



OPEN

Identification of tumor-related genes via RNA sequencing of tumor tissues in *Xenopus tropicalis*

Kazuki Kitamura¹, Takayoshi Yamamoto¹, Haruki Ochi², Makoto Suzuki³, Nanoka Suzuki³, Takeshi Igawa³, Tadashi Yoshida¹, Mitsuru Futakuchi⁴, Hajime Ogino³ & Tatsuo Michiue^{1,5}✉

Cancer treatment is still challenging because the disease is often caused by multiple mutations. Although genomic studies have identified many oncogenes and tumor suppressor genes, gene sets involved in tumorigenesis remain poorly understood. *Xenopus*, a genus of aquatic frogs, is a useful model to identify gene sets because it can be genetically and experimentally analyzed. Here, we analyzed gene expression in tumor tissues of three individuals in *Xenopus tropicalis* and identified 55 differentially expressed genes (DEGs). Gene ontology (GO) analysis showed that the upregulated genes in the tumor tissues were enriched in GO terms related to the extracellular matrix and collagen fibril organization. Hierarchical clustering showed that the gene expression patterns of tumor tissues in *X. tropicalis* were comparable to those of human connective, soft, and subcutaneous tissue-derived cancers. Additionally, pathway analysis revealed that these DEGs were associated with multiple pathways, including the extracellular matrix, collagen fibril organization, MET signaling, and keratan sulfate. We also found that the expression tendency of some DEGs that have not been well analyzed in the cancer field clearly determines the prognosis of human cancer patients. This study provides a remarkable reference for future experimental work on *X. tropicalis* to identify gene sets involved in human cancer.

Gene mutations occur at a certain frequency. For instance, they are caused by environmental cues, including smoking, alcohol consumption, and dietary habits¹⁻³. The accumulation of genetic mutations causes various diseases, among which cancer is highly common⁴. It is estimated that approximately half of all humans develop cancer during their lifetime, and the early detection of cancer is particularly crucial for reducing the risk of death. Therefore, genetic testing has gained increasing attention in cancer treatment.

Recently, many disease-associated genes have been identified using next-generation sequencing (NGS). For example, 736 genes have been defined as “cancer-driving genes” by the Cancer Gene Census (CGC) (version 97, November 2022), which is a key resource within the Catalogue of Somatic Mutations in Cancer (COSMIC), comprising a long-term, ongoing effort to catalog and describe all genes with causal impact in human cancer⁵. For instance, in breast cancer, BRCA1/2 genes are used for genetic testing, and for patients with mutations in these genes, treatments such as mastectomy are performed even prior to the onset of cancer progression⁶. However, in many cases, a single mutation is insufficient to induce cancer, i.e., the accumulation of gene mutations, rather than a single gene mutation, causes cancer^{7,8}. Nevertheless, the gene sets critical for cancer development and progression are not yet well determined. Therefore, determining these gene sets enables diagnosis and treatment before the onset of cancer progression, thereby reducing the risk of cancer-related deaths.

Many genes exhibit differences in expression levels between normal and cancer tissues. In addition, certain genes demonstrate significant changes in expression levels in cancer tissues with a poor prognosis^{9,10}. It is possible that subsequent treatment could be changed or modified depending on the gene expression pattern in the cancer. Thus, in cancer research, not only mutation analysis but also expression analysis is important. However, although gene expression levels have been just analyzed for tumor tissues, studies linked the expression analysis to cancer treatment are limited.

¹Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-Ku, Tokyo 153-8902, Japan. ²Institute for Promotion of Medical Science Research, Faculty of Medicine, Yamagata University, 2-2-2 Iida-Nishi, Yamagata City, Yamagata 990-9585, Japan. ³Amphibian Research Center, Hiroshima University, 1-3-2 Kagamiyama, Higashi-Hiroshima City, Hiroshima 739-8511, Japan. ⁴Department of Pathology, Faculty of Medicine, Yamagata University, 2-2-2 Iida-Nishi, Yamagata City, Yamagata 990-9585, Japan. ⁵Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan. ✉email: tmichiue@bio.c.u-tokyo.ac.jp

Genetic and experimental analyses are required to determine gene sets associated with cancer. In this regard, genome-editing technologies, such as the CRISPR/Cas9 system, are useful¹¹. Most of these screenings have been performed using cultured cells. However, gene expressions *in vitro* differ from those *in vivo*¹². Thus, there is still a gap between *in vivo* and *in vitro* screening, and *in vivo* screening is crucial for biomedical research. Because *in vivo* assays in humans are difficult due to ethical dilemmas, useful model organisms are required.

Xenopus tropicalis (*X. tropicalis*), a genus of aquatic frogs, has a diploid genome, rendering it an ideal organism for genome-editing experiments¹³. Recently, many tumors have been found across several generations of the Nigerian H (NH) strain of *X. tropicalis*. In this study, we performed RNA-seq analysis of three *X. tropicalis* individuals of the NH strain with tumors to identify tumorigenic genes and compared them with human cancer tissues to provide novel responsible genes for oncogenesis.

Results

Observation of *X. tropicalis* tumor tissues. Tumors developed spontaneously over several generations in the NH strain (Fig. S1), one of which (NH-VIII-7) has already been histologically analyzed¹⁴. In this study, we analyzed the gene expression profiles of three tumor-bearing individuals, including NH-VIII-7. Similar to NH-VIII-7, tumors were found in the flanks of NH-III-4 and NH-V-1 (Fig. 1a,b). The tumors were hard and marginally different in color from the surrounding skin tissue, i.e., purple, gray, or reddish brown, and had a dome shape (Fig. 1a',b').

For further detailed observation, the tissues were sectioned and stained with Hematoxylin and Eosin (HE). In normal tissues, the granules and glands are localized just inside the epidermis. The dermis, which is derived from



Figure 1. External view and histological analysis of *X. tropicalis* tumors. **(a,b)** External view of NH-III-4 and NH-V-1 tumor individuals. Arrowheads indicate the tumors. **(a',b')** Enlarged view of the tumors. Scale bar indicates 2 mm **(a',b')**. **(c)** Histological analysis of normal tissue, stained by HE. Skin is composed of the epidermis (E) and dermis, which is subdivided into the dense regular connective tissue (C.T.) (arrow), dense irregular C.T. (arrow) and loose C.T. (L). Arrowheads indicate the pigment. The pigment adjacent to the epidermis is commonly referred to as the chromatophore. GG, granular gland; MG, mucous gland. **(d,e)** Histological analysis of tumors of NH-III-4 **(d)** and NH-V-1 **(e)**. Arrowheads in **(d,e)** indicate the pigment. Scale bar indicates 100 μ m **(c-e)**.

the mesoderm, is layered under the granules and glands (Fig. 1c). In the tumor tissue, the mucous glands and serous glands were missing in NH-III-4 (Fig. 1d). Instead of this, connective tissue (C.T.) of the dermis seemed to have increased, particularly in dense regular C.T. of the dermis. Similar to this, in NH-V-1 (Fig. 1e), the dermis showed remarkable thickening. In addition, pigments were scattered in the tumor tissues. These observations suggest that these individuals are considered to have defects mainly in the mesoderm-derived tissues.

Differences in gene expression patterns among samples using hierarchical clustering. We performed RNA-seq analysis to investigate the gene expression patterns of tumor tissues (Tumor Tissue (flank) #1-#3, defined as TT-body1-3 hereafter). We also collected skin tissues, which have no tumors, from the legs of three tumor-bearing individuals (non-tumor tissue (leg) in Tumor-bearing individual #1-#3, TI-leg1-3) and the flank and leg of three no-tumor-bearing individuals (non-tumor tissue (flank) in No-tumor-bearing individual #1-#3, NI-body1-3; non-tumor tissue (leg) in No-tumor-bearing individual #1-#3, NI-leg1-3), as negative control samples (Table 1).

To identify differences in gene expression patterns among samples, we performed hierarchical clustering using Pearson's correlation coefficients¹⁵ (Fig. 2a, Fig. S2a). The three tumor tissues were positioned close to each other in the dendrogram, suggesting that their gene expression patterns were similar. Because the results of both hierarchical clustering and histological analysis showed that the three samples were similar, TT-body1-3 were analyzed as biological replicates in subsequent analyses.

In addition, because the flank and leg skin tissues of no-tumor-bearing individuals were also closely positioned in the dendrogram (Fig. 2a), the expression patterns are similar independent to the region, flank or leg, of the skin. Thus, we used the leg skin of tumor-bearing individuals as a control sample. However, the leg skins possibly show abnormal gene expression patterns. Therefore, both the leg skin of tumor-bearing individuals and the flank skin of no-tumor-bearing individuals were used as control samples for subsequent analyses (Fig. 2b).

Analysis of differential gene expression between the flank skin (tumor) and the leg skin (non-tumor) in tumor-bearing individuals.

To identify differentially expressed genes (DEGs), we compared gene expression between TT-body1-3 and TI-leg1-3 (Fig. 2c) with Trimmed Means of M values (TMM) method¹⁶. We identified 268 DEGs (DEG1; Supplementary Table 1) (FDR < 0.05) and visualized the gene expression differences using MA plot (Fig. 2d). DEG1 possibly includes DEGs due to region difference, flank or leg. To only detect differences in gene expression due to tumorigenesis, we decided to exclude the DEGs of NI-leg1-3 and NI-body1-3 from DEG1 (Fig. 2e). 9 genes were output as DEGs (DEG2; Fig. 2f, Supplementary Table 2). We thought of removing the common genes of both DEG1 and DEG2, which were considered region-independent genes, that is, tumor-specific genes (DEG3) (Fig. 2g) but, the number of common genes was zero. So, in this case, DEG1 was identical to DEG3 (Fig. 1g, Supplementary Table 1).

Gene Ontology (GO) enrichment analysis was performed using DEG3, referred to the human GO annotation file. Human orthologous genes were determined by the Reciprocal best BLAST Hit (RBH) using OrthoFinder^{17,18}. The number of genes that DEG3 finally converted to human genes was 138 (113 upregulated genes and 25 downregulated genes in the tumor tissue; DEG4; Supplementary Table 3). These genes were then subjected to GO analysis. Upregulated DEG4 genes were enriched in GO terms related to extracellular matrix (ECM), collagen fibril organization and osteogenesis (Fig. 3a). The cancer-driving genes⁵ COL1A1, COL3A1, FAT4, and FKBP9, were commonly included in the DEGs with enriched GO terms. Although the number of downregulated DEG4 genes was small, GO analysis was performed. GO term of alpha-amino acid catabolic process and cellular amino acid catabolic process were enriched in the down regulated genes (Fig. 3b).

In DEG3 (268 genes in total), there were lots of gene (130 genes) that were not converted to human genes by RBH. Of these, 79 genes remain unannotated in *Xenopus* genome database. We then examined their orthologs just by best blast hits (the gene that came up first by the BLAST search), not by "reciprocal" best blast hit (RBH). Out of 130 genes, 73 genes were identified. The remaining 57 genes were putative homologs of the 73 genes or did not have any hits with the BLAST search. GO analysis was performed with the 73 genes in addition to the 138 genes (DEG4). The results in upregulated genes showed that the top enriched GO terms were almost consistent with the analysis conducted using the 138 genes alone (Fig. S2b,c).

Analysis of differential gene expression of tumor tissue (flank skin) and the equivalent regions in no-tumor-bearing individuals.

As mentioned above, because the negative control leg skin was derived from tumor-bearing individuals, gene expression may differ from that in no-tumor-bearing individuals, even though the leg skin did not have tumors. Therefore, we compared gene expression with NI-body1-3 as a negative control (Fig. 4a). 198 genes were identified as DEGs (DEG5; Fig. 4b, Supplementary Table 4). Because these genes might include gene expression differences that caused by innate mutation throughout the body of tumor-bearing individuals, we excluded DEGs of TI-leg1-3 and NI-leg1-3 (DEG6; Supplementary Table 5) from DEG5 (Fig. 4c) to identify tumor-specific genes. 14 genes were identified as DEG6 (Fig. 4d). 1 gene was common to

Region	Tumor-bearing individual	No-tumor-bearing individual
Flank	Tumor tissue: TT-body	Non tumor tissue: NI-body
Leg	Non tumor tissue: TI-leg	Non tumor tissue: NI-leg

Table 1. Sample names of *Xenopus* tumor tissues.

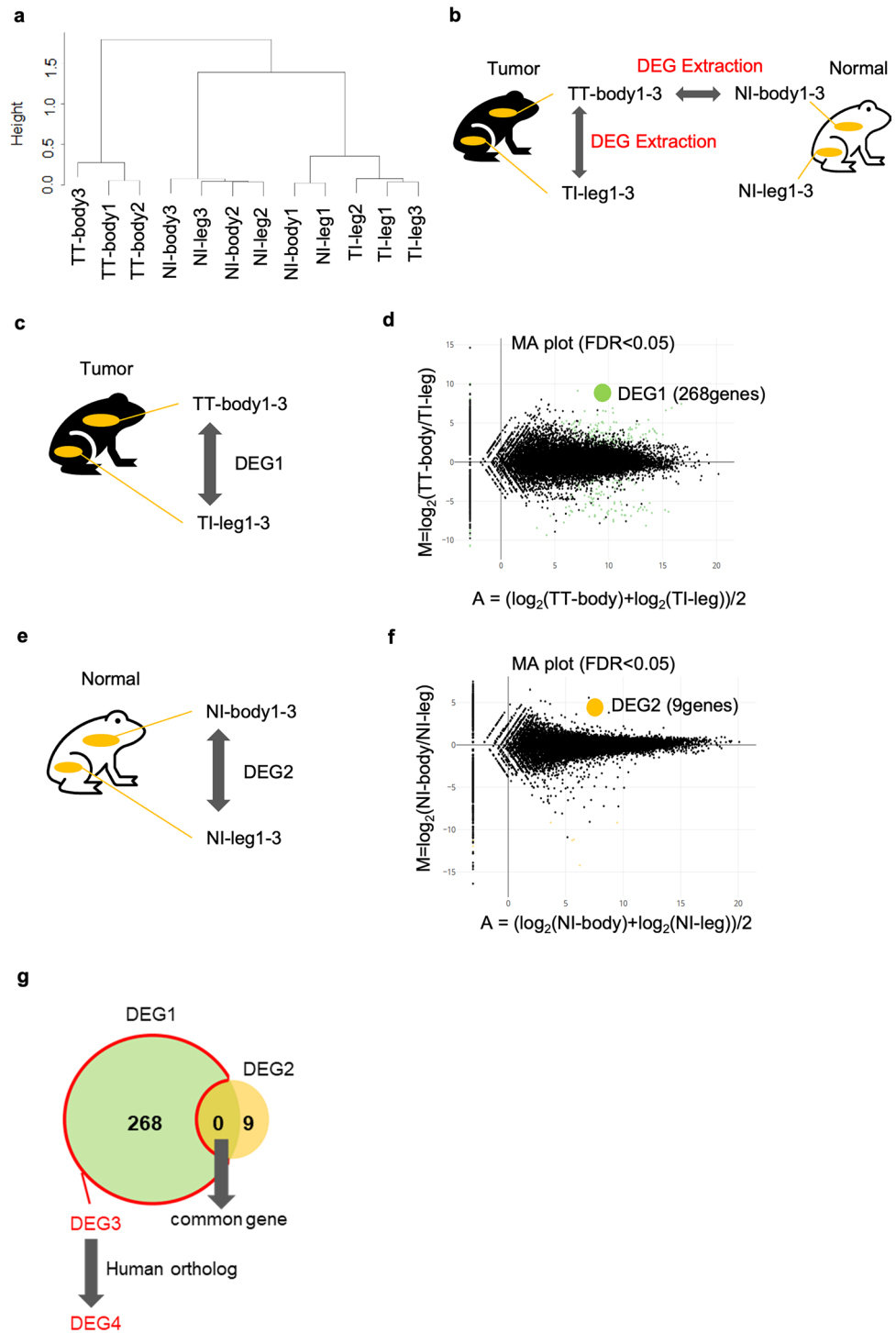


Figure 2. Hierarchical clustering of all analyzed tumors in *Xenopus* and DEGs of TT-body1-3 vs TI-leg1-3. (a) Hierarchical clustering of 12 analyzed samples. Difference of the expression pattern among the samples was indicated with a tree diagram. (b) Schematic diagram of extraction of differentially expressed genes (DEGs) in *Xenopus*, which we analyzed in this study. (c,d) Schematic diagram of DEGs extraction (TT-body1-3 vs TI-leg1-3) (c) and the MA plot (d). Green dots indicate DEGs with statistical significance (FDR<0.05) in (d). (e,f) Schematic diagram of DEGs extraction (NI-body1-3 vs NI-leg1-3) (e), and the MA plot (f). Orange dots indicate DEGs with statistical significance (FDR<0.05) in (f). (g) Venn diagram showing DEG3. DEG3 is $\text{DEG1} \cap \text{DEG2}$.

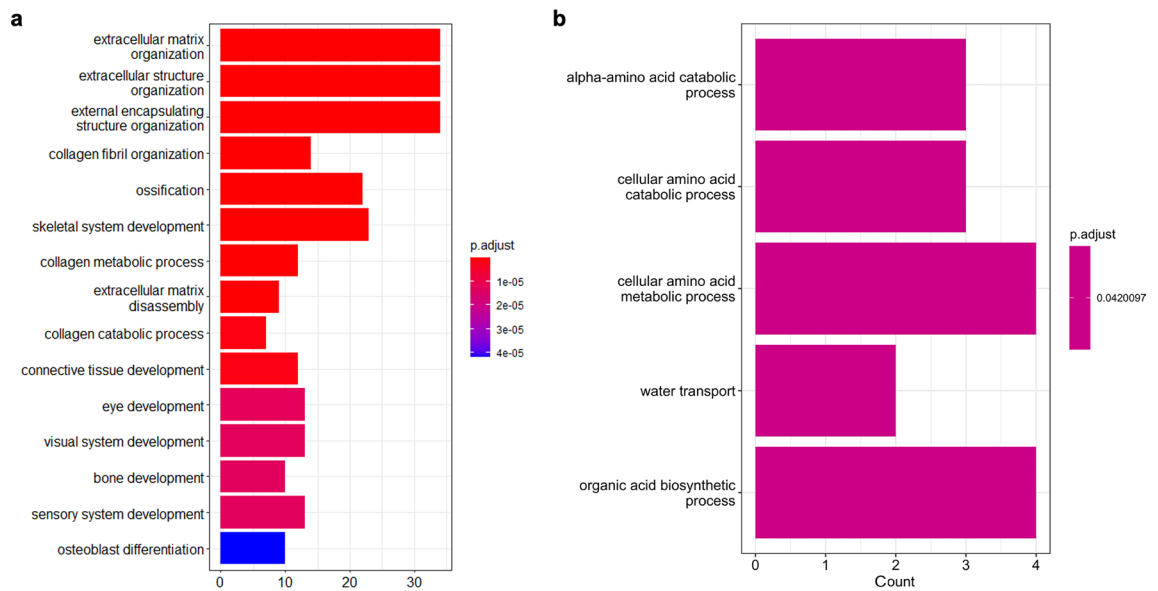


Figure 3. GO analysis of up- or down-regulated genes of DEG4. GO analysis of upregulated genes (**a**) or downregulated genes (**b**). Bar plots represents number of genes involved in each term. Adjusted p-value was represented by color scale, and the statistically significance level decreased from red (higher significance) to blue (lower significance).

both DEG5 and DEG6 (Supplementary Table 6). Thus, by excluding the 1 gene from DEG5, 197 genes were defined as DEG7 (Supplementary Table 7), which are tumor-specific genes (Fig. 4e).

We determined human orthologs of DEG7 by RBH, and 96 genes were identified (DEG8; Supplementary Table 8). Among the DEG8, 58 and 38 genes were upregulated and downregulated in tumor tissues, respectively. Upregulated DEG8 genes were enriched in GO terms related to the ECM, collagen metabolism, and bone formation, and these results were similar to those for DEG4 (Fig. 5a). In addition, the cancer-driving genes⁵ COL3A1 and FAT4 were commonly included in DEG8 with enriched GO terms, similar to those of DEG4. Downregulated DEG8 genes were enriched in GO terms related to myogenic processes (Fig. 5b).

Non-tumor tissues (leg skin) of tumor-bearing individuals are possibly in a pre-tumor state. To examine the possibility, we extracted common genes between DEG5 and DEG6 and found only one gene, BAAT. To further examine it, we extracted common genes with gaining FDR threshold (FDR < 0.1). 22 genes were extracted: of them, 7 genes were reported to have roles in promoting or suppressing cancer in humans. For instance, AMPD3 is known to be highly expressed in gastrointestinal stromal tumors¹⁹; BGLAP is expressed in pancreatic cancer cells and promotes their growth and invasion²⁰; COL6A6 inhibits the growth of non-small cell lung cancer²¹; PODN inhibits the growth of osteosarcoma²²; and SOD3 is downregulated in breast cancer²³.

To further narrow down responsible genes of the tumor in *Xenopus*, we investigated the genes commonly included in DEG3 and DEG7, which used different negative controls (Fig. 5c). Of these 119 genes (DEG9; Supplementary Table 9), 62 genes were upregulated, and 57 genes were downregulated in the tumor tissue. 55 genes were identified as human orthologs by RBH (DEG10; Supplementary Table 10), of which 46 genes were upregulated and 9 genes were downregulated. GO terms related to the ECM, collagen fibril organization and ossification were enriched in the upregulated genes (Fig. 5d). In the downregulated genes, GO term could not be properly extracted because of the small number of genes.

Comparison of DEGs between *Xenopus* tumor tissues and human mesoderm-derived cancer tissues.

We compared these expression patterns with those of human mesoderm-derived cancers because the tumor in *Xenopus* shows dermal expansion. Using human cancer expression data registered in The Cancer Genome Atlas (TCGA), we performed hierarchical clustering with all the expression data of TT-body1-3. TT-body1-3 seems to exhibit a similar pattern to human cancer tissues derived from bone, joint, and articular cartilage of the extremities (hereafter referred to as "cancer tissues derived from bone tissue"), connective tissue, subcutaneous tissue, and other soft tissues (Fig. S3a).

To compare the DEGs in *Xenopus* tumor tissues with those in human mesoderm-derived cancer, we identified DEGs in human cancer. Read-count data of bone marrow cells and muscle tissues were obtained as controls from the atlas of RNA-seq profiles for normal human tissues²⁴. Consequently, 11,555 genes (7484 upregulated and 4071 downregulated genes) were identified as DEGs in the cancer tissues derived from bone tissue (DEG11, Fig. 6a, Supplementary Table 11), and 7432 genes (3787 upregulated and 3645 downregulated genes) were identified as DEGs in cancer tissues derived from connective, subcutaneous, and other soft tissues (DEG12, Fig. 6b, Supplementary Table 12). Similar to the *Xenopus* tumor, upregulated genes of DEG11 were enriched in GO terms related to ECM and skeletal system (Fig. 7a). Downregulated genes were enriched in GO terms related to T-cell activation and cell-cell adhesion (Fig. 7b). Upregulated genes of DEG12 were enriched in GO terms related to protein targeting to membrane (Fig. 7c). Downregulated genes were enriched in GO terms related to

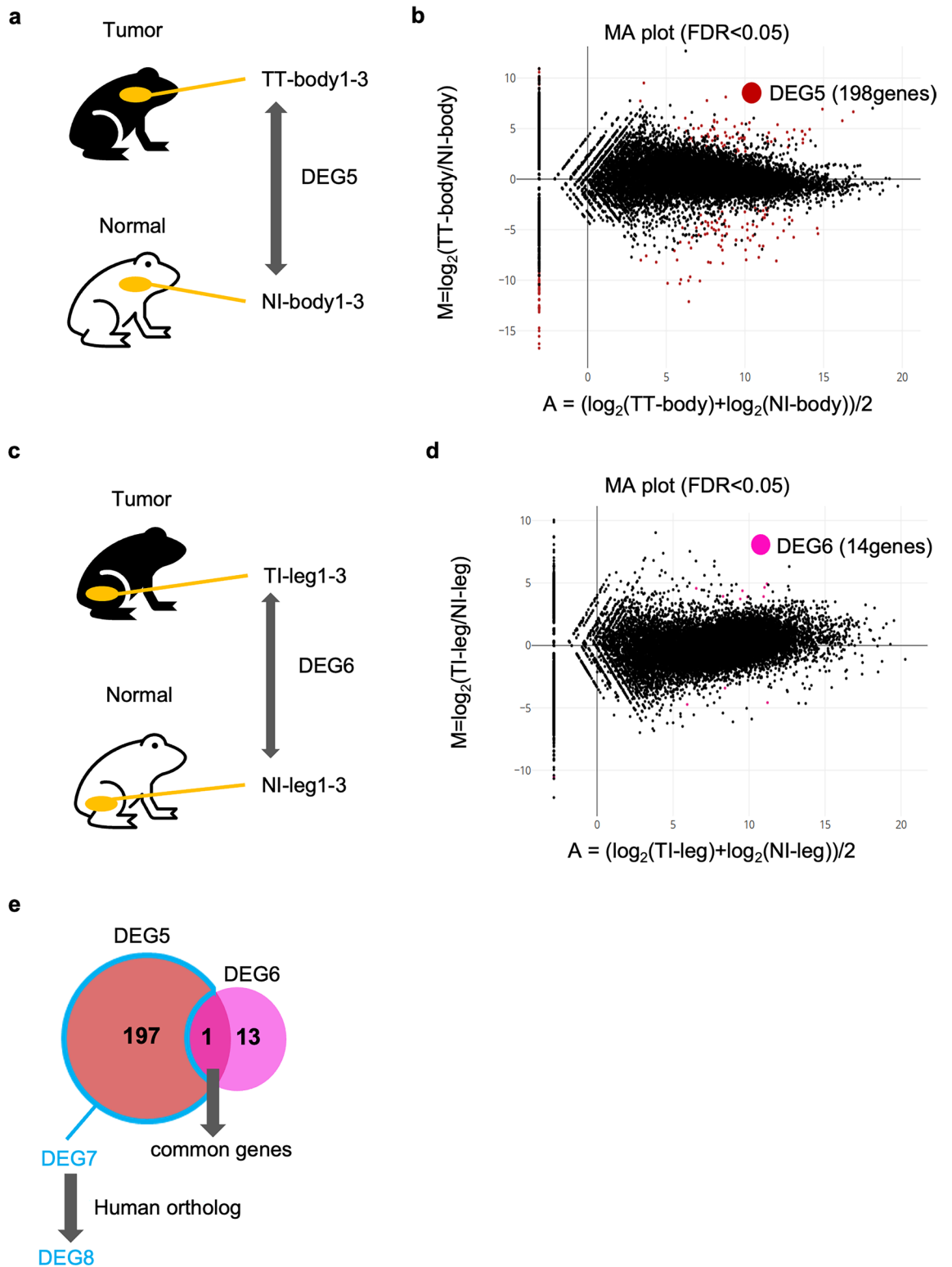


Figure 4. DEGs of TT-body1-3 vs NI-body1-3. (a,b) Schematic diagram of DEGs extraction (TT-body1-3 vs NI-body1-3) (a) and the MA plot (b). (c,d) Schematic diagram of DEGs extraction (TI-leg1-3 vs NI-leg1-3) (c) and the MA plot (d). Red (b) and pink (d) dots indicate DEGs with statistical significance (FDR<0.05). (e) Venn diagram showing DEG7. DEG7 is DEG5—DEG5∩DEG6.

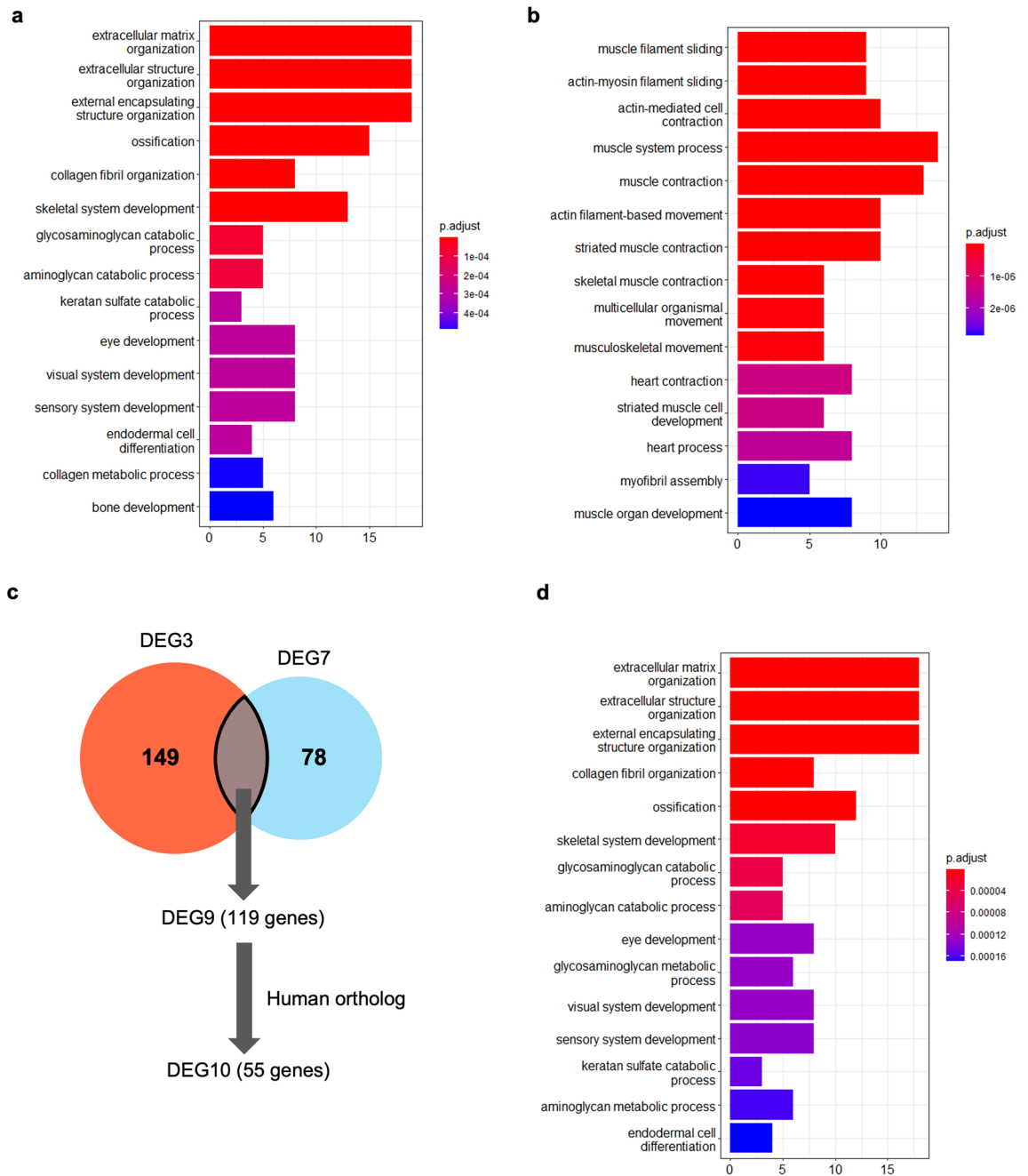


Figure 5. GO analysis of DEG8 and DEG10 genes. GO analysis of upregulated genes (a) or downregulated genes (b) of DEG8. Bar plots represents number of genes involved in each term. Adjusted p-value was represented by color scale, and the statistically significance level decreased from red (higher significance) to blue (lower significance). (c) Venn diagram showing DEG9, which is common genes of DEG3 and DEG7. (d) GO analysis of upregulated genes of DEG10.

muscle tissue, such as the development of muscle structures and muscular system processes, as in the *Xenopus* tumor tissue (Figs. 5b, 7d).

We then identified common genes between *Xenopus* DEGs (DEG10) and human DEGs (DEG11) (Fig. 6c), which was 45 genes (DEG13; Supplementary Table 13). Of these genes, 41 were upregulated and 4 were downregulated in *Xenopus* tumor tissues. The cancer-driving genes⁵, COL3A1 and FAT4, were included in upregulated genes. We examined whether the expression of 45 genes was higher or lower in *Xenopus* tumor and human cancer than in each control. Most genes (44) showed similar expression patterns, upregulated or downregulated, in *Xenopus* and human tumor tissues (Fig. S3b,c). The remaining one gene, SOD3 was downregulated in *Xenopus* tumor whereas it was upregulated in human cancer.

We then extracted common DEGs from DEG10 and DEG12 and identified 24 common genes (DEG14; Supplementary Table 14; Fig. 6d), which included 22 upregulated and 2 downregulated genes in *Xenopus* tumor

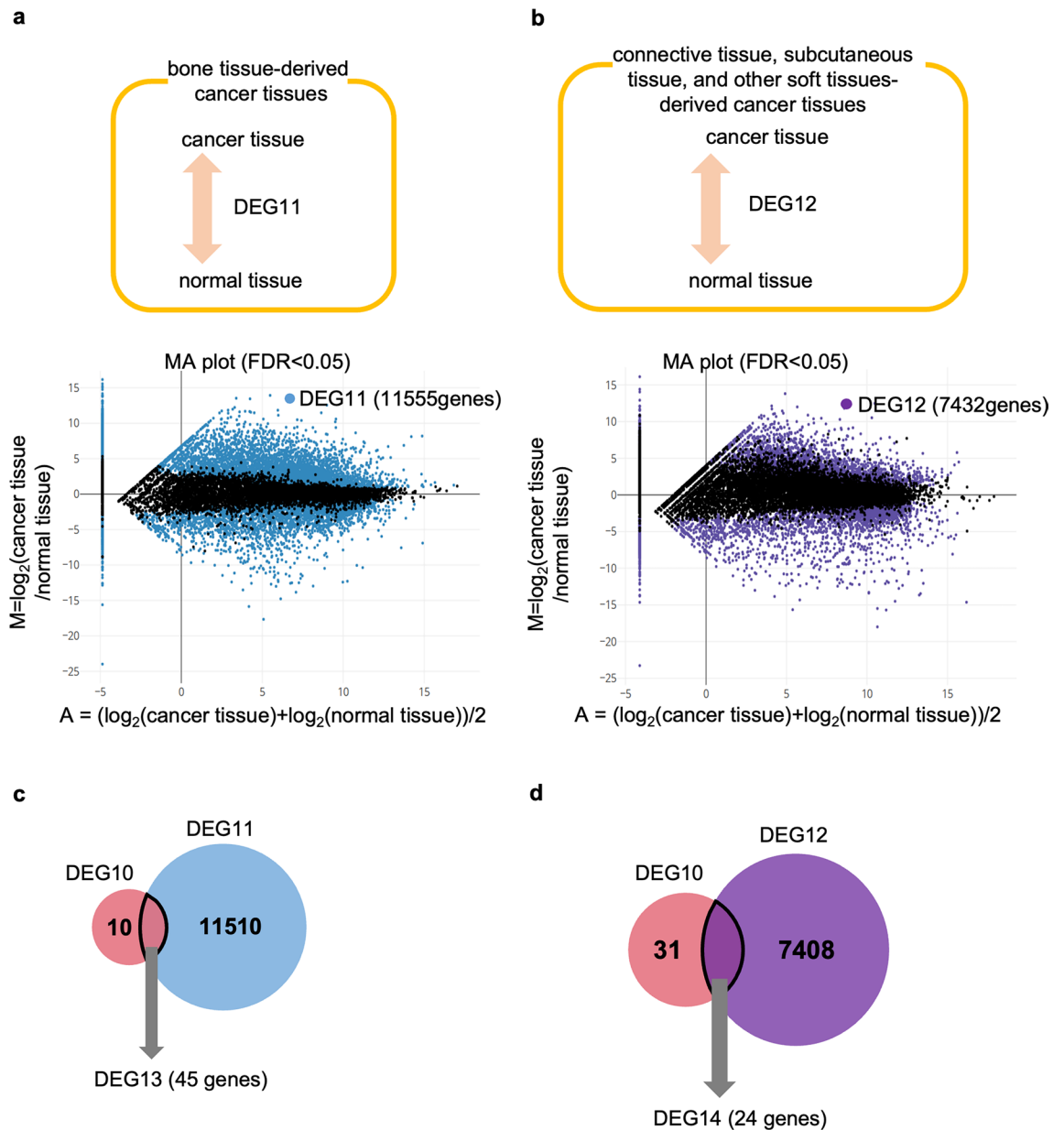


Figure 6. MA plot comparing gene expression pattern of cancer tissues with that of normal tissues in human. (a,b) MA plot of cancer tissue vs normal tissue derived from bone tissue (a) and connective tissues, subcutaneous tissue, and other soft tissues (b) in human. Blue (a) and purple (b) dots indicate DEGs with statistical significance (FDR < 0.05). (c,d) Venn diagram of the extraction of DEG13 (c) and DEG14 (d).

tissue. Similar to DEG13, DEG14 included the cancer-driving genes⁵ COL3A1 and FAT4. In addition, 19 of the 22 genes showed similar gene expression patterns, upregulated or downregulated, in *Xenopus* tumor and human cancer derived from connective, subcutaneous, and other soft tissues (Fig. S3d,e). About the remaining genes, FAT4 was upregulated in *Xenopus* tumor and downregulated in human cancer. Two genes, TFF3 and FCGBP, were downregulated in *Xenopus* tumor and upregulated in human cancer.

Identification of DEGs in each *Xenopus* tumor tissue. Although the three samples derived from the NH strain in *Xenopus* were considered biological replicates in the above analyses, each tumor tissue may be in a different tumor progression state. At least, based on the external view, TT-body3 (NH-V-1) may be different from the other tumors: TT-body3 exhibited a spherical tumor (Fig. 1b'), while TT-body1 (NH-VIII-7) and TT-body2 (NH-III-4) exhibited distorted shapes with multiple overlapping protrusions (Fig. 1a', Fig. S4a).

To investigate this, we identified DEGs of each TT-body with the same negative control samples (leg skin of tumor-bearing individuals [TI-leg1-3] or flank skin of no-tumor-bearing individuals [NI-body1-3]; Fig. S4b-g, Supplementary Tables 15–18). The number of DEGs in TT-body3 was higher than in TT-body #1 (TT-body1) or TT-body2, especially the DEGs in tumor tissue vs. leg skin (Fig. S4b,d,f). To further determine the differences among these DEGs, we extracted the common genes of each DEGs identified in the leg skin as controls

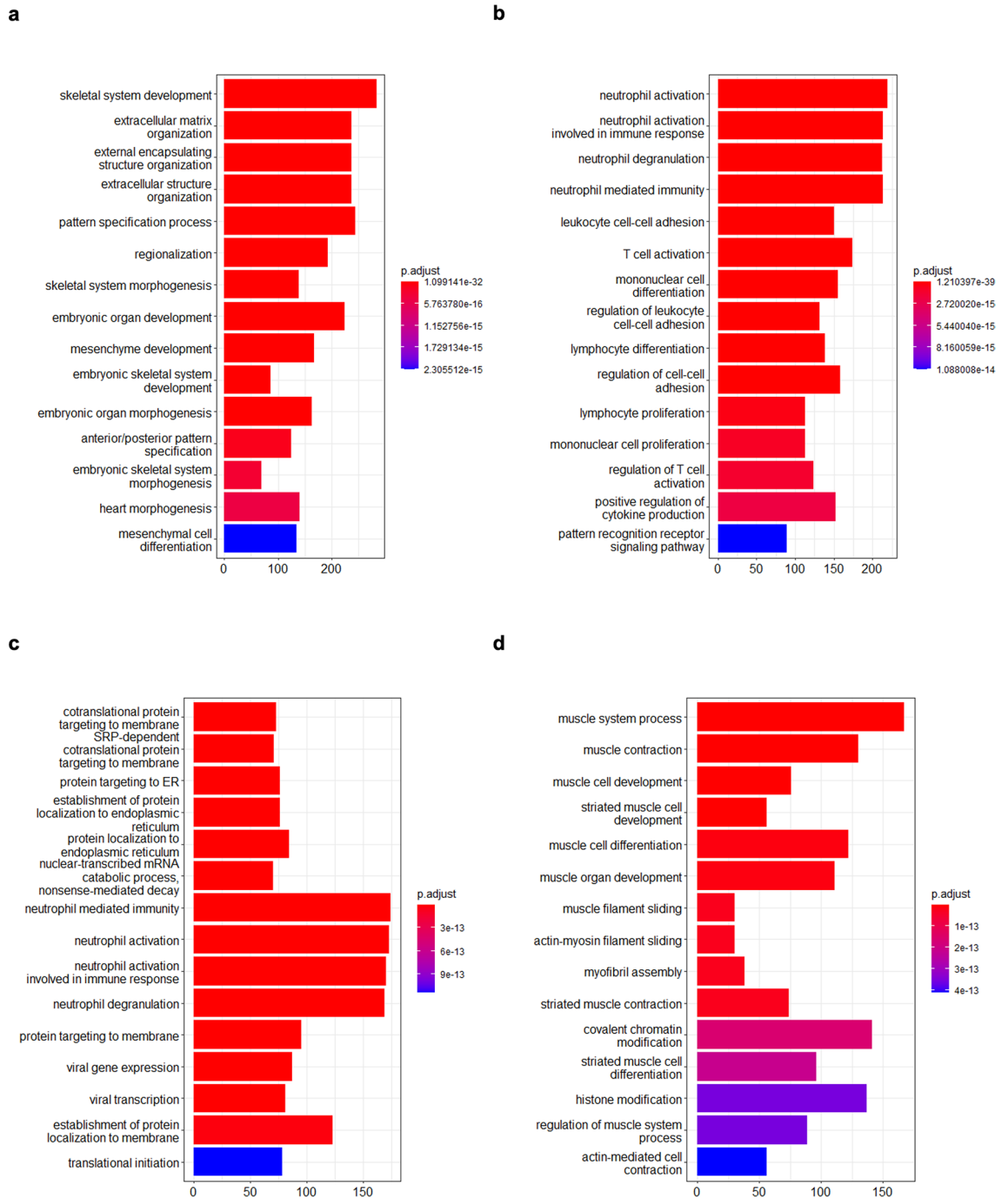


Figure 7. GO analysis of DEG11 and DEG12 genes. (a,b) GO analysis of upregulated genes (a) or downregulated genes (b) of DEG11. (c,d) GO analysis of upregulated genes (a) or downregulated genes (b) of DEG12. Bar plots represents number of genes involved in each term. Adjusted p-value was represented by color scale, and the statistically significance level decreased from red (higher significance) to blue (lower significance).

(Fig. S5a,b). The percentage of common DEGs between TT-body1 and TT-body2 was larger than 50% (Fig. S5a, bars 1–2). The percentage of common DEGs from TT-body1 and 3 was 65% of that from TT-body1 (Fig. S5a, bar 3), and 62% of the DEGs from TT-body2 were common to DEGs from TT-body2 and TT-body3 (Fig. S5a, bars 5). These results suggest that DEGs from TT-body1 and TT-body2 share many common genes and that DEGs of TT-body3 vs. TI-leg1-3 also share genes with TT-body1 and TT-body2.

The number of DEGs in TT-body3 was higher than that in TT-body1 and TT-body2, although there were many common genes among the DEGs in TT-body1 to TT-body3, suggesting that TT-body3 is in a higher state of tumor progression than TT-body1 and TT-body2. To examine TT-body3 specific genes, 3903 genes not included in DEGs of TT-body1 and TT-body2 were extracted and converted to human genes (2575 genes). In the 2575 genes, it was found that 123 genes were cancer-driving genes⁵.

Pathway analysis of DEG in *Xenopus*. As described above, we found that the tumor-specific genes in *Xenopus*, DEG10 (common genes between DEG4 (tumor tissue vs. leg skin) and DEG8 (tumor tissue vs. flank skin)), shared many genes with the cancer-specific genes in humans (DEG11 and 12). To further investigate DEG10, we performed pathway analysis of DEG10 using ReactomePA²⁵, a database of human biological reactions and pathways. Genes involved in ECM and collagen fibers were enriched (Fig. 8). PDGF and MET signaling factors related to angiogenesis in tumor tissues were also enriched. As genes participating in multiple pathways are probably effective in tumorigenesis, we determined the number of genes participating in multiple pathways (>2 pathways) in DEG10 (Fig. S6a,b). Of the 55 genes (DEG10), 19 genes, most of which were upregulated in *Xenopus* tumor tissue, were involved in multiple pathways (Fig. S6a). For instance, COL1A2 and COL3A1 function in many pathways such as ECM organization, collagen fibers, MET signaling, and PDGF signaling. LUM is involved in multiple functions such as collagen fiber organization, epithelial cell migration, and tissue repair and exhibits dual functionality as both an oncogene and a tumor suppressor gene^{26,27}.

To confirm the involvement of collagen in the tumors, we performed Elastica van Gieson staining with *Xenopus* tumors that have similar characteristics (Fig. S7a–c). This was clearly stained, indicating the presence of collagen fiber hyperplasia in the tumor (Fig. S7d). The tumor tissue in the dermis was also clearly stained by Azan staining (Fig. S7e), indicating that the collagen fibers in the tumor include type I collagen. The dermis was also stained by silver, indicating that this collagen fibers also include type III collagen (Fig. S7f). These characteristics are similar to dermatofibromas in human, which sometimes metastasizes²⁸.

Survival curves of human patients with defects in the expression of DEG10 genes. Among all the genes of DEG10 (55 genes), 9 genes have not been well analyzed in the cancer field. We analyzed the survival curves of human patients with defects in the expression levels of these genes using Survival Genie²⁹ (Fig. S8, Supplementary Table 19), which uses human expression data from many cancer tissues obtained from The Cancer Genome Atlas (TCGA). For example, in patients with kidney cancer, higher expression of LACTBL1 was correlated with a decrease in survival rates (Fig. S8d). Consistent with this, LACTBL1 was highly expressed in *Xenopus* tumors. In patients with colon, connective, subcutaneous, and other soft tissue cancers, higher expression of KERA, GUSB, and LACTBL1 had decreased survival rates (Fig. S8i), which were consistently highly expressed in

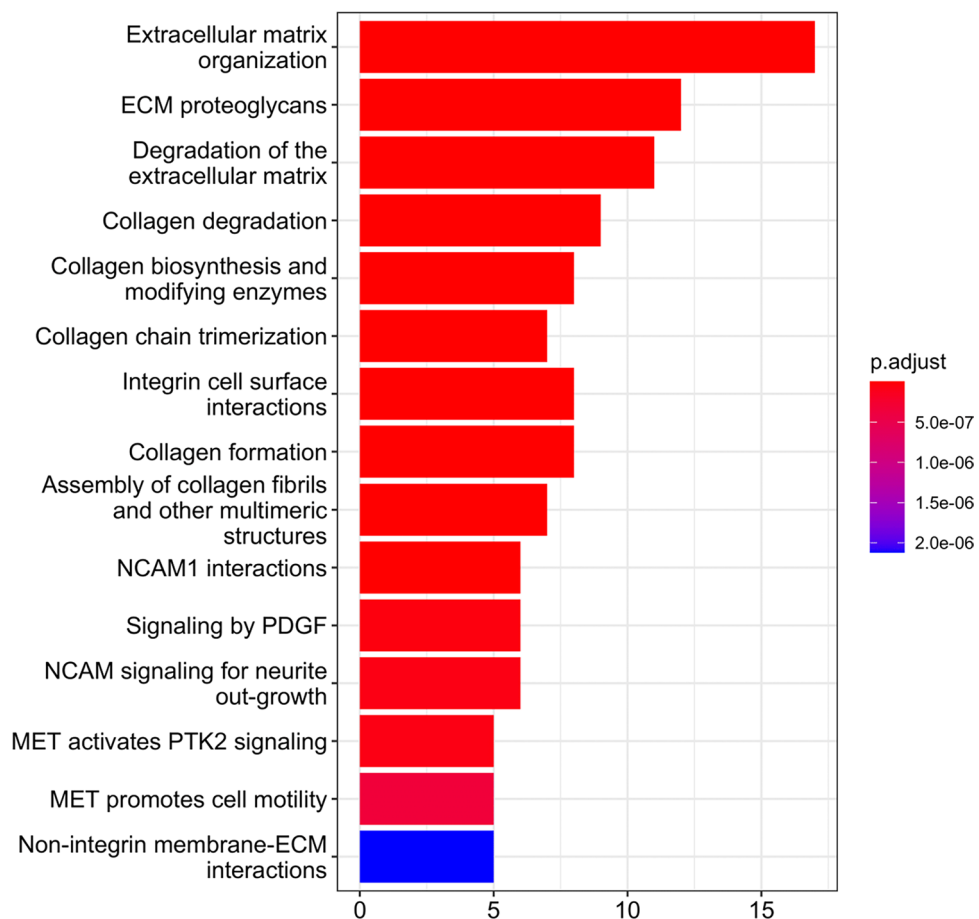


Figure 8. Reactome Pathway Analysis of genes of DEG10. Bar plots represents number of DEGs involved in each term. Adjusted p-value was represented by color scale, and the statistically significance level decreased from red (higher significance) to blue (lower significance).

Xenopus tumors. In patients with cancer of the bones, joints, the articular cartilage of limbs, lower expression of KLHDC7A, whose expression was also lower in *Xenopus* tumors, had decreased survival rates (Fig. S8h). Taken together, 6 out of the 9 genes were proposed to influence the survival rate in human cancer.

Discussion

In this study, we performed histological and RNA-seq analyses on tumor tissues in *Xenopus*. We have identified putative responsible genes for tumorigenesis, some of which correlate with prognosis in human cancer patients.

In the historical analyses, dermal region seemed to have increased in the tumor tissues (Fig. 1d,e). Considering the dermis consists of various cell types, it is important to determine which part was increased. In the GO analyses of upregulated genes in tumor tissues, the terms of ossification and collagen fibril organization were enriched. The layer below the granules in the dermis, the Eberth–Katschenko layer, is known for its high levels of calcium³⁰ and collagen^{31,32}, and the surrounding region of the layer is also suggested to be high in the expression of osteogenic genes³⁰. Taken together, it is plausible that these regions have expanded in the tumor tissues. To elucidate the underlying mechanism of this expansion, two potential reasons can be considered: promoted proliferation or enhanced/changed differentiation. Taking the absence of proliferation-related terms in the GO analysis into consideration, the latter has higher possibility.

Although most cancer research focuses on gene mutations, alterations in gene expression, must have a substantial impact on tumorigenesis. By focusing on gene expression of tumors, we have successfully identified novel putative responsible genes (DEG10) involved in cancer formation. In addition, most genes (46 genes) in DEG10 (55 genes) related to *Xenopus* tumorigenesis are involved in human oncogenesis, suggesting that there are high similarities between *Xenopus* and humans in the mechanism of tumorigenesis, and *Xenopus* could be a model organism for human cancer research. In addition, we found that the expression levels of 6 genes in DEG10, which have not been well analyzed in relation to human cancer, impacted the subsequent survival rate of cancer patients in human (Fig. S8). The reason why these genes have not been identified as cancer-driving genes⁵ to date may be partly attributed to their infrequent occurrence of mutations in human cancers or mutation-based analyses primarily focus on the coding sequences, not on the expression levels. Therefore, a drug capable of regulating the expression level of these genes could potentially control cancer progression. This could be a novel option for drug selection, which based on the gene expression in the tumor. Thus, this study provides a new perspective in the field of human cancer treatment.

Materials and methods

***Xenopus tropicalis* (*X. tropicalis*).** All animal experiments were approved by the Office for Life Science Research Ethics and Safety, at the University of Tokyo (#2020-6). Tumor and normal individuals of *X. tropicalis* NH (Nigerian H) strain, a fully inbred strain³³, were provided by the Amphibian Research Center (Hiroshima University, Japan). Detailed information on the analyzed individuals is shown in Supplementary Table 20. In this study, we analyzed only females of *Xenopus*.

Imaging, histology. Tissues were fixed at room temperature overnight using Bouin's Fixative. The fixed tissues were then embedded in paraffin, and 8 μ m sections were generated by microtomy. Standard methods were used to stain the sections with hematoxylin and eosin (HE). Collagen was stained with Elastica van Gieson (EvG) staining. Type I collagen fibers were stained blue with Azan staining. Type III collagen was stained black with silver staining.

RNA purification. Tumor tissues and skin of the leg and flank were surgically collected and subsequently crushed using a bead crusher (MS-100R, TOMY) with 6 mm diameter beads. The beads and samples were placed in a tube together and crushed five times (3000 rpm, 1 min, 2 °C). RNA was then purified using the RNeasy RNA Cell Miniprep System (Promega).

RNA sequencing. Sequencing was performed with Illumina NovaSeq6000. Sequences were aligned by HISAT2³⁴. Samtools³⁵ was used to convert SAM files to indexed, sorted, and merged BAM files. Transcripts were assembled and quantified using StringTie³⁶.

Hierarchical clustering. To compare gene expression patterns between samples, hierarchical clustering with Pearson's correlation coefficient¹⁵ was performed using R software. R function 'hclust' was used for sample clustering based on gene expression matrices. The distance matrix is based on $1 - r$, where r is the Pearson's correlation coefficient between sample pairs. Ward's minimum variance method was used as the agglomeration method. The Ward method is used to calculate the distance matrix between clusters, and the length of each branch of the dendrogram is represented by "Height".

DEG output and MA plot. The read counts were normalized using the trimmed mean of M values (TMM) method¹⁶ with the edgeR Bioconductor package. This method scales read counts by the weighted log fold-change values of a reference sample with genes that have extreme log-fold-changes (M values) and extreme absolute expression levels (A values) removed from data. Differentially expressed genes (DEGs) were identified using a false discovery rate (FDR < 0.05). The TCC^{37,38} package was used to calculate M and A values for the MA plot. TCC is an R package that provides functions for differential expression analysis of tag count data. We used the TCC-GUI³⁹, a graphical user interface for TCC, to analyze the data.

OrthoFinder. Orthofinder is an algorithm for inferring orthologs across multiple species^{17,18}. FASTA files of protein sequence in *Xenopus* and human were downloaded from the NCBI Reference Sequence Database (RefSeq), and the sequences were used as the query for the followings. Reciprocal best BLAST Hits (RBH) method was applied to identify human orthologs for each gene in *Xenopus*.

RNA-seq data from human cancer patients. RNA-seq read count information for 43 different tumor types, which are all the categories of cancer available in the database, were collected from Genomic Data Commons (GDC) Data Portal of the Cancer Genome Atlas Program (TCGA) (10 data were obtained from each cancer data collection). We used TCGA's "Open Access data," which does not require user certification. These read count data were normalized by the TMM method¹⁶.

Gene Ontology analysis and pathway analysis. To perform Gene Ontology (GO) analysis, we referred to GO-term annotation for human genome. We examined human orthologs of the *Xenopus* gene and Gene Ontology enrichment analysis was performed using clusterprofiler⁴⁰, a package implemented in R. Reactome pathway analysis was performed using ReactomePA, an R package based on the human REACTOME pathway database²⁵.

Survival rate analysis. We utilized Survival Genie²⁹, a web-based tool, to perform Kaplan–Meier analysis, which is a statistical method to evaluate survival rates, on single-cell RNA-seq (scRNA-seq) data and a variety of other molecular inputs such as gene sets, genes ratio, tumor infiltrating immune cells proportion, gene expression profile scores, and tumor mutation burden, based on cancer patient data derived from 33 projects of TCGA's Open Access data. We defined high/low expression groups using "median" among selected samples. We also defined as a significant difference when it has > 0.2 difference in survival probability between high and low expression of a certain gene.

Data availability

The raw RNA-seq datasets of *Xenopus* tissue (fastq format) have been deposited to NCBI-GEO with the accession number GSE233287 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE233287>).

Received: 27 March 2023; Accepted: 6 August 2023

Published online: 14 August 2023

References

- Palladino-Davis, A. G., Mendez, B. M., Fisichella, P. M. & Davis, C. S. Dietary habits and esophageal cancer. *Dis. Esophagus* **28**, 59–67 (2015).
- Gandini, S. *et al.* Tobacco smoking and cancer: A meta-analysis. *Int. J. Cancer* **122**, 155–164 (2008).
- Rumgay, H., Murphy, N., Ferrari, P. & Soerjomataram, I. Alcohol and cancer: Epidemiology and biological mechanisms. *Nutrients* **13**, 3173 (2021).
- Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18545–18550 (2010).
- Sondka, Z. *et al.* The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
- Ford, D. *et al.* Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.* **62**, 676–689 (1998).
- Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
- Zhou, Q. *et al.* Genes that predict poor prognosis in breast cancer via bioinformatical analysis. *Biomed. Res. Int.* **2021**, 1–8 (2021).
- Zhang, J., Liu, D., Li, L., Wang, Q. & Lan, Y. Detection of critical genes associated with poor prognosis in breast cancer via integrated bioinformatics analyses. *JBUON* **25**, 2537–2545 (2020).
- Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Han, K. *et al.* CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities. *Nature* **580**, 136–141 (2020).
- Nakayama, T. *et al.* Cas9-based genome editing in *Xenopus tropicalis*. *Methods Enzymol.* **546**, 355–375 (2014).
- Suzuki, M., Igawa, T., Suzuki, N., Ogino, H. & Ochi, H. Spontaneous neoplasia in the western clawed frog *Xenopus tropicalis*. *MicroPubl Biol* **2020**, (2020).
- Kim, T. *et al.* Impact of similarity metrics on single-cell RNA-seq data clustering. *Brief Bioinform.* **20**, 2316–2326 (2019).
- Maza, E. In Papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-Seq experimental design. *Front. Genet.* <https://doi.org/10.3389/fgene.2016.00164> (2016).
- Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* <https://doi.org/10.1186/s13059-019-1832-y> (2019).
- Emms, D. M. & Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* <https://doi.org/10.1186/s13059-015-0721-2> (2015).
- Wong, M. *et al.* AMPD3 is associated with the malignant characteristics of gastrointestinal stromal tumors. *Oncol. Lett.* **13**, 1281–1287 (2017).
- Kayed, H. *et al.* BGLAP is expressed in pancreatic cancer cells and increases their growth and invasion. *Mol. Cancer* <https://doi.org/10.1186/1476-4598-6-83> (2007).
- Qiao, H., Feng, Y. & Tang, H. COL6A6 inhibits the proliferation and metastasis of non-small cell lung cancer through the JAK signalling pathway. *Transl. Cancer Res.* **10**, 4514–4522 (2021).
- Han, X. *et al.* Hypermethylated PODN represses the progression of osteosarcoma by inactivating the TGF- β /Smad2/3 pathway. *Pathol. Res. Pract.* **238**, 154075 (2022).
- Zhang, Y. *et al.* The effect of extracellular superoxide dismutase (SOD3) gene in lung cancer. *Front. Oncol.* <https://doi.org/10.3389/fonc.2022.722646> (2022).
- Suntsova, M. *et al.* Atlas of RNA sequencing profiles for normal human tissues. *Sci. Data* <https://doi.org/10.1038/s41597-019-0043-4> (2019).

25. Yu, G. & He, Q. Y. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
26. Chen, X. *et al.* LUM expression and its prognostic significance in gastric cancer. *Front. Oncol.* <https://doi.org/10.3389/fonc.2020.00605> (2020).
27. Nikitovic, D., Katonis, P., Tsatsakis, A., Karamanos, N. K. & Tzanakakis, G. N. Lumican, a small leucine-rich proteoglycan. *IUBMB Life* **60**, 818–823 (2008).
28. Mentzel, T. *et al.* Malignant dermatofibroma: Clinicopathological, immunohistochemical, and molecular analysis of seven cases. *Mod. Pathol.* **26**, 256–267 (2013).
29. Dwivedi, B., Mumme, H., Satpathy, S., Bhasin, S. S. & Bhasin, M. Survival Genie, a web platform for survival analysis across pediatric and adult cancers. *Sci. Rep.* <https://doi.org/10.1038/s41598-022-06841-0> (2022).
30. Okada, M., Tazawa, I., Nakajima, K. & Yaoita, Y. Expression of the amelogenin gene in the skin of *Xenopus tropicalis*. *Zoolog. Sci.* **30**, 154–159 (2013).
31. Greven, H., Zanger, K. & Schwinger, G. Mechanical properties of the skin of *Xenopus laevis* (Anura, Amphibia). *J. Morphol.* **224**, 15–22 (1995).
32. Taylor, R. E., Taylor, H. C. & Barker, S. B. Chemical and morphological studies on inorganic phosphate deposits in *Rana catesbeiana* skin. *J. Exp. Zool.* **161**, 271–285 (1966).
33. Igawa, T. *et al.* Inbreeding ratio and genetic relationships among strains of the western clawed frog, *Xenopus tropicalis*. *PLoS One* **10**, e0133963 (2015).
34. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
35. Li, H. *et al.* The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
36. Perlea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
37. Sun, J., Nishiyama, T., Shimizu, K. & Kadota, K. TCC: An R package for comparing tag count data with robust normalization strategies. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-14-219> (2013).
38. Tang, M., Sun, J., Shimizu, K. & Kadota, K. Evaluation of methods for differential expression analysis on multi-group RNA-seq count data. *BMC Bioinform.* <https://doi.org/10.1186/s12859-015-0794-7> (2015).
39. Su, W., Sun, J., Shimizu, K. & Kadota, K. TCC-GUI: A Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res. Notes* <https://doi.org/10.1186/s13104-019-4179-2> (2019).
40. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

Acknowledgements

We thank Mr. Sho Nakamura (Hosei University, Japan) and Dr. Yui Uchida (RIKEN, Japan) for technical assistance of RNA-seq analysis, Dr. Masayuki Tsuzuki (University of Kochi, Japan) and Ms. Yuko Fukuda (University of Tokyo, Japan) for help of RNA extraction, Dr. Takashi Ariizumi (Tamagawa University, Japan), Dr. Ichiro Tazawa (Hiroshima University, Japan), Mr. Toki Kaneshima (University of Tokyo, Japan), Mr. Masaki Ogawa (University of Tokyo, Japan) and Dr. Ayumi Horikawa (University of Tokyo, Japan) for critical comments on this project, Dr. Hiroki Danno (RIKEN, Japan) for help in starting this work. This work was supported in part by MEXT/JSPS KAKENHI (17KT0049 to HOc and TM, 19K16138 to TYa, 18K06244/21K06183 to TYa and TM) and JST SPRING (JPMJSP2108 to KK). We also thank the National BioResource Project (NBRP) for providing *Xenopus* adults and the NBRP Information Center (National Institute for Genetics) for providing the *Xenopus* genomic database (<http://viewer.shigen.info/xenopus/>).

Author contributions

K.K., T.Ya., H.Oc. and T.M. conceived this project. K.K., H.Oc., T.Yo. and M.F. performed histological analysis; K.K., T.Ya. and H.Oc. performed RNA-extraction and RNA-seq analysis; K.K., M.S., N.S., T.I. and H.Og. conceived strain management and extracted tumor tissues. K.K., T.Ya., H.Oc. and T.M. wrote the manuscript with help from the co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-40193-7>.

Correspondence and requests for materials should be addressed to T.M.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 Author(s) 2023