structure in tumors that broadly reflected transcriptomic variation among localities and may reflect different tumor strains. However, we acknowledge that the persistence of a small number of host variants in the tumor dataset may produce a residual signal of host population structure, despite

rigorous bioinformatic filtering to exclude known devil variants genotyped from lip RNAseq data as well high-coverage whole genome sequences (see Additional file 3: Text S3). Further, although we identified a number of variants in tumors that were predicted to affect the function of genes differentially expressed among localities, no significant associations were detected. Thus, transcriptional variation in tumors may be purely regulatory in origin, or we may have lacked sufficient statistical power (19 samples) to identify genotype-expression associations. Alternatively, population genetic structure and adaptive and plastic responses to the local environment in hosts can produce variation in immune function that drive differential responses to wildlife disease [14, 89]. Our study sites have similar vegetation communities and experience similar climatic conditions but decrease in elevation from east to west. Region-specific adaptation of devils to local environmental conditions exists, as well as significant selection imposed by DFTD following its arrival in naïve populations [37, 39, 72]. Different transcriptomic responses of DFTD to devils from different localities may thus also be explained by immunological variation among devil populations that is driven by environmental differences that lead to differential exposures to non-DFTD infectious agents.

## Conclusions

Characterization of DFTD gene expression and our broad analysis of transcriptomic variation in devils and DFTD points to a potential link between the relative degrees of DFTD-characteristic gene expression patterns and host population. Whether this variation is due to distinct tumor strains with differing phenotypes or by locally specific differences in host tolerance/resistance remains unclear. Nonetheless, geographic variation in gene expression among tumors highlights the potential for ongoing devil-tumor evolution. DFTD arrived at each of the populations in this study at different times, and thus the coevolutionary processes that influence the nature of host-tumor molecular interactions may be at different stages for each. Alternatively, or perhaps concurrently, environmental variation among localities may also be influencing spatial variation in tumor gene expression. To separate these alternative explanations would require replicated sampling of time points with respect to DFTD arrival across different localities. Given the well-documented east-to-west spread of DFTD over a 25-year period, this system is highly suited to exploration of host-pathogen coevolutionary relationships at various times since disease arrival.

## Methods

### Sample collection

Lip and tumor tissues were collected as 3 mm biopsies from live, wild Tasmanian devils between 2016 and 2018

from three locations in northern Tasmania, Australia: Black River (BR), Takone (TKN), and West Pencil Pine (WPP) (Fig. 1; Additional file 1: Table S1). Detailed descriptions of ethically-approved trapping protocols can be found in Hawkins et al. [19] and Hamede et al. [86]. All research involving use of animals and field activity was performed under University of Tasmania ethics approval A13326 and WSU IACUC approval ASAF 6796. Normal (i.e., noncancerous) tissue from inside of the lip was chosen due to its proximity to where DFTD tumor allografts typically implant, increasing the likelihood of detecting localized tissue responses in DFTD-infected animals, while ensuring both animal and handler safety. Where individuals had multiple tumors, biopsies were taken from the largest tumor. All tumor biopsies were taken from ulcerated tumors and comprised only solid tissue that was free from necrosis or secondary infection. Sampled tumors were measured, and tumor volumes were calculated using the ellipsoid volume equation ( $volume = \frac{4}{3} \times \pi \times \frac{length}{2} \times \frac{width}{2} \times \frac{depth}{2}$ ). Among-population differences in tumor volume were evaluated using an analysis of variance. For each population, lip biopsies from at least 3 DFTD-infected and 3 uninfected devils of each sex were collected, along with corresponding tumor biopsies from every infected devil. All biopsies were immediately preserved in RNAlater, kept at  $-20^{\circ}\text{C}$  for up to two weeks while in the field, and subsequently stored at  $-80^{\circ}\text{C}$  until RNA extraction.

### RNA extraction, library preparation, and sequencing

We extracted whole RNA from lip and tumor samples in two batches. The first batch contained samples collected in 2016, and RNA extraction was performed using a combination of Nucleospin RNA extraction kit (Macherey-Nagel, Easton, PA, USA), Qiagen Allprep DNA/RNA Mini Kit (Qiagen Inc., Hilden, Germany) and a standard Trizol-chloroform protocol. The second batch contained samples collected in 2018, and RNA extraction for this batch was performed using the Trizol-chloroform protocol. All extracted RNA was treated with DNase prior to library preparation to remove DNA contamination.

We prepared and sequenced mRNA-seq libraries in three batches. Batches 1 ( $n = 4$ ) and 2 ( $n = 12$ ) comprised samples collected in 2016; we prepared these libraries using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs, Ipswich, MA, USA) according to Fraik et al. [73]. Batch 3 ( $n = 42$ ) comprised all samples collected in 2018, with library preparation performed by the Washington State University Genomics Core in Spokane using the Illumina TruSeq Stranded mRNA Library Prep Kit and assessed for quality using a Fragment Analyzer (Agilent, Santa Clara,

CA, USA). Both library preparation kits employ a PolyA tail selection approach that isolates mRNA from ribosomal and globin RNAs. For each batch, samples were pooled together and sequenced for 100 bp paired-end reads using an Illumina HiSeq 2500: batch 1 on a single lane, batch 2 across two lanes, and batch three across four lanes. All sequencing was performed by the Washington State University Genomics Core, Spokane, WA. We accounted for potential batch effects attributable to use of different kits and sequencing lane effects throughout all analyses (see below).

### Sequence alignment and transcript assembly

We conducted initial quality screening of raw sequencing reads using FastQC and assessed study-wide sequencing quality using MultiQC. To trim and filter reads for quality, we used the TrimGalore wrapper with relatively relaxed settings due to the documented negative effects of stringent filtering on RNAseq analyses [90]. Specifically, we trimmed adapter sequences and ends with a Phred quality score < 10 and removed reads < 30 bp long.

We performed sequence alignments and assembled transcripts according to protocols recommended by Pertea et al. [91]. Using Hisat2 v 2.1.0 [92], we aligned the trimmed reads to the published Tasmanian devil reference genome (Murchison 2012), specifying the `--dta` flag to produce alignments suitable for transcript assembly. Aligned reads were checked for DNA contamination, firstly by visualization using the Broad Institute's (Cambridge, MA, USA) Integrative Genomics Viewer, and then by using the DepthOfCoverage function from the Genome Analysis Toolkit (GATK) v 4.1.8.1 to quantify the proportions of different genomic regions (e.g., exonic, intergenic) that were covered by at least one sequencing read (see Additional file 12, Table S12 for results). We used samtools to sort the alignments prior to assembling the transcripts using Stringtie v 2.1.0 [93] and the Ensembl reference annotation v 7.0.97 [94]. Then, using Stringtie, we merged all transcripts into a non-redundant assembly and used gffcompare v 0.11.7 [95] to annotate the merged assembly and examine how the assembled transcripts compare with the reference annotation. We then repeated the transcript assembly for each sample, using the merged annotation as the reference, and generated tables of transcript abundances including only transcripts that appeared in the merged reference annotation.

### Differential expression analysis

Prior to differential expression analysis, tables of transcript abundances were converted into read tables of gene-level abundances using the R package tximport [96]. All differential expression analyses and preparatory

steps were conducted in R using EdgeR [97] unless otherwise indicated, with the entire pipeline run three times as distinct analysis sets: 1) tumor-lip contrasts to investigate differential expression in tumors compared to host lip tissues; 2) lip-only contrasts to investigate variation in gene expression among hosts (e.g., with respect to sex and/or DFTD infection status); 3) tumor-only contrasts to investigate variation in expression among tumors (e.g., with respect to host sex or locality). For the tumor-lip contrasts, all lip and tumor transcriptomes were analyzed together, while the lip-only contrasts only included lip transcriptomes and the tumor-only contrasts only included tumor transcriptomes.

To ensure our analysis contained only genes that were expressed across experimental groups, we removed genes lacking transcript counts > 10 in at least three samples. Read counts were then normalized among libraries using the `calcNormFactors` function. We included the following factors in our analysis: *sex* (of the host devil, in the case of tumors), *infected* (i.e., with or without DFTD), *tissue* (normal lip or DFTD tumor), *population* (sampling locality; BR, TKN, or WPP), and *batch* (batches 1 or 2 comprising 2016 samples, or batch 3 comprising 2018 samples). Differential expression analyses were performed using a quasiliikelihood generalized linear model framework [98], using a different model formula depending on the tissues being analyzed. To simplify the design for each model and more easily facilitate comparisons between groups of potentially interacting factors, two factors of interest were combined into a single *group* factor. For the lip-tumor and lip-only models, this *group* factor comprised *sex* and *infected* (female-uninfected, female-infected, male-uninfected, or male-infected). For the tumor-only model, the *group* factor comprised *sex* and *population* (female-BR, female-TKN, female-WPP, male-BR, male-TKN, or male-WPP). For all models, to account for any effect that differences in lab protocols (see above) may have had on gene expression measurements, we included *batch* as a factor. Accordingly, we specified the formula for the host lip tissue vs DFTD tumor tissue model as:

$$0 \sim \text{tissue} + \text{group} + \text{population} + \text{batch}$$

The model formula for comparisons among lip tissues only was:

$$0 \sim \text{group} + \text{population} + \text{batch}$$

The model formula for comparisons among tumor tissues only was:

$$0 \sim \text{group} + \text{batch}$$

For each model, we calculated log gene counts per million (logCPM) and constructed multidimensional

scaling (MDS) plots using the R package *limma* [99] to identify clustering of samples according to our factors of interest. We observed MDS clustering of samples with respect to sequencing batch; this effect was subsequently corrected in all data visualization using the *limma* function *removeBatchEffect* to remove any variation in logCPM associated with the *batch* factor (Additional file 4: Fig. S4). MDS plots represented the top 500 differentially expressed genes defined as those with the largest standard deviations between all samples.

To identify differentially expressed genes between lip and tumor tissues, we performed a single contrast comparing all lip samples with all tumor samples, together with three separate contrasts comparing lip samples with tumor samples within in each locality. For the tissue-specific models, we performed a number of different contrasts: among lip samples, we performed 10 contrasts comprising various combinations of sex and infection status (Additional file 2: Table S2); among tumors, we performed seven contrasts comprising various combinations of host sex and sampling locality (Additional file 2: Table S2). Each contrast produced a table of results for all expressed genes quantifying the magnitude of differential expression between groups (log2 fold change; log2FC) and the significance of this difference given as nominal and false-discovery-rate-adjusted [FDR; 100] *P* values. We defined significantly differentially expressed genes at a false discovery rate threshold of 0.05, with no cutoff for log2FC. To verify effective control of batch effects, all differential expression results were compared with repeated analysis of batch 3 samples only (Additional file 11: File S11).

### Enrichment analysis

For contrasts with large numbers of differentially expressed genes, we performed a PANTHER Overrepresentation Test to test for biological processes (as defined by the PANTHER GO-Slim Biological Process annotation dataset) that are overrepresented among sets of differentially expressed genes. This comprises a Fisher's Exact Test for GO-terms that are over- or underrepresented relative to a 'background' reference list, which we specified as the full list of expressed genes from which the differentially expressed genes were detected. Significantly enriched GO-terms were identified at an FDR threshold of 0.05. We performed overrepresentation tests on sets of significantly over- and under-expressed genes separately as this has been shown to be more powerful for detecting differentially expressed pathways than analyzing all differentially expressed genes together [101]. To reduce redundancy among significantly enriched GO-terms and thus reduce the size and increase interpretability of our GO-term lists, we used REVIGO [102] to cluster highly related GO-terms

together and identify single representative GO-terms for each cluster.

To test for general patterns of up- or downregulation of biological pathways, we performed Gene Set Enrichment Analyses [GSEA; 103, 104] using pre-ranked lists of genes from differential expression analyses. Each expressed gene, irrespective of whether it was significantly differentially expressed, was ranked using a metric that measures the magnitude of the gene's log2FC as well as the nominal statistical significance of that change, calculated as  $\log_2FC * -\log_{10}(P)$ . This produced a list whereby, for a given comparison between conditions, significantly upregulated genes had higher positive scores, significantly downregulated genes had lower negative scores, and genes that were neither up- nor downregulated had scores close to zero. Given such a ranked list, GSEA tests for known sets of related genes that have disproportionately higher or lower rankings, calculating an enrichment score (ES) and FDR-adjusted *p*-value for each gene set tested. We performed the analysis using GSEA v4.0.3 [103, 104], testing gene sets from the most recent Reactome pathways database [105]. All GSEA contrasts were performed across 1000 permutations and restricted to gene sets containing between 15 and 5000 genes, with the 'meandiv' normalization mode selected. For contrasts with relatively few (< 50) differentially expressed genes, we used the weighted enrichment statistic, which places greater emphasis on genes at the top and bottom of the ranked list. For contrasts with larger numbers of differentially expressed genes (thus having important genes ranked further down the list), we used the classic enrichment statistic, which weights rank positions equally.

To visualize and aid interpretation of our GSEA results, we produced similarity networks that cluster enriched gene sets according to numbers of overlapping genes and annotated each cluster of gene sets by common functions using the *EnrichmentMap* [106] and *AutoAnnotate* [107] add-ons to Cytoscape v 3.8.0 [108]. This accounts for redundancy between gene sets, reduces the overall complexity of the GSEA results, and facilitates identification of both major functional themes as well as unique or distinct pathways from among potentially hundreds of significant yet largely redundant gene sets. Gene sets were filtered to exclude those above an FDR *q*-value cutoff of 0.01. Gene set clusters were defined using the Markov Cluster (MCL) algorithm with default settings. For each cluster, the top three most frequently occurring words among gene set names – normalized according to their frequency across all gene sets – were used to concisely annotate each cluster for generalized functions. We made minor manual edits to annotations where automatic annotation produced misleading results.

## Genotype analysis

We performed genotyping of mRNA sequences according to the Broad Institute's Best Practices Workflow for RNAseq short variant discovery. Detailed descriptions of the genotyping and variant filtration workflow are provided in Additional file 3: Text S3. To characterize population genetic structure in both devils and DFTD tumors that may be associated with among-population variation in tumor gene expression, we performed separate discriminant analyses of principal components (DAPC), implemented in the R package *adeigenet*, using the respective SNP datasets. We defined groups according to sample locality to characterize the extent of genetic differentiation among them; however, we also used the *find.clusters* function to infer genetic clusters purely from SNP variation, without a priori designation.

To identify specific SNPs and indels potentially associated with differentially expressed genes, we used SnpEff [47] to annotate all genotyped tumor variants according to their positions with respect to known genes (using the Ensembl v 7.0.86 annotation for *S. harrisii*) and quantifying each variant's predicted impact on gene function. For example, a variant leading to loss or gain of a stop codon would have a predicted high impact, a nonsynonymous variant would be predicted to have a moderate impact, and a synonymous variant would be predicted to have a low impact [47]. We evaluated these annotations against our lists of differentially expressed genes to identify variants potentially contributing to or associated with up- or downregulation of genes. We tested all variants predicted to affect any gene that was differentially expressed between any pair of localities for their effects on expression of their respective genes (measured as counts per million and normalized as above) using a Kruskal-Wallis test. *P*-values were adjusted for multiple comparisons using the Benjamini-Hochberg approach, with an FDR threshold of 0.05.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07994-4>.

**Additional file 1 Table S1.** Sample metadata.

**Additional file 2 Table S2.** Differential gene expression contrasts performed for each analysis set.

**Additional file 3 Text S3.** Genotyping and variant filtration workflow for RNAseq data.

**Additional file 4 Fig. S4.** MDS plots showing expression variation with respect to sequencing batch for each analysis. Left-hand plots show batch effects prior to correction, while right-hand plots show batch-corrected expression variation.

**Additional file 5 Fig. S5.** Heat map indicating relative expression of 14,807 genes among 19 DFTD tumor samples and 39 normal lip tissues from Tasmanian devils (19 DFTD-infected, 20 putatively uninfected). Samples are arranged as columns, with genes arranged as rows. Differential expression analyses additionally accounted for sex, locality, batch, and

tissue type – indicated for each sample as colored bars above each column. Relative gene expression is shown as a gradation from red (overexpressed) to blue (underexpressed). Dendrograms indicate clustering of samples (top) and genes (right) by similarity of expression.

**Additional file 6 Fig. S6.** Genes differentially expressed in DFTD tumors relative to normal lip tissues are enriched for various biological functions (GO-terms). Enriched biological functions are shown for genes up- and downregulated in DFTD, including functions that were enriched for both. Log2(fold enrichment) > 0 indicates overrepresented functions, while < 0 indicates underrepresented functions.

**Additional file 7 Fig. S7.** Venn diagram showing differentially expressed genes shared between each between-locality contrast among DFTD tumors.

**Additional file 8 S8 File.** .xlsx file containing tables of gene lists for all differential expression contrasts.

**Additional file 9 S9 File.** .xlsx file containing tables of significantly over- and underrepresented biological processes for up- and downregulated genes for each differential expression contrast.

**Additional file 10 S10 File.** .xlsx file containing tables of gene set enrichment results, including annotated gene set clusters.

**Additional file 11 S11 File.** .xlsx file containing tables of gene lists for all differential expression contrasts using batch 3 samples only.

**Additional file 12 S12 Table.** Table of results for GATK DepthOfCoverage analysis of sequencing coverage across different genomic regions: exons, 3'UTR, 5'UTR, < 5 kb downstream, < 5 kb upstream, introns, and intergenic regions. Total length of each region is provided, along with mean coverage across each region, as well as the percentage of each region covered at several coverage thresholds: 1x, 3x, 5x, and 10x.

**Additional file 13 S13 Table.** List of gene symbols mentioned in the manuscript text, with full gene names and short descriptions of each summarized from GeneCards.

## Acknowledgements

We thank S. Hotaling for thoughtful comments on our manuscript and J. Kelley for advice during the early stages of analysis. We are grateful to M. and C. Walsh from Discovery Holiday Parks—Cradle Mountain for providing accommodation and logistic support during fieldwork. Forico Pty Ltd. facilitated land access during our study. We thank S. Peck for veterinary support, B. and P. Hopcroft and a large number of volunteers who helped to collect data.

## Authors' contributions

C.P.K. designed the study, processed samples, performed all bioinformatic and statistical analyses, prepared figures, and wrote the manuscript with contributions from all authors. M.R.A., D.G.H., and A.K.F. collected samples. A.K.F. contributed to the study design and processed samples. A.K.F., A.H.P., and M.J.M. provided statistical input. A.S. provided infrastructure. R.H., H.M., P.A.H., M.E.J., and A.S. conceived of and oversaw the project.

## Funding

This work was funded by NIH grant no. R01-GM126563 as part of the NIH-NSF-USDA Ecology and Evolution of Infectious Diseases Program (to A.S., M.E.J., H.M., and P.A.H.), the National Science Foundation (DEB 2027446) to A.S., M.E.J., H.M., and M.J.M., the Australian Research Council (ARC) Future Fellowship (FT100100250) and ARC Large Grants (A00000162) to M.E.J., and the ARC DECRA (DE 170101116; LP 170101105) and Agence Nationale de la Recherche (ANR Blanc Project TRANSCAN) to R.H.

## Availability of data and materials

All sequence data are available at the NCBI Sequence Read Archive under BioProject PRJNA693818 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA693818>). All other sample data are provided in the Supplementary Materials.

## Declarations

### Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. All research involving use of animals and field activity was performed in compliance with the ARRIVE guidelines, under University of Tasmania ethics approval A13326 and Washington State University IACUC approval ASAF 6796.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>School of Biological Sciences, Washington State University, Pullman, Washington, USA. <sup>2</sup>Department of Integrative Biology, University of California, Berkeley, California, USA. <sup>3</sup>Department of Microbiology & Immunology, Montana State University, Bozeman, MT, USA. <sup>4</sup>School of Natural Sciences, University of Tasmania, Hobart, Tasmania, Australia. <sup>5</sup>CANECEV, Centre de Recherches Ecologiques et Evolutives sur le Cancer, 34090 Montpellier, France. <sup>6</sup>Environmental Futures Research Institute, Griffith University, Nathan, Queensland, Australia. <sup>7</sup>Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA. <sup>8</sup>Department of Integrative Biology, University of South Florida, Tampa, Florida, USA.

Received: 10 February 2021 Accepted: 8 September 2021

Published online: 27 September 2021

## References

- Melo FDSE, Vermeulen L, Fessler E, Medema JP. Cancer heterogeneity—a multifaceted view. *EMBO Rep.* 2013;14(8):686–95. <https://doi.org/10.1038/embor.2013.92>.
- Zardavas D, Irtthum A, Swanton C, Piccart M. Clinical management of breast cancer heterogeneity. *Nat Rev Clin Oncol.* 2015;12(7):381–94. <https://doi.org/10.1038/nrclinonc.2015.73>.
- Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer.* 2007;7(7):545–53. <https://doi.org/10.1038/nrc2173>.
- Cieślak M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. *Nat Rev Genet.* 2018;19(2):93–109. <https://doi.org/10.1038/nrg.2017.96>.
- Li M, Sun Q, Wang X. Transcriptional landscape of human cancers. *Oncotarget.* 2017;8:34534–51. <https://doi.org/10.18632/oncotarget.15837>.
- Espinal-Enríquez J, Fresno C, Anda-Jáuregui G, Hernández-Lemus E. RNA-Seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. *Sci Rep.* 2017;7:1–19.
- Calon A, Lonardo E, Berenguer-Llargo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet.* 2015;47(4):320–9. <https://doi.org/10.1038/ng.3225>.
- Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Grait K, et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc Natl Acad Sci U S A.* 2015;112(47):E6544–52. <https://doi.org/10.1073/pnas.1518007112>.
- Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 2013;10(5):e1001453. <https://doi.org/10.1371/journal.pmed.1001453>.
- Zhao J, Guo C, Xiong F, Yu J, Ge J, Wang H, et al. Single cell RNA-seq reveals the landscape of tumor and infiltrating immune cells in nasopharyngeal carcinoma. *Cancer Lett.* 2020;477:131–43. <https://doi.org/10.1016/j.canlet.2020.02.010>.
- Hotelling S, Shah AA, McGowan KL, Tronstad LM, Giersch JJ, Finn DS, et al. Mountain stoneflies may tolerate warming streams: Evidence from organismal physiology and gene expression. *Glob Chang Biol.* 2020. <https://doi.org/10.1111/gcb.15294>.
- Oleksiak MF. Changes in gene expression due to chronic exposure to environmental pollutants. *Aquat Toxicol.* 2008;90(3):161–71. <https://doi.org/10.1016/j.aquatox.2008.08.010>.
- Bonneaud C, Balenger SL, Russell AF, Zhang J, Hill GE, Edwards SV. Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. *Proc Natl Acad Sci.* 2011;108:7866 LP–7871. <https://doi.org/10.1073/pnas.1018580108>.
- Ellison A, Zamudio K, Lips K, Mulet-Wolz C. Temperature-mediated shifts in salamander transcriptomic responses to the amphibian-killing fungus. *Mol Ecol.* 2020;29(2):325–43. <https://doi.org/10.1111/mec.15327>.
- Ostrander EA, Davis BW, Ostrander GK. Transmissible tumors: breaking the cancer paradigm. *Trends Genet.* 2016;32(1):1–15. <https://doi.org/10.1016/j.tig.2015.10.001>.
- Metzger MJ, Goff SP. A sixth modality of infectious disease: contagious cancer from devils to clams and beyond. *PLoS Pathog.* 2016;12:1–7.
- Pyecroft SB, Pearce AM, Loh R, Swift K, Belov K, Fox N, et al. Towards a case definition for devil facial tumour disease: what is it? *Ecohealth.* 2007;4(3):346–51. <https://doi.org/10.1007/s10393-007-0126-0>.
- McCallum H, Tompkins DM, Jones M, Lachish S, Marvanek S, Lazenby B, et al. Distribution and impacts of Tasmanian devil facial tumor disease. *Ecohealth.* 2007;4(3):318–25. <https://doi.org/10.1007/s10393-007-0118-0>.
- Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, et al. Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biol Conserv.* 2006;131(2):307–24. <https://doi.org/10.1016/j.biocon.2006.04.010>.
- Lazenby BT, Tobler MW, Brown WE, Hawkins CE, Hocking GJ, Hume F, et al. Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. *J Appl Ecol.* 2018;55(3):1368–79. <https://doi.org/10.1111/1365-2664.13088>.
- Storfer A, Hohenlohe PA, Margres MJ, Patton A, Fraik AK, Lawrance M, et al. The devil is in the details: genomics of transmissible cancers in Tasmanian devils. *PLoS Pathog.* 2018;14(8):e1007098. <https://doi.org/10.1371/journal.ppat.1007098>.
- Murchison EP, Tovar C, Hsu A, Bender HS, Kheradpour P, Rebbeck CA, et al. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. *Science* (80- ). 2010;327:84–7.
- Siddle HV, Kaufman J. How the devil facial tumor disease escapes host immune responses. *Oncoimmunology.* 2013;2(8):e25235. <https://doi.org/10.4161/onci.25235>.
- Stammnitz MR, Coorens THH, Gori KC, Hayes D, Fu B, Wang J, et al. The origins and vulnerabilities of two transmissible cancers in Tasmanian devils. *Cancer Cell.* 2018;33:607–619.e15. <https://doi.org/10.1016/j.ccell.2018.03.013>.
- Brown GK, Kreiss A, Lyons AB, Woods GM. Natural killer cell mediated cytotoxic responses in the tasmanian devil. *PLoS One.* 2011;6(9):1–10. <https://doi.org/10.1371/journal.pone.0024475>.
- Patchett A, Woods G. Targeting transmissible cancers in animals. *Science* (80- ). 2019;365:438–40.
- Siddle HV, Kreiss A, Tovar C, Yuen CK, Cheng Y, Belov K, et al. Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. *Proc Natl Acad Sci U S A.* 2013;110(13):5103–8. <https://doi.org/10.1073/pnas.1219920110>.
- Hamilton DG, Jones ME, Cameron EZ, McCallum H, Storfer A, Hohenlohe PA, et al. Rate of intersexual interactions affects injury likelihood in Tasmanian devil contact networks. *Behav Ecol.* 2019;30(4):1087–95. <https://doi.org/10.1093/beheco/arz054>.
- Hamede RK, McCallum H, Jones M. Biting injuries and transmission of Tasmanian devil facial tumour disease. *J Anim Ecol.* 2013;82(1):182–90. <https://doi.org/10.1111/j.1365-2656.2012.02025.x>.
- Lachish S, Jones M, McCallum H. The impact of disease on the survival and population growth rate of the Tasmanian devil. *J Anim Ecol.* 2007;76(5):926–36. <https://doi.org/10.1111/j.1365-2656.2007.01272.x>.
- McCallum H, Jones M, Hawkins C, Hamede R, Lachish S, Sinn DL, et al. Transmission dynamics of Tasmanian devil facial tumour disease may lead to disease-induced extinction. *Ecology.* 2009;90(12):3379–92. <https://doi.org/10.1890/08-1763.1>.
- Wells K, Hamede RK, Kerlin DH, Storfer A, Hohenlohe PA, Jones ME, et al. Infection of the fittest: devil facial tumour disease has greatest effect on individuals with highest reproductive output. *Ecol Lett.* 2017;20(6):770–8. <https://doi.org/10.1111/ele.12776>.
- Wells K, Hamede RK, Jones ME, Hohenlohe PA, Storfer A, McCallum HI. Individual and temporal variation in pathogen load predicts long-term

- impacts of an emerging infectious disease. *Ecology*. 2019;100(3):e02613. <https://doi.org/10.1002/ecy.2613>.
34. Ruiz-Aravena M, Jones ME, Carver S, Estay S, Espejo C, Storfer A, et al. Sex bias in ability to cope with cancer: Tasmanian devils and facial tumour disease. *Proc R Soc B Biol Sci*. 2018;285(1891):20182239. <https://doi.org/10.1098/rspb.2018.2239>.
  35. Jones ME, Cockburn A, Hamede R, Hawkins C, Hesterman H, Lachish S, et al. Life-history change in disease-ravaged Tasmanian devil populations. *Proc Natl Acad Sci U S A*. 2008;105(29):10023–7. <https://doi.org/10.1073/pnas.0711236105>.
  36. Lachish S, McCallum H, Jones M. Demography, disease and the devil: life-history changes in a disease-affected population of Tasmanian devils (*Sarcophilus harrisii*). *J Anim Ecol*. 2009;78(2):427–36. <https://doi.org/10.1111/j.1365-2656.2008.01494.x>.
  37. Epstein B, Jones M, Hamede R, Hendricks S, McCallum H, Murchison EP, et al. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nat Commun*. 2016;7(1):12684. <https://doi.org/10.1038/ncomms12684>.
  38. Hubert JN, Zerjal T, Hospital F. Cancer- and behavior-related genes are targeted by selection in the Tasmanian devil (*Sarcophilus harrisii*). *PLoS One*. 2018;13:1–15.
  39. Margres MJ, Jones ME, Epstein B, Kerlin DH, Comte S, Fox S, et al. Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisii*) infected with a transmissible cancer. *Mol Ecol*. 2018;27(21):4189–99. <https://doi.org/10.1111/mec.14853>.
  40. Margres MJ, Ruiz-Aravena M, Hamede R, Jones ME, Lawrance MF, Hendricks SA, et al. The genomic basis of tumor regression in Tasmanian devils (*Sarcophilus harrisii*). *Genome Biol Evol*. 2018;10(11):3012–25. <https://doi.org/10.1093/gbe/evy229>.
  41. Margres MJ, Ruiz-Aravena M, Hamede R, Chawla K, Patton AH, Lawrance MF, et al. Spontaneous tumor regression in Tasmanian devils associated with RASL11A activation. *Genetics*. 2020;genetics.303428.2020. <https://doi.org/10.1534/genetics.120.303428>.
  42. Flies AS, Blackburn NB, Lyons AB, Hayball JD, Woods GM. Comparative analysis of immune checkpoint molecules and their potential role in the transmissible tasmanian devil facial tumor disease. *Front Immunol*. 2017;8 MAY:1–27.
  43. Flies AS, Bruce Lyons A, Corcoran LM, Papenfuss AT, Murphy JM, Knowles GW, et al. PD-L1 is not constitutively expressed on tasmanian devil facial tumor cells but is strongly upregulated in response to IFN- $\gamma$  and can be expressed in the tumor microenvironment. *Front Immunol*. 2016;7 DEC:1–13.
  44. Kosack L, Wingelhofer B, Popa A, Orlova A, Agerer B, Vilagos B, et al. The ERBB-STAT3 axis drives Tasmanian devil facial tumor disease. *Cancer Cell*. 2019;35:125–139.e9. <https://doi.org/10.1016/j.ccell.2018.11.018>.
  45. Kozakiewicz CP, Ricci L, Patton AH, Stahlke AR, Hendricks SA, Margres MJ, et al. Comparative landscape genetics reveals differential effects of environment on host and pathogen genetic structure in Tasmanian devils (*Sarcophilus harrisii*) and their transmissible tumour. *Mol Ecol*. 2020;29(17):3217–33. <https://doi.org/10.1111/mec.15558>.
  46. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, et al. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc*. 2019;14(3):703–21. <https://doi.org/10.1038/s41596-019-0128-8>.
  47. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. 2012;6(2):80–92. <https://doi.org/10.4161/fly.19695>.
  48. Adetunji MO, Lamont SJ, Abasht B, Schmidt CJ. Variant analysis pipeline for accurate detection of genomic variants from transcriptome sequencing data. *PLoS One*. 2019;14(9):e0216838–8. <https://doi.org/10.1371/journal.pone.0216838>.
  49. Coudray A, Battenhouse AM, Bucher P, Iyer VR. Detection and benchmarking of somatic mutations in cancer genomes using RNA-seq data. *PeerJ*. 2018;6:e5362. <https://doi.org/10.7717/peerj.5362>.
  50. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214–8. <https://doi.org/10.1038/nature12213>.
  51. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–74. <https://doi.org/10.1016/j.cell.2011.02.013>.
  52. Hanahan D, Weinberg RA. The hallmarks of Cancer. *Cell*. 2000;100(1):57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
  53. Discher DE, Mooney DJ, Zandstra PW. Growth factors, matrices, and forces combine and control stem cells. *Science* (80- ). 2009;324:1673–1677. <https://doi.org/10.1126/science.1171643>.
  54. Gritsenko PG, Ilna O, Friedl P. Interstitial guidance of cancer invasion. *J Pathol*. 2012;226(2):185–99. <https://doi.org/10.1002/path.3031>.
  55. Walker C, Mojares E, Del Río Hernández A. Role of extracellular matrix in development and cancer progression. *Int J Mol Sci*. 2018;19(10):3028. <https://doi.org/10.3390/ijms19103028>.
  56. Foster I. Cancer: a cell cycle defect. *Radiography*. 2008;14(2):144–9. <https://doi.org/10.1016/j.radi.2006.12.001>.
  57. Kirsch DG, Kastan MB. Tumor-suppressor p53: implications for tumor development and prognosis. *J Clin Oncol*. 1998;16(9):3158–68. <https://doi.org/10.1200/JCO.1998.16.9.3158>.
  58. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat*. 2007;28(6):622–9. <https://doi.org/10.1002/humu.20495>.
  59. Curtin NJ. DNA repair dysregulation from cancer driver to therapeutic target. *Nat Rev Cancer*. 2012;12(12):801–17. <https://doi.org/10.1038/nrc3399>.
  60. Chae YK, Anker JF, Carneiro BA, Chandra S, Kaplan J, Kalyan A, et al. Genomic landscape of DNA repair genes in cancer. *Oncotarget*. 2016;7:23312–21. <https://doi.org/10.18632/oncotarget.8196>.
  61. Gavande NS, VanderVere-Carozza PS, Hinshaw HD, Jalal SI, Sears CR, Pawelczak KS, et al. DNA repair targeted therapy: the past or future of cancer treatment? *Pharmacol Ther*. 2016;160:65–83. <https://doi.org/10.1016/j.pharmthera.2016.02.003>.
  62. Varga T, Aplan PD. Chromosomal aberrations induced by double strand DNA breaks. *DNA Repair (Amst)*. 2005;4:1038–46. <https://doi.org/10.1016/j.dnarep.2005.05.004>.
  63. Taylor RL, Zhang Y, Schöning JP, Deakin JE. Identification of candidate genes for devil facial tumour disease tumorigenesis. *Sci Rep*. 2017;7(1):8761. <https://doi.org/10.1038/s41598-017-08908-9>.
  64. Czerwińska P, Mazurek S, Wiznerowicz M. The complexity of TRIM28 contribution to cancer. *J Biomed Sci*. 2017;24(1):63. <https://doi.org/10.1186/s12929-017-0374-4>.
  65. Jovčevska I, Zupanec N, Urlep Ž, Vranič A, Matos B, Stokin CL, et al. Differentially expressed proteins in glioblastoma multiforme identified with a nanobody-based anti-proteome approach and confirmed by OncoFinder as possible tumor-class predictive biomarker candidates. *Oncotarget*. 2017;8(27):44141–58. <https://doi.org/10.18632/oncotarget.17390>.
  66. Wang Y, Jiang J, Li Q, Ma H, Xu Z, Gao Y. KAP1 is overexpressed in hepatocellular carcinoma and its clinical significance. *Int J Clin Oncol*. 2016;21(5):927–33. <https://doi.org/10.1007/s10147-016-0979-8>.
  67. Brennan A, Dean CH, Zhang AL, Cass DT, Minsky R, Jessen KR. Endothelins control the timing of Schwann cell generation in vitro and in vivo. *Dev Biol*. 2000;227(2):545–57. <https://doi.org/10.1006/dbio.2000.9887>.
  68. Woodhoo A, Alonso MBD, Droggiti A, Turmaine M, D'Antonio M, Parkinson DB, et al. Notch controls embryonic Schwann cell differentiation, postnatal myelination and adult plasticity. *Nat Neurosci*. 2009;12(7):839–47. <https://doi.org/10.1038/nn.2323>.
  69. Pye R, Hamede R, Siddle HV, Caldwell A, Knowles GW, Swift K, et al. Demonstration of immune responses against devil facial tumour disease in wild Tasmanian devils. *Biol Lett*. 2016;12(10):20160553. <https://doi.org/10.1098/rsbl.2016.0553>.
  70. Aran D, Camarda R, Odegard J, Paik H, Oskotsky B, Krings G, et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat Commun*. 2017;8(1):1077. <https://doi.org/10.1038/s41467-017-01027-z>.
  71. Gutkin DW, Shurin MR. Clinical evaluation of systemic and local immune responses in cancer: time for integration. *Cancer Immunol Immunother*. 2014;63(1):45–57. <https://doi.org/10.1007/s00262-013-1480-0>.
  72. Fraik AK, Margres MJ, Epstein B, Barbosa S, Jones M, Hendricks S, et al. Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. *Evolution (N Y)*. 2020. <https://doi.org/10.1111/evo.14023>.
  73. Fraik AK, Quackenbush C, Margres MJ, Comte S, Hamilton DG, Kozakiewicz CP, et al. Transcriptomics of Tasmanian devil (*Sarcophilus harrisii*) ear tissue reveals homogeneous gene expression patterns across a heterogeneous landscape. *Genes (Basel)*. 2019;10(10):1–15. <https://doi.org/10.3390/genes10100801>.

74. Zhang Z, Chang Y, Zhang J, Lu Y, Zheng L, Hu Y, et al. HMGB3 promotes growth and migration in colorectal cancer by regulating WNT/ $\beta$ -catenin pathway. *PLoS One*. 2017;12(7):e0179741. <https://doi.org/10.1371/journal.pone.0179741>.
75. Nemeth MJ, Kirby MR, Bodine DM. Hmgb3 regulates the balance between hematopoietic stem cell self-renewal and differentiation. *Proc Natl Acad Sci*. 2006;103(37):13783–8. <https://doi.org/10.1073/pnas.0604006103>.
76. Sun D, Cao H, Yang L, Lin L, Hou B, Zheng W, et al. MiR-200b in heme oxygenase-1-modified bone marrow mesenchymal stem cell-derived exosomes alleviates inflammatory injury of intestinal epithelial cells by targeting high mobility group box 3. *Cell Death Dis*. 2020;11(6):480. <https://doi.org/10.1038/s41419-020-2685-8>.
77. Yanai H, Ban T, Wang Z, Choi MK, Kawamura T, Negishi H, et al. HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature*. 2009;462(7269):99–103. <https://doi.org/10.1038/nature08512>.
78. Ewald SE, Barton GM. Nucleic acid sensing toll-like receptors in autoimmunity. *Curr Opin Immunol*. 2011;23(1):3–9. <https://doi.org/10.1016/j.coi.2010.11.006>.
79. Scaffidi P, Misteli T, Bianchi ME. Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature*. 2002;418(6894):191–5. <https://doi.org/10.1038/nature00858>.
80. Neupane M, Clark AP, Landini S, Birkbak NJ, Eklund AC, Lim E, et al. MECP2 is a frequently amplified oncogene with a novel epigenetic mechanism that mimics the role of activated RAS in malignancy. *Cancer Discov*. 2016;6(1):45–58. <https://doi.org/10.1158/2159-8290.CD-15-0341>.
81. Zajchowski DA, Bartholdi MF, Gong Y, Webster L, Liu H-L, Munishkin A, et al. Identification of Gene Expression Profiles That Predict the Aggressive Behavior of Breast Cancer Cells. *Cancer Res*. 2001;61:5168 LP – 5178. <http://cancerres.aacrjournals.org/content/61/13/5168.abstract>.
82. Zhang D, Park D, Zhong Y, Lu Y, Rycak J, Gong S, et al. Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. *Nat Commun*. 2016;7(1):1–15. <https://doi.org/10.1038/ncomms10798>.
83. Ruzyczki PA, Tran NM, Kefalov VJ, Kolesnikov AV, Chen S. Graded gene expression changes determine phenotype severity in mouse models of CRX-associated retinopathies. *Genome Biol*. 2015;16(1):171. <https://doi.org/10.1186/s13059-015-0732-z>.
84. Ho GYF, Jung HJ, Schoen RE, Wang T, Lin J, Williams Z, et al. Differential expression of circulating microRNAs according to severity of colorectal neoplasia. *Transl Res*. 2015;166(3):225–32. <https://doi.org/10.1016/j.trsl.2015.02.004>.
85. Hamede RK, Lachish S, Belov K, Woods G, Kreiss A, Pearse AM, et al. Reduced effect of Tasmanian devil facial tumor disease at the disease front. *Conserv Biol*. 2012;26(1):124–34. <https://doi.org/10.1111/j.1523-1739.2011.01747.x>.
86. Hamede RK, Pearse AM, Swift K, Barmuta LA, Murchison EP, Jones ME. Transmissible cancer in Tasmanian devils: localized lineage replacement and host population response. *Proc R Soc B Biol Sci*. 2015;282(1814):20151468. <https://doi.org/10.1098/rspb.2015.1468>.
87. Patton AH, Lawrence M, Margres MJ, Kozakiewicz CP, Hamede R, Ruiz-Aravena M, et al. A transmissible cancer shifts from emergence to endemism in Tasmanian devils. *Science* (80- ). 2020;370:eabb9772.
88. Kwon YM, Gori K, Park N, Potts N, Swift K, Wang J, et al. Evolution and lineage dynamics of a transmissible cancer in Tasmanian devils 2020. <https://doi.org/10.1371/journal.pbio.3000926>, 18, 11, e3000926.
89. Becker DJ, Albery GF, Kessler MK, Lunn TJ, Falvo CA, Cziráj GÁ, et al. Macroimmunology: the drivers and consequences of spatial patterns in wildlife immune defence. *J Anim Ecol*. 2020;89(4):972–95. <https://doi.org/10.1111/1365-2656.13166>.
90. Williams CR, Baccarella A, Parrish JZ, Kim CC. Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics*. 2016;17(1):1–13. <https://doi.org/10.1186/s12859-016-0956-2>.
91. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11:1650–1667. doi:<https://doi.org/10.1038/nprot.2016-095>, 9.
92. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
93. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. 2019;20(1):1–13. <https://doi.org/10.1186/s13059-019-1910-1>.
94. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. *Nucleic Acids Res*. 2019;48:D682–8. <https://doi.org/10.1093/nar/gkz966>.
95. Pertea M, Pertea G. GFF Utilities: GffRead and GffCompare. *F1000Research*. 2020;9:1–17.
96. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. 2015;4.
97. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
98. Lun ATL, Chen Y, Smyth GK. It's DE-licious: a recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. In: *Statistical Genomics*. Springer; 2016. p. 391–416.
99. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–7. <https://doi.org/10.1093/nar/gkv007>.
100. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
101. Hong G, Zhang W, Li H, Shen X, Guo Z. Separate enrichment analysis of pathways for up- and downregulated genes. *J R Soc Interface*. 2014;11(92):20130950. <https://doi.org/10.1098/rsif.2013.0950>.
102. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6(7):e21800. <https://doi.org/10.1371/journal.pone.0021800>.
103. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267–73. <https://doi.org/10.1038/ng1180>.
104. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102>.
105. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2018;46(D1):D649–55. <https://doi.org/10.1093/nar/gkx1132>.
106. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010;5(11):e13984. <https://doi.org/10.1371/journal.pone.0013984>.
107. Kucera M, Isserlin R, Arkhangorodsky A, Bader GD. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research*. 2016;5:1717. doi:<https://doi.org/10.12688/f1000research.9090.1>.
108. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models. *Genome Res*. 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303.metabolite>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

