HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2023 March 29.

Published in final edited form as:

Cell Rep. 2023 January 31; 42(1): 112013. doi:10.1016/j.celrep.2023.112013.

VGLL2-NCOA2 leverages developmental programs for pediatric sarcomagenesis

Sarah Watson[1,2,17], **Collette A. LaVigne**[3,17], **Lin Xu**[4], **Didier Surdez**[1,5], **Joanna Cyrta**[6], **Delia Calderon**[7,8], **Matthew V. Cannon**[7], **Matthew R. Kent**[7], **Katherine M. Silvius**[7], **Jack P. Kucinski**[7,8], **Emma N. Harrison**[7], **Whitney Murchison**[9], **Dinesh Rakheja**[9,10], **Franck Tirode**[11], **Olivier Delattre**[1,12,13], **James F. Amatruda**[14,15,18,*], **Genevieve C. Kendall**[7,16,18,19,*]

[1]Institut Curie Research Center, Paris Sciences et Lettres (PSL) Research University, INSERM U830, 75005 Paris, France

[2]Institut Curie, Paris Sciences et Lettres (PSL) Research University, Medical Oncology Department, 75005 Paris, France

[3]Department of Molecular Biology, UT Southwestern Medical Center, Dallas, TX 75390, USA

[4]Quantitative Biomedical Research Center, Department of Population and Data Sciences, UT Southwestern Medical Center, Dallas, TX 75390, USA

[5]Balgrist University Hospital, Faculty of Medicine, University of Zürich (UZH), 8008 Zürich, Switzerland

[6]Institut Curie, Paris Sciences et Lettres (PSL) Research University, Department of Pathology, 75005 Paris, France

[7]Center for Childhood Cancer, The Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, OH 43205, USA

[8]Molecular, Cellular, and Developmental Biology Ph.D. Program, The Ohio State University, Columbus, OH 43210, USA

[9]Department of Pediatrics, UT Southwestern Medical Center, Dallas, TX 75390, USA

*Correspondence: jamatruda@chla.usc.edu (J.F.A.), genevieve.kendall@nationwidechildrens.org (G.C.K.).

AUTHOR CONTRIBUTIONS

S.W., F.T., O.D., J.F.A., and G.C.K. conceived and supervised the study. F.T. provided computational analysis, insight, and expertise of human and zebrafish RNA-seq data and their similarity in Figures 2, 6, and S3. L.X. provided computational analysis, insight, and expertise into developmental studies and human and zebrafish RNA-seq data in Figures 3, 7, and S4. M.V.C. provided computational analysis, insight, and expertise in Figures 2, 4, S3, and S6. D.R. and J.C. provided pathological assessment. S.W., C.A.L., D.S., J.C., D.C., M.K., K.S., J.K., E.H., W.M., and G.C.K. performed experiments. C.A.L., J.C., D.C., M.K., K.S., J.K., E.H., and G.C.K. analyzed experimental data. C.A.L., J.F.A., and G.C.K. drafted the manuscript and figures. All authors reviewed and edited the final manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112013.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper received support from a program designed to increase minority representation in their field of research. One or more of the authors of this paper self-identifies as living with a disability. We support inclusive, diverse, and equitable conduct of research.

[10]Department of Pathology, UT Southwestern Medical Center, Dallas, TX 75390, USA

[11]University Lyon, Université Claude Bernard Lyon 1, Cancer Research Center of Lyon, INSERM 1052, CNRS 5286, Centre LéonBérard, 69008 Lyon, France

[12]Institut Curie, SIREDO Pediatric Center, 75005 Paris, France

[13]Institut Curie Hospital Group, Unité de Génétique Somatique, 75005 Paris, France

[14]Cancer and Blood Disease Institute, Children's Hospital Los Angeles, Los Angeles, CA 90027, USA

[15]Departments of Pediatrics and Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

[16]Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH 43205, USA

[17]These authors contributed equally

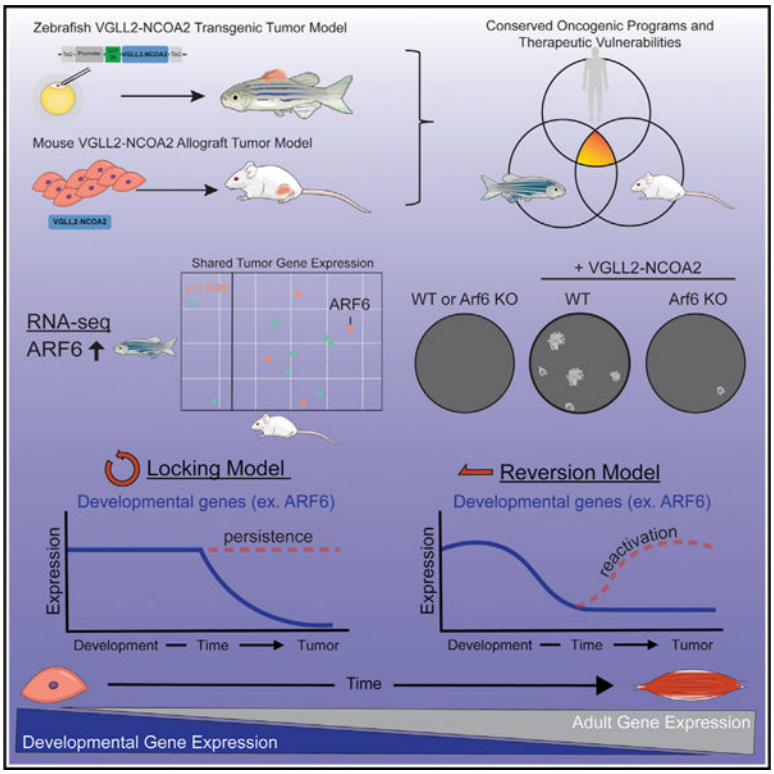
[18]These authors contributed equally

[19]Lead contact

SUMMARY

Clinical sequencing efforts are rapidly identifying sarcoma gene fusions that lack functional validation. An example is the fusion of transcriptional coactivators, VGLL2-NCOA2, found in infantile rhabdomyosarcoma. To delineate VGLL2-NCOA2 tumorigenic mechanisms and identify therapeutic vulnerabilities, we implement a cross-species comparative oncology approach with zebrafish, mouse allograft, and patient samples. We find that VGLL2-NCOA2 is sufficient to generate mesenchymal tumors that display features of immature skeletal muscle and recapitulate the human disease. A subset of VGLL2-NCOA2 zebrafish tumors transcriptionally cluster with embryonic somitogenesis and identify VGLL2-NCOA2 developmental programs, including a RAS family GTPase, ARF6. In VGLL2-NCOA2 zebrafish, mouse, and patient tumors, ARF6 is highly expressed. ARF6 knockout suppresses VGLL2-NCOA2 oncogenic activity in cell culture, and, more broadly, ARF6 is overexpressed in adult and pediatric sarcomas. Our data indicate that VGLL2-NCOA2 is an oncogene that leverages developmental programs for tumorigenesis and that reactivation or persistence of ARF6 could represent a therapeutic opportunity.

Graphical abstract



In brief

Watson et al. use a cross-species comparative approach to develop zebrafish and mouse models of a rare fusion-driven pediatric sarcoma. These models recapitulate the human disease, and the integration of these systems identifies a conserved developmental target, ARF6. ARF6 cooperates with the primary oncogenic driver, representing a potential therapeutic opportunity.

INTRODUCTION

With the broader implementation of clinical sequencing efforts, there has been a significant increase in the number of gene fusions that have been identified in patient tumors. Particularly in sarcomas, there is a predilection for fusion oncogenes with a predicted 20%–49% of sarcomas containing a gene fusion event.[1–4] In pediatric sarcomas, there are typically few other somatic mutations, highlighting fusions as the defining oncogenic drivers in these diseases.[5–7] However, developing cell culture or animal models of these newly identified fusions is challenging, hindering the verification of their role in tumorigenesis and the identification of potential targeted therapeutics. A pressing need exists for tractable genetic models to understand the shared and divergent biology between genetic drivers, and ultimately provide a platform to identify mechanism based therapies.

One example of a recently discovered fusion gene is VGLL2-NCOA2, which was identified from a cohort of congenital pediatric sarcomas that did not contain the classical pediatric sarcoma gene fusions.[4,8] Both vestigial-like family member 2 (VGLL2) and nuclear receptor coactivator 2 (NCOA2) are transcriptional coactivators, but their function in this

gene fusion has not been studied. VGLL2 (previously named VITO-1) plays a role in muscle development and targets TEA domain transcription factor (TEAD) proteins to their appropriate contexts during the differentiation process.[9–11] Additionally, in zebrafish models, *vgll2a* is required for neural crest cell survival and craniofacial development.[12] NCOA2 is a common 3′ fusion partner in non-spindle cell sarcoma contexts; examples include HEY1-NCOA2 in mesenchymal chondrosarcoma, MEIS1-NCOA2 in genitourinary and gynecologic tract spindle cell sarcomas and intraosseous rhabdomyosarcomas, and PAX3-NCOA2 in alveolar rhabdomyosarcoma.[13–17] Suggestively, TEAD1-NCOA2 fusions have also been identified in infantile rhabdomyosarcoma, indicating converging developmental processes for this subtype.[18,19]

The VGLL2-NCOA2 fusion has been described in multiple patients by multiple groups. There are two forms of the gene fusion, *VGLL2* exons 1–2 and either *NCOA2* exons 13–23 or 14–23. The translocation involves two genomic events, an inversion of *VGLL2*, followed by a break and chromosomal translocation of chromosomes 6 and 8. These fusions are the defining genetic features of the disease; however, pathologically and transcriptionally, the tumors are heterogeneous.[4,8,20] The biological explanation for this heterogeneity remains unclear.

Clinically, VGLL2-NCOA2 tumors typically present between 0 and 1 year of age and have a spindle and sclerosing cellular morphology.[4,8,18,20–22] Fusion-driven congenital rhabdomyosarcomas generally have a favorable prognosis with successful surgery and chemotherapy.[8] However, non-resectable tumors are challenging to treat, and chemotherapy exposure carries a significant risk of short- and long-term adverse effects. Moreover, clinical outcomes of VGLL2-NCOA2 rearranged tumors are variable. Recently, a case study with four VGLL2 rearranged tumors described three patients that had multi-metastatic spread, and two that died from disease. This study suggests that complete surgical resection is critical (but not always attainable) for long-term benefit, and that VGLL2 fusions are capable of aggressive disease.[22] Given that these tumors occur at a developmentally sensitive age, effective targeted therapies are critically needed to improve disease outcomes and ameliorate off-target effects from toxic and generalized therapies. Therefore, we set out to test the tumorigenic capacity of this fusion gene, define its oncogenic program, and identify potential therapeutic targets.

Previously, we described a mosaic strategy using vertebrate transgenic zebrafish models that addresses the need to readily generate animal models that recapitulate the genetics and presentation of the human disease. This strategy has been successful for PAX3-FOXO1-driven rhabdomyosarcoma and CIC-DUX4-mediated Ewing-like sarcoma, indicating that zebrafish are a powerful tool for personalized medicine approaches and for the study of rare cancers and disease.[23,24] Here, we implement a functional genomics approach to generate human VGLL2-NCOA2 fusion-oncogene-driven zebrafish and mouse allograft models of tumorigenesis. We find that the fusion is indeed sufficient for aggressive tumorigenesis in the zebrafish and mouse systems, resulting in tumors that recapitulate histological and molecular features of the human disease. Zebrafish VGLL2-NCOA2 tumors express markers indicative of arrested skeletal muscle development and poor differentiation. RNA sequencing (RNA-seq) of zebrafish and mouse allograft tumors identified ARF6, a small

GTPase, as an overexpressed and potentially druggable target. ARF6 is expressed not only in human VGLL2-NCOA2 tumors but also across a panel of pediatric sarcomas and other cancers, whereas there is no expression in normal tissue counterparts. We demonstrate in cell culture systems that ARF6 and VGLL2-NCOA2 genetically cooperate to inhibit myogenesis, and that ARF6 knockout suppresses VGLL2-NCOA2-mediated colony formation. Our study suggests that this congenital rhabdomyosarcoma fusion oncogene coopts developmental programs to induce its effects, and that zebrafish are high-fidelity models of these systems. Further, these models have the potential to identify tumor-specific therapeutic opportunities for these rare diseases.

RESULTS

VGLL2-NCOA2 is an oncogene and is tumorigenic in zebrafish

Previously, we and others have shown that VGLL2-NCOA2 is a fusion gene exclusively found in a subset of congenital rhabdomyosarcomas.[4,8] Five patients under the age of five presented with soft tissue tumors that were originally diagnosed as embryonal rhabdomyosarcomas. Clinical sequencing of patient tumors identified the presence of a VGLL2-NCOA2 fusion. Table S1 describes the clinical presentation of each case that was identified in Watson et al.[4] The fusion in all tumors contains an intronic breakpoint in *VGLL2* that results in the incorporation of the first two exons of *VGLL2*. However, there were two intronic breakpoints seen in *NCOA2*; five tumors had a breakpoint that resulted in inclusion of exons 14 through 23 of *NCOA2*, whereas one tumor had an additional fusion product with the inclusion of exons 13–23 (Table S1; Figure S1).[4]

The coding sequence of the human fusion was cloned out of a primary patient tumor that harbored the *NCOA2* exon 14 breakpoint (Figure S1A). Gateway cloning was used to generate a cytomegalovirus (CMV)-GFP2A-*VGLL2NCOA2* construct for expression of the human fusion gene in zebrafish systems. The cassette is flanked on both sides by Tol2 sites, which facilitates genomic integration when injected with Tol2 mRNA (Figure 1A).[25] The viral 2A sequence allows for GFP and *VGLL2-NCOA2* to be transcribed on the same mRNA but to function as independent proteins. We injected this expression construct and Tol2 mRNA into one-cell-stage wild-type zebrafish embryos and observed robust GFP expression at 24 h post fertilization (hpf) in >95% of embryos (data not shown). Zebrafish were monitored for tumor development by observing gross morphology and GFP expression, which served as a proxy for fusion-oncogene expression. Zebrafish presented with tumors as early as 1 month, with 20% of zebrafish developing tumors by 50 days, and ~30% by 6 months. In control cohorts injected with CMV-GFP-pA, tumors were never observed (Figure 1B). The sufficiency of the VGLL2-NCOA2 fusion to generate tumors in zebrafish confirms its function as an oncogene.

We classified all generated tumors by location on the fish, binning the fish as depicted in Figure 1C. We observed that tumors arising on the tail were the most common (44%; n = 33 out of 75), back and head tumors occurred in equal numbers (23%; n = 17 out of 75 and 20%; n = 15 out of 75 respectively), and ventral tumors were the least common (9%; n = 7 out of 75) (Figure 1C). The tumor onset was similar for each location with the exception of the ventral tumors, which arose later compared with the other tumor locations (log rank, p <

0.05) (Figure 1D). Interestingly, three zebrafish presented with more than one tumor around 47 days of age (Figures 1C, 3C, Table S2).

Shown in Figure 1E is the gross morphology and GFP fluorescence of four representative zebrafish tumors and their corresponding hematoxylin and eosin (H&E) stains of transverse sections through the GFP-positive (GFP+) region. All four of these tumors are mesenchymal tumors diagnosed as sarcomas, with infiltrating cells, and a variety of cell types including spindle cells, round cells, and pleomorphic cells, although in differing proportions. All tumors are mitotically active. The tumors contain dispersed nuclear chromatin and poorly defined cell membranes; prominent nucleoli are seen in the head tumor. Fascicular arrangements of cells are observed in the tail and back tumors, and there is a notable lack of background stroma. In the ventral tumor, there is a higher proportion of stroma and more pleomorphic and spindle cells. Collagen is present in variable amounts in the ventral tumor. In the head tumor, the predominant cell types are pleomorphic and round cells with accompanying necrosis. Altogether, these tumors are consistent with the histology of human sarcomas. Further examples of this histology are presented as higher-magnification images in Figure 1F.

We also generated constructs in which GFP-tagged VGLL2-NCOA2 was driven by the following zebrafish promoters: MCS-beta-globin-Splice Acceptor, β-actin, ubiquitin, and unc503. Following zebrafish microinjection, we observed GFP expression at 24 h post fertilization, indicating successful genomic integration of the constructs (Figure S2A). With MCS-beta-globin-SA-GFP2A-VGLL2NCOA2, we observed a total of three fish out of 42 that presented with tumors, two diagnosed as sarcomas, and one as neuroblastoma/retinoblastoma (Figure S2B). The other promoters failed to produce any tumors, although some GFP+ adult fish were observed (Figure S2C). Given the low incidence, we focused our efforts on the CMV-driven model, which recapitulated the human disease with the highest fidelity. All zebrafish tumors in the remainder of the study are generated from the CMV-GFP2A-VGLL2NCOA2 construct.

VGLL2-NCOA2 zebrafish tumors recapitulate the human disease

To validate that zebrafish tumors expressed the VGLL2-NCOA2 fusion at the protein level, we performed immunohistochemistry on serial transverse sections of zebrafish tumors using a human NCOA2 antibody (Figure 2A). The tumor cells identified with H&E staining were also positive for VGLL2-NCOA2. Additionally, a wild-type zebrafish control section was largely negative for NCOA2, indicating that VGLL2-NCOA2 is specifically expressed in the GFP+ tumors. We confirmed by RNAscope that VGLL2-NCOA2 tumors expressed the zebrafish orthologs of diagnostic rhabdomyosarcoma markers, including *myod1*, *myog*, and *desma*. Shown are examples of their expression patterns in wild-type skeletal muscle compared with tumors (Figure 2B). These muscle markers are used to clinically characterize the human disease (Table S2).[4,8] Further, tumor samples had high levels of Pcna expression (Figure 2B). These findings were consistent across multiple tumors, with synthesized data presented in Table S3. The RNAscope findings are consistent with our RT-PCR data showing that VGLL2-NCOA2 tumors expressed the fusion at the mRNA level, and, except for one tumor, expressed the muscle markers *myod1*, *myog*, and *desma* (Figure S3A).

We then performed RNA-seq on a cohort of zebrafish VGLL2-NCOA2 sarcomas to determine how well they transcriptionally recapitulate the human disease. To do this, we used a comparison of zebrafish or human VGLL2-NCOA2 tumors to mature skeletal muscle. VGLL2-NCOA2 zebrafish and human tumor expression signatures relative to mature skeletal muscle were then compared in an agreement of differential expression (AGDEX) analysis with a minimum of 5,000 permutations (Figure 2C).[4,24,27] There were 10,661 genes that were shared and differentially expressed between the zebrafish and human VGLL2-NCOA2 tumor types compared with mature skeletal muscle. In Figure 2C, the cosine of the angle (cos) and difference of proportions (dop) statistics are both positive, suggesting the zebrafish model recapitulates the human disease. This is supported by a permutation analysis by randomly sampling from kernel density distributions matching the distributions of the original data indicating an enrichment of differentially expressed genes shared between human and zebrafish tumors (Figure 2C). Gene set enrichment analysis of these 10,661 overlapping and differentially expressed genes revealed multiple biological pathway terms related to activation of extracellular structure organization and inhibition of skeletal muscle differentiation processes (Figures 2D and S3). This significant overlap between zebrafish and human VGLL2-NCOA2 sarcomas is also observed when compared with CIC-DUX4 Ewing-like sarcoma rather than skeletal muscle (Figures S3B and S3C). Importantly, in contrast, there was little overlap between a KRAS-driven rhabdomyosarcoma zebrafish tumor model and VGLL2-NCOA2 human or zebrafish tumors (Figures S3D and S3E), highlighting the power of this strategy. Overall, these data support implementing zebrafish tumor models to understand human VGLL2-NCOA2 disease biology.

VGLL2-NCOA2 leverages developmental genes and programs during tumor formation

Given the early onset of the tumors in the human and zebrafish cohorts, and the possible link to development observed in the RNA-seq data, we hypothesized that the fusion oncogene was co-opting developmental genes to promote tumorigenesis. RNA-seq data from 18 zebrafish developmental time points from one-cell to 5 days post fertilization were analyzed alongside VGLL2-NCOA2 zebrafish tumors and were plotted using the first three principal components.[26] This analysis revealed a large cohort of tumors clustered with the developmental time points of segmentation (Figure 3A; stage S9–S11). This specific developmental time point is striking as it corresponds to somitic muscle development in zebrafish (Figure 3B). The tumors that clustered with segmentation include the tumors on the head (green, Figures 3A and 3C), tail (purple, Figures 3A and 3C), and several of the ventral tumors (blue, Figures 3A and 3C). However, the tumors on the back (orange, Figures 3A and 3C) did not cluster with this group, instead forming a second cluster distant from the developmental trajectory.

To identify genes in VGLL2-NCOA2 tumors that direct this clustering with segmentation, we performed a differential gene expression analysis from the RNA-seq data. First, the cohort of n = 18 VGLL2-NCOA2 zebrafish tumors was compared with n = 7 mature zebrafish skeletal muscle samples, identifying 385 differentially regulated genes. Next, n = 3 samples of embryonic zebrafish during somitogenesis (1–4 somites, 14–19 somites, and 20–25 somites) were compared with mature zebrafish skeletal muscle; and 303 differentially regulated genes were found. Then, these 385 and 303 gene sets were intersected to

identify 27 genes shared in both development and tumor datasets as differentially expressed compared with mature skeletal muscle (Figure 4A). These 27 genes represent potential VGLL2-NCOA2 signatures that are reactivated, inappropriately persist, or are never turned on in tumors. To demonstrate the magnitude of this expression, the fragments per kilobase of exon per million mapped fragments (FPKM) of these 27 genes was plotted, including genes downregulated in the tumors and developing muscle compared with mature skeletal muscle (Figure 4B), and genes that were conversely upregulated in tumors and development compared with mature skeletal muscle (Figure 4C).

The expression levels of these 27 genes were then evaluated in a complementary mouse allograft VGLL2-NCOA2 model and human VGLL2-NCOA2 tumors compared with mature skeletal muscle. Plotted is the fold change of these 27 genes in zebrafish relative to their expression in our mouse VGLL2-NCOA2 allograft model (Figure 4D) and in human VGLL2-NCOA2 tumors (Figure 4E). This allowed us to further refine our search for genes that are commonly dysregulated in all three VGLL2-NCOA2 contexts. Interestingly, although muscle regulatory factors were expressed in VGLL2-NCOA2 tumors, there was not a significant difference in the expression level compared with mature skeletal muscle or developmental samples (Figure S4A). Overall, this analysis identified genes with expression in tumors that more closely matched embryos rather than mature skeletal muscle, suggesting that VGLL2-NCOA2 could co-opt embryonic gene expression programs to promote tumor growth.

Cross-species comparative oncology analysis identifies ARF6 as upregulated in zebrafish, mouse allograft, and human VGLL2-NCOA2 tumors

Gene ontology analysis of genes differentially expressed between VGLL2-NCOA2 zebrafish tumors and mature skeletal muscle highlighted several GTPases and their functions (Figure S4B). One of the genes significantly overexpressed in all three VGLL2-NCOA2 tumor contexts (zebrafish, mouse allograft, and human), as well as in development, was *arf6b*, a small GTPase involved in endocytosis and membrane receptor recycling and in actin reorganization (Figure 4C).[31] In zebrafish, the *arf6* gene is duplicated and is referred to as *arf6a* and *arf6b*. The zebrafish Arf6a and Arf6b protein sequences are identical and share 99% identity with the ARF6 human ortholog (Figure S4C). We performed qRT-PCR on both *arf6a* and *arf6b* to determine mRNA expression levels in an independent zebrafish tumor cohort, and found that *arf6a* was modestly elevated and *arf6b* was significantly elevated in development and tumor samples compared with mature skeletal muscle (Figure 5A). This pattern was even more striking at the protein level, where Arf6 protein expression was significantly upregulated in the context of VGLL2-NCOA2 tumors compared with embryos and mature skeletal muscle (Figures 5B and 5C). To verify the spatial expression of Arf6, we performed Arf6 immunohistochemistry on serial sections of zebrafish VGLL2-NCOA2 tumors and found that Arf6 protein expression overlaid with tumor cells and human anti-NCOA2 positive cells (Figure 5C). A control H&E and Arf6 IHC stain from a wild-type zebrafish transverse section is shown in Figure S4D, and the control anti-NCOA2 stain is shown in Figure 2A. These data suggest that Arf6 is highly expressed in VGLL2-NCOA2 established tumors and could be contributing to the disease process.

To implement a cross-species comparative oncology approach, we generated a C2C12 mouse myoblast cell line that constitutively and stably overexpressed the human form of the VGLL2-NCOA2 fusion that is used in our zebrafish models. We found that the fusion was highly expressed at the protein level even after passage and selection in culture (Figure 6A). Further, when either C2C12-pCDNA3.1 control or C2C12-VGLL2NCOA2 cells were allografted into severe combined immunodeficiency (SCID) or Nude mice, cells overexpressing the VGLL2-NCOA2 fusion generated aggressive tumors that rapidly grew with initial detection at 11 days post injection and termination of the experiment at 26–28 days post injection (Figures 6B and 6C). This indicates that the human form of the fusion oncogene is transforming in multiple systems.

On pathological analysis, C2C12-pcDNA3.1 control allografts resulted only in small lesions at the site of injection, composed of bland spindle cells (Figures 6D and S5A). Conversely, C2C12-VGLL2NCOA2 allografts consistently produced tumors showing features of high-grade sarcoma, with both spindle cell and pleomorphic morphology (Figures 6D; S5B). C2C12-VGLL2NCOA2 allografts also exhibited muscle invasion and tumor necrosis (Figures S5C–S5F). Further, C2C12-VGLL2NCOA2 allograft tumors expressed clinically diagnostic markers of infantile rhabdomyosarcoma, including heterogeneous expression of Desmin and Myogenin (Figure 6E; Table S4). The fusion-driven tumors also exhibited elevated staining for Ki67, suggesting an increase in proliferation compared with C2C12-pcDNA3.1 controls (Figure 6E). We performed RNA-seq on a subset of these generated C2C12-pCDNA3.1 control and C2C12-VGLL2NCOA2 allografts and compared their transcriptional profiles with that of the human disease, in a similar strategy that was used for zebrafish VGLL2-NCOA2 tumors. Using an AGDEX analysis, we found that our allograft C2C12-VGLL2NCOA2 model was consistent with the human disease, and shared 9,763 differentially expressed genes (Figure S6A). A UMAP analysis of transcriptional signatures shows that C2C12-VGLL2NCOA2 tumors cluster with human VGLL2-NCOA2 tumors, whereas C2C12-pcDNA3.1 controls cluster with mouse skeletal muscle (Figure S6B). We performed a gene set enrichment analysis of these 9,763 shared mouse and human differentially expressed genes, which revealed similarly regulated pathways as in our zebrafish model, including enrichment of extracellular matrix and suppression of skeletal muscle development (Figure S6C). In cell culture, C2C12-VGLL2NCOA2 exhibits an increase in H3K27 acetylation at the Arf6 promoter in the context of the fusion, and, in VGLL2-NCOA2 allograft tumors, *Arf6* was significantly overexpressed compared with controls (Figures S6D, 4D, and 6F). Further, we found evidence of Arf6 protein expression by immunohistochemistry in the C2C12-VGLL2NCOA2 allograft tumor models of the disease (Figure 6G; Table S4).

Functional cooperation of VGLL2-NCOA2 and ARF6 for sarcomagenesis

Our cross-species comparative oncology approach identified ARF6 as a developmentally regulated gene that is highly expressed in VGLL2-NCOA2 tumors. Next, we wanted to evaluate the functional contribution of ARF6 to the disease process. To do this, we generated C2C12 Arf6 knockout clones using a CRISPR-Cas9 strategy with a guide targeting *Arf6*. These cells have a premature stop in Arf6 resulting in a 109-amino-acid protein product and no detectable protein expression (Figure 7A). We then expressed the human *VGLL2-*

NCOA2 construct using the pcDNA3.1 vector as described in Figure 6, and, after selection in culture, generated stable cell lines with or without Arf6 knockout or VGLL2-NCOA2 expression (Figure 6A). First, we assessed whether the VGLL2-NCOA2 fusion or Arf6 knockout had any impact on cellular proliferation, and found there was no difference across all groups (Figures S7A and S7B), suggesting an alternative mechanism for cooperation. Next, we assessed how the VGLL2-NCOA2 fusion with or without Arf6 knockout affected the myogenic process. We seeded confluent C2C12 myoblasts onto porcine gelatin-coated dishes and serum starved them for 5 days after plating. C2C12 myoblasts should fuse into multinucleated myotubes that express myogenin (*MyoG*), an earlier fusion marker, and myosin light and heavy chain (*Myl1*, *Myh1*), more terminal markers of differentiation. We found that the VGLL2-NCOA2 fusion significantly reduced the mRNA expression of *MyoG*, *Myl1*, and *Myh1* (Figures 7B, S7C, and S7D) by day 6 of fusion. However, Arf6 knockout mitigated this effect, restoring *MyoG*, *Myl1*, and *Myh1* expression levels and rescuing the overall morphology of the cells (Figures 7B, 7C, S7C, and S7D). These assays were complemented with a soft agar colony formation assay to determine the tumorigenic capacity of these cells in cultures. We seeded C2C12-VGLL2NCOA2 or control with or without Arf6 knockout in soft agar and allowed colonies to develop for 2 weeks, after which they were counted. Shown are examples of colonies formed with each condition, with both C2C12 wild-type and Arf6 knockout transfected with control pcDNA3.1 empty plasmid failing to generate colonies (Figure 7D). Quantification of the number of colonies per well indicates that Arf6 knockout suppresses VGLL2-NCOA2 transformation capacity by 2.8-fold (Figure 7E). Together, these data suggest that Arf6 is a genetic cooperating event in the disease and could be a potential therapeutic target.

Next, we analyzed RNA-seq data from patient tumors, and performed differential expression analysis between human VGLL2-NCOA2 tumors and human mature skeletal muscle. We found that *ARF6* mRNA was modestly overexpressed in the human disease (Figures 4E and 7F). To more broadly assess the potential applicability of ARF6 inhibition in sarcoma, we analyzed *ARF6* expression in a panel of sarcomas and found that this observation may be translatable to many other sarcomas. There was similar overexpression of *ARF6* compared with mature skeletal muscle in adult sarcomas in general, and in pediatric sarcomas, including Ewing sarcoma, osteosarcoma, and rhabdomyosarcoma. Notably, this did not hold true for clear cell sarcoma of kidney (Figure 7G).

Overall, development is a tightly controlled and orchestrated process involving appropriate gene activation and inactivation to dictate migration, cell division, and structure formation. Our data suggest that VGLL2-NCOA2 is an oncogene that potentially leverages developmental programs in order to promote sarcomagenesis. There are two distinct, but not mutually exclusive, possibilities for the mechanism: one in which, during maturation, developmental genes continuously decrease in expression and are then are reactivated either at the onset or during tumorigenesis. Alternatively, developmental genes may never decrease in expression and instead inappropriately persist throughout the course of tumorigenesis. Further studies will be required to delineate the mechanisms by which VGLL2-NCOA2 mediates these changes in gene expression and their functional role in tumorigenesis.

DISCUSSION

Each sarcoma gene fusion produces a slightly different disease that warrants the development of a genetic animal model system to understand the biology and to identify potential therapeutic targets. This is difficult to accomplish. Often, the sarcoma cell of origin is unknown, making screening putative cellular origins in mouse models costly and burdensome. Zebrafish complement mouse models as a timely approach to make significant contributions to understanding rare genetic diseases and pediatric cancer. Zebrafish capture a relevant developmental context and have conserved genetics and molecular signaling pathways.[32–34] By applying these models, one can complement clinical sequencing efforts and rapidly define the biology of identified gene fusions to determine the most significant features of the disease. Previously, we described a genetic approach focused on PAX3-FOXO1 rhabdomyosarcoma and CIC-DUX4 sarcoma transgenic zebrafish modeling.[23,24] We have now broadened our approach to incorporate additional rhabdomyosarcoma fusion oncogenes. This modular platform can be implemented for any rhabdomyosarcoma fusion identified from clinical sequencing efforts that has yet to be studied in a laboratory setting.

Here, we applied this strategy to demonstrate that *VGLL2-NCOA2* is an oncogene that generates sarcoma in genetic zebrafish and mouse allograft systems. When the human form of the fusion is integrated into the zebrafish genome, zebrafish develop tumors that are histologically and molecularly consistent with the human disease. Introduction of the VGLL2-NCOA2 fusion into C2C12 mouse myoblasts and allografts of these models result in aggressive tumors by 3 weeks post injection. These data indicate the fusion alone is sufficient for transformation and does not require cooperating events. This highlights the capacity of the zebrafish system to faithfully capture distinct sarcoma subtypes based on the driving oncogene. For example, zebrafish VGLL2-NCOA2 tumors present at a much earlier age and have distinct histology and unique gene expression profiles compared with our PAX3-FOXO1 zebrafish tumor model.[23] This is consistent with the human disease, in which patients with VGLL2-NCOA2 tumors present between 0 and 1 year with distinct histology from PAX3-FOXO1-driven rhabdomyosarcoma. The CMV promoter efficiently generated VGLL2-NCOA2 tumors in our zebrafish and mouse models and drove high expression levels of VGLL2-NCOA2, resulting in the aggressive pathology and higher levels of expression than is observed in the human disease (data not shown). However, the molecular features and general pathology are consistent, but this opens the possibility that utilizing more nuanced promoters could reflect the heterogeneity observed in the human disease. Our data underscore the importance of creating and studying animal models of rare genetic diseases to determine biological differences and similarities in pediatric sarcomas.

In infantile rhabdomyosarcoma, there are other fusion partners for VGLL2, including CITED2.[4,8,18,20] However, transcriptional clustering indicates VGLL2-fused tumors have shared signatures, suggesting that VGLL2 and its regulatory elements are the predominant drivers for this disease.[4,20] Insight into normal VGLL2 expression and its targets may help understand its activity in the fusion. VGLL2 is highly expressed during mouse embryonic development in the somitic myotome and pharyngeal pouch, but, in adult tissues, its expression is restricted to skeletal muscle.[35] In human development, VGLL2 is solely expressed in fetal skeletal muscle, and is absent from cardiac, liver, thymus, or

brain.[35] This specific expression and interaction partners could explain the immature muscle features of the disease. A *Vgll2* knockout mouse has helped elucidate its role in skeletal muscle development, and *Vgll2* knockout mice exhibit defects in skeletal muscle fiber type composition and exercise exhaustion.[36,37] This model has confirmed direct interaction with Vgll2 and Tead1/4 by co-immunoprecipitation in embryonic skeletal muscle.[36] Intriguingly, TEAD1-NCOA2 fusions are another defining genetic event in congenital sarcoma with a similar clinical presentation.[18,19] Perhaps VGLL2 is responsible for regulating targets through TEAD1/4 proteins for precise developmental control, and this could be co-opted or deregulated by the VGLL2-NCOA2 fusion.

Transcriptional analysis of VGLL2-NCOA2 zebrafish tumors identified a subset that clustered with somitogenesis in the developing zebrafish, distinct from zebrafish mature skeletal muscle. These results suggest that tumor programs either reactivate developmental genes or that developmental genes inappropriately persist. Our differential gene expression analysis highlighted small GTPases in zebrafish, mouse allograft, and human VGLL2-NCOA2 tumor samples. Specifically, of these GTPases, Arf6 had increased protein expression in zebrafish tumors and was completely absent in mature skeletal muscle. The sufficiency of the VGLL2-NCOA2 fusion and overexpression of Arf6 in the context of the fusion was confirmed in C2C12 mouse allograft models. In the human disease, *ARF6* mRNA is overexpressed in VGLL2-NCOA2 patient tumors compared with mature skeletal muscle. Our functional analyses indicate that ARF6 and VGLL2-NCOA2 genetically cooperate to inhibit the skeletal muscle differentiation process and suppress colony formation *in vitro*. More broadly, in a panel of pediatric and adult sarcomas, *ARF6* is overexpressed compared with mature skeletal muscle, suggesting that ARF6 could be a potential therapeutic target in many sarcoma contexts. Perhaps this has been overlooked because *ARF6* mRNA levels do not always correlate with ARF6 protein expression (Figures 5A–5C),[38] and ARF6 is not a defining feature of any sarcoma subtype but represents a shared feature of the disease.

ARF6 is a highly conserved mediator of developmental processes, myogenesis, and carcinogenesis, all of which could contribute to its function in VGLL2-NCOA2 tumors. Studies in sea urchins suggest that *Arf6* knockdown impairs gastrulation and alters directed migration of primary mesenchyme cells.[39] This is complemented from data in *Arf6* null mice, which are embryonic lethal indicating a critical developmental role.[40] ARF6 is also a core member of the complex required for myoblast fusion and muscle development.[41] Impaired muscle differentiation is a hallmark of rhabdomyosarcoma, and these processes are aberrant in VGLL2-NCOA2 tumors, which express the early markers of myogenesis but appear histologically undifferentiated. ARF6 activation or overexpression has a role in other cancer contexts and promotes proliferation, invasion, and metastasis or predicts poor prognosis in breast cancer, melanoma, lung cancer, and other solid tumors.[38,42–46] A remaining question is the contribution of ARF6 to tumor initiation versus maintenance (or both), which is not clear and will be a focus of future lines of research. ARF6 has also been proposed as a therapeutic opportunity in cancer.[47] In uveal melanoma, an ARF6 inhibitor, NAV-2729 inhibited tumor growth in xenograft models of the disease.[43] Although ARF6 is broadly expressed, the tolerability of NAV-2729 in these studies indicates that there may be a

therapeutic window. Whether this is true in VGLL2-NCOA2-driven rhabdomyosarcoma, or sarcoma in general, remains to be seen.

Overall, we developed transgenic zebrafish and mouse allograft models of VGLL2-NCOA2 fusion-driven sarcoma and our cross-species comparative oncology approach identified a potential therapeutic opportunity, ARF6. Our findings also highlight how studying rare disease can illuminate molecular mechanisms or systems that are more broadly applicable. We have now presented a strategy for building fusion-oncogene-driven animal models to define biology, identify therapeutic opportunities, and intersect data to determine conserved sarcoma disease drivers.

Limitations of the study

Our findings demonstrate a conserved role for the human VGLL2-NCOA2 fusion oncogenic function in zebrafish and mouse allograft models of the disease. We utilized the CMV promoter to generate these models, which drives high levels of expression of the VGLL2-NCOA2 fusion. Although the CMV promoter driving VGLL2-NCOA2 generated tumors that recapitulate the human disease, a more nuanced approach would involve restricting the expression of the fusion to various cell lineages to determine their capacity for transformation. Such approaches could better reflect the heterogeneity that is observed in the human disease.

Further, integration of transcriptional data from our VGLL2-NCOA2 zebrafish and mouse allograft models with patient tumor data identified overexpression of ARF6 as a shared disease feature. To investigate the role of ARF6 in the disease, we generated a C2C12 Arf6 CRISPR-Cas9 knockout cell line. However, the Arf6 cell knockouts that were generated only address the role of Arf6 in tumor initiation and transformation events. The long-term goal is to understand Arf6 therapeutically, which will require temporal control of knockdown during different stages of tumorigenesis, or as a complementary approach, utilizing a small-molecule NAV-2729 Arf6 inhibitor.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and request for resources and reagents should be directed to and will be fulfilled by the lead contact, Genevieve Kendall (Genevieve.Kendall@NationwideChildrens.org).

Materials availability—Requests for materials generated in this study should be directed to the lead contact, Genevieve Kendall (Genevieve.Kendall@NationwideChildrens.org), and are available with a completed Materials Transfer Agreement.

Data and code availability

- Zebrafish RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. The accession number is listed in the key resources table. This paper also analyzes existing, publicly available data or data with

access that is available upon request. Accession numbers for the datasets are listed in the key resources table. Original western blot images and microscopy data will be shared by the lead contact upon request.

- This paper does not report original code. The code used to analyze the data is available on GitHub (https://github.com/MVesuviusC/2022_VGLL2_NCOA2_paper).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Zebrafish—*Danio rerio* were housed in an AAALAC-accredited, USDA-registered, OLAW-assured Aquaneering facility in compliance with the Guide for the Care and Use of Laboratory Animals. Vertebrate animal work was overseen by the UT Southwestern Medical Center or the Abigail Wexner Research Institute at Nationwide Children's Hospital IACUC committee and Animal Resources Core. Zebrafish are free of *Pseudoloma neurophilia*, *Pleistophora hyphessobryconis*, *Pseudocapillaria tomentosa*, *Mycobacterium spp.*, *Edwardsiella ictalurid*, *Ichthyophthirius multifilis*, *Flavobacterium columnare*, and zebrafish picornavirus (ZfPV1) as determined by a quarterly sentinel monitoring program. The fish were housed at a density of 5–12 fish per liter in mixed-sex groups in 0.8-L, 1.8-L, 2.8-L, or 6-L tanks on a recirculating system (Aquaneering, San Diego, CA) in 28°C water (conductivity, 510 to 600 μS; pH, 7.3 to 7.7; hardness, 80 ppm; alkalinity, 80 ppm; dissolved oxygen, greater than 6 mg/L; ammonia, 0 ppm; nitrate, 0 to 0.5 ppm; and nitrite, 0 ppm) in a room with a 14:10-h light:dark cycle. System water was carbon-filtered municipal tap water, filtered through a 20-μm pleated particulate filter, and exposed to 40W UV light. As juveniles (5–30 days) the fish were fed live rotifer reeds, and as adults (>30 days) fish were fed twice daily with a commercial pelleted diet. Both male and female zebrafish were utilized in single-cell injection experiments since zebrafish sex is not determined until 30–60 days of age in our facility. Wild-type lines used in this study were AB, WIK, TL and were obtained from the Zebrafish International Resource Center (ZIRC). AB/TL were bred in our facility by crossing AB and TL. The *p53* mutant line, $tp53^{M214K}$, was a kind gift from Tom Look.[52]

Mice—Animal care and use for this study were performed in accordance with the recommendations of the European Community (2010/63/UE) for the care and use of laboratory animals. Experimental procedures were specifically approved by the ethics committee of the Institut Curie CEEA-IC #118 (Authorization APAFIS#11206–2017090816044613-v2 given by National Authority) in compliance with international guidelines. Mouse strains used were SCID (Charles River Laboratories CB17/Icr-$Prkdc^{scid}$/IcrIcoCrl) or Swiss Nude (Charles River Laboratories Crl:NU(Ico)-$Foxn1^{nu}$) mice. All injected mice were female and, except for one Nude mouse at 136 days of age, were 46 days old at the time of injection.

Cells—C2C12 mouse myoblast cells were obtained from American Type Culture Collection (#CRL-1722) and were cultured in growth media consisting of 10% or 20%

fetal bovine serum in DMEM (GE healthcare #SH30022.01 or Gibco #11995065). Cells were maintained in a humidified incubator at 37°C with 5% CO2. Cell identity was verified by STR profiling and was confirmed negative by qPCR for mycoplasma through Genetica (Lab Corp).

Human sarcoma and skeletal muscle samples—Previously generated human sarcoma or mature skeletal muscle RNA-seq data was used in this paper. The accession numbers are provided in the key resources table and data are available upon request from the original depositor.

METHOD DETAILS

Plasmids and cloning—*VGLL2-NCOA2* coding sequence was cloned out of a patient tumor. In the fusion used to develop the zebrafish model, exons 1 and 2 of *VGLL2* is fused to exon 14 through 23 of *NCOA2*. Sanger sequencing revealed that serine at position 544 of the resulting fusion protein had the nucleotide sequence TCT, a silent mutation compared to the human reference sequence of TCC. The fusion was cloned into the Gateway expression system by adding 5′ and-3′ ATT sites (attb2r/attb3) using a high-fidelity polymerase. Purified PCR product was then cloned into a 3′ entry clone as described in Kendall and Amatruda, 2016.[70] The Tol2 kit components: destination vector pDestTol2pA2, 3′ SV40 late polyA signal construct, beta actin promoter, cmv promoter, and multiple cloning site were used to generate constructs for expression in zebrafish.[53] The ubi promoter was a kind gift from Len Zon (Addgene #27320),[54] and the unc503 promoter from Peter Currie (Addgene #27320).[55] Middle entry beta globin intron and splice acceptor was from Koichi Kawakami.[56] The GFP-viral2A sequence were a gift from Steven Leach[71] and were sub-cloned into a middle entry Gateway expression system.[23] Tol2 mRNA was synthesized *in vitro* from pCS2FA-transposase which is from Koichi Kawakami.[25] The CMV-GFP2A-pA construct was generated as described in Kendall et al., 2018.[23] The zebrafish expression constructs generated for this study include: CMV-GFP2A-VGLL2NCOA2, BetaActin-GFP2A-VGLL2NCOA2, ubi-GFP2A-VGLL2NCOA2, unc503-GFP2A-VGLL2NCOA2, MCS-beta-globin-SpliceAcceptor-GFP2A-VGLL2NCOA2. For expression in C2C12 cells, the VGLL2-NCOA2 fusion was sub-cloned into pcDNA3.1 using primers that added BamH1 and Xho1 digest sites to the 5′ and-3′ ends of the transcript, respectively (Table S5).

Zebrafish embryo injections—Zebrafish embryos were injected at the single-cell stage with an injection mix containing 50 ng/μL of Tol2 transposase mRNA, 50 ng/μL of the VGLL2-NCOA2 DNA construct and equimolar amounts of control constructs, 0.1% phenol red and 3x Danieau's buffer.

Zebrafish tumor collection and processing—Zebrafish with tumors were euthanized in 2 mg/mL Tricaine-S solution and screened using the Nikon SMZ25 fluorescent stereoscope to detect GFP expression. Tumors were resected and flash frozen in liquid nitrogen. Total RNA isolation was performed on frozen tissue using the RNeasy Micro Kit (Qiagen #74004). The remaining zebrafish was fixed in 4% paraformaldehyde in PBS for 48 h shaking at 4°C, subsequently decalcified in 0.5M EDTA for 5 days at room

temperature and mounted in paraffin blocks for microtome sectioning. Sections were taken at 5–10 micron intervals. De-paraffinized slides were stained with hematoxylin and eosin as described in Kendall and Amatruda, 2016.[70]

Zebrafish tumor incidence—All zebrafish that survived past thirty days were counted for tumor incidence curves. Zebrafish with no GFP fluorescence were counted as negative for transgene-dependent tumor formation. Zebrafish with GFP fluorescence were processed as described earlier and tumor formation was confirmed by visual review by a pathologist of hematoxylin and eosin stained slides. Fish tumors were classified by location (Figure 1C): tumors anterior to the gill were considered head tumors, and tumors posterior to the largest part of the ventral fin were considered tail tumors. Tumors between these two boundaries were divided into back tumors and ventral tumors at the midcoronal plane. Tumors spanning more than one area were classified by the area in which the tumor predominately presented.

Generation of genetically modified C2C12 cell lines—Knockout of Arf6 in the C2C12 was performed as follows: guides were designed targeting Arf6 in Benchling and the most efficient predicted gRNA was utilized with the sequence TCTTCGGGAACAAGGAAATG (Synthego). The 4D-Nucleofector Kit (Lonza #1032) was used to electroporate HiFi Cas9 nuclease (IDT #1081061), Arf6 gRNA, and the Electroporation Enhancer (IDT #1075916) into 200,000 C2C12 cells using the Lonza 4D-Nucleofector system. Control wild-type cells were electroporated without the Cas9/RNP complex. To confirm the gene editing efficiency, PCR was performed around the target sequence, Sanger sequenced, and Inference of CRISPR Edits (ICE) analysis (Synthego) suggested multiple knockout alleles. These populations (both control and Arf6 knockout cells) were utilized for generating monoclonal cell populations using cloning by limiting dilution. From monoclonal populations, DNA was isolated using the DNeasy Blood and Tissue Kit (QIAGEN #69504), PCR and Sanger sequencing performed to identify a one base pair A insertion at the 40th nucleotide relative to the start codon, resulting in premature stop and a 109 amino acid Arf6 protein as compared to the full-length 176 amino acid protein. Arf6 protein knockout was confirmed by western blot. In Figure 6, wild-type C2C12 cells were transfected with pcDNA3.1 empty vector or pcDNA3.1 containing the human coding sequence of the VGLL2-NCOA2 fusion using the AMAXA cell line Nucleofection kit V (Lonza #VCA-1003) and were selected in growth media containing 1 mg/mL concentration of G418 (ThermoFisher #10131035). Cells expressed the VGLL2-NCOA2 fusion at 24 h post transfection and stably after passage in culture under G418 selection. In Figure 7, wild-type and Arf6 knockout cells were transfected using the GenJet In Vitro DNA Transfection Reagent for C2C12 Cells (SignaGen Laboratories # SL100489-C2C12) with either the pcDNA3.1 empty plasmid or the VGLL2-NCOA2 expressing pcDNA3.1 plasmid. Cells were selected for over two weeks in growth media (DMEM + 10% FBS) with 1 mg/mL Geneticin/G418 (Gibco #10131027).

Western blot—In Figure 6, C2C12 cell lysate was generated using 5 million cells in RIPA lysis buffer with protease/phosphatase inhibitors (Roche #11836145001 and #04906837001), and 20μg of protein was loaded on a denaturing SDS page gel for detection of the VGLL2-NCOA2 fusion. Western blot was performed using primary antibodies against

human NCOA2 at 1:5000 dilution (Abcam #ab10508) and anti Beta Actin at 1:10000 dilution (Sigma #A5316), and were visualized using monoclonal anti-Rabbit (1:3000) secondary, and chemiluminescence on film. In Figure 7, C2C12 cells were lysed in RIPA buffer and 1X protease/phosphatase inhibitors with brief use of a pestle (Sigma #R0278 and Thermo Scientific #78446), rocked at 4°C for 20 min, and centrifuged at 15000 g at 4°C for 15 min. Protein concentration was determined by Pierce BCA Protein Assay Kit (ThermoFisher Scientific, #23225). A pre-cast 4–15% gradient gel (Bio-Rad #4561084) was loaded with 10μL of Precision Plus Protein Dual Color Standard (Bio-Rad #1610374), 18μg of protein lysate, and ran in 1X Tris/Glycine/SDS buffer (Bio-Rad #1610772). The gel was transferred to an Immun-Blot PVDF membrane (Bio-Rad #1620177), blocked with 5% milk + 0.1% Tween in 1X Tris-Buffered Saline (TBS), and incubated with 1:1000 primary rabbit polyclonal anti-KAT13C/NCOA2 antibody (Abcam #ab10508) in Casein + 0.05% Tween overnight at 4°C (ThermoScientific #37528). Membranes were then washed and incubated with 1:10000 goat anti-rabbit IgG (Bio-Rad #1721019) secondary antibody in Casein + 0.05% Tween for 1 h at room temperature. For chemiluminescence visualization we used SuperSignal™ West Atto Ultimate Sensitivity Substrate (ThermoScientific # PI34577) and imaged on the LI-COR C-DiGit Blot Scanner. After imaging, the membrane was stripped for 15 min with the Thermo Scientific™ Restore™ Western Blot Stripping Buffer (ThermoScientific #PI21059), washed, and incubated with either 1:1000 alpha-Tubulin (DM1A) Mouse mAb (Cell Signaling #3873S) or 1:1000 Anti-ARF6 antibody (Abcam #ab226389) in Casein + 0.05% Tween at 4°C overnight. For Tubulin, a dilution of 1:10000 goat anti-mouse IgG secondary antibody was used (Bio-Rad #1706516) and for ARF6, a dilution of 1:10000 goat ant-rabbit IgG secondary antibody (Bio-Rad #1721019). Secondary antibodies were diluted in Casein + 0.05% Tween and the membrane was incubated for 1 h at room temperature. Imaging was performed as above. In Figure 5, adult zebrafish were euthanized, and skeletal muscle or tumors extracted and snap-frozen. To isolate protein, skeletal muscle and tumors were homogenized in the T-PER™ Tissue Protein Extraction Reagent (ThermoFisher Scientific #78510) and rotated at 4°C for 2 h. Protein concentrations were obtained by Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific #23225). Wild-type AB embryos at 24 h post fertilization were collected, dechorionated, deyolked with 55 mM NaCl, 1.8 mM KCl, 1.25 mM $NaHCO_3$, washed with 0.5X Danieau's buffer (29 mM NaCl, 0.35 mM KCl, 0.2 mM $MgSO_4 \cdot 4H_2O$, 0.3 mM $Ca(NO_3)_2 \cdot 4H_2O$, 2.5 mM HEPES), and snap-frozen.[72] Western blot was performed with 25μg of skeletal muscle protein, 25μg of tumor protein, and 12, 24 h post fertilization embryos. Samples were run in duplicate to quantify loading with QC Colloidal Coomassie (Bio-Rad cat#1610803) overnight. Membrane was incubated with 1:1000 anti-ARF6 (Abcam #ab226389) and visualized with 1:10000 anti-rabbit (Bio-Rad #1721019) on LI-COR C-DiGit Blot Scanner. Western blots were quantified by densitometry analysis in ImageJ.

ChIP-qPCR—ChIP-qPCR protocol is adapted from Sunkel et al., 2021.[73] Transfected C2C12 cells were cultured in T175 flasks, detached with TrypLE Express, and washed with dPBS. Cells were re-suspended in dPBS and fixed in 1% formaldehyde solution for 10 min at room temperature. Fixation was quenched by adding glycine to a final concentration of 125mM and incubating on ice for 5 min. Cells were aliquoted, snap-frozen with dry ice, and stored at −80°C until sonication. Six million cells per sample were thawed and suspended

in 800μL of TE buffer, pH 8.0 with protease inhibitors. Cells were placed into a cooling block in a polystyrene micro centrifuge tube and sonicated with an Active Motif EpiShear probe sonicator at 30% amplitude with a 30 s on/off cycle for 25 min. 5 μL of sonicated DNA was taken from each sample as an input. Crosslinks were reversed by adding 20 μL TE, 1 μL 10% SDS, and 1 μL 18.5 mg/mL Proteinase K and incubating overnight at 65°C. Sonication and DNA fragmentation were evaluated using E-Gel 2% EX agarose gel system after purification with a Qiagen MinElute PCR Purification kit. Following fragmentation, 300μL of chromatin in TE buffer was taken and modified into a ChIP buffer with a final concentration of 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, and 200mM NaCl. Chromatin was incubated on ice for 5 min and centrifuged at 13,000 rpm at 4°C for 10 min. Supernatant was moved to a new micro centrifuge tube and incubated with 2 μg of H3K27Ac antibody (Active Motif # 39133) for 2 h with overhead rotation at 4°C. 40μL of Protein G Dynabeads, suspended in the previous mentioned ChIP buffer, were added to each sample and incubation continued overnight. Samples were wash twice in TE with 0.1% SDS, 0.1% sodium deoxycholate, and 1% Triton X-100, twice in TE with 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, and 200mM NaCl, and twice in TE with 250mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate. Samples were suspended in 100 μL TE buffer and reverse cross-linked overnight at 65°C with the addition of 2.5 μL 10% SDS and 5 μL 18.5 mg/mL Proteinase K. DNA was purified with a Qiagen MinElute PCR Purification kit. ChIP was analyzed by qPCR using iTaq Universal SYBR Green Supermix on a BioRad CFX384™ Real-Time Thermal Cycler. Equal volumes of ChIP and input DNA were diluted and run in triplicate using the primers in Table S5. Fold enrichment was calculated compared to a negative locus control (Active Motif #71012).

Mouse allograft experiments and tumor processing—C2C12-pcDNA3.1 control and C2C12-pcDNA3.1-VGLL2NCOA2 cell lines were utilized in allograft experiments. Two million cells were resuspended in 50μL of sterile 1X PBS and injected intramuscularly into the leg of SCID (Charles River Laboratories CB17/Icr-*Prkdc*^{scid}/IcrIcoCrl) or Swiss Nude (Charles River Laboratories Crl:NU(Ico)-*Foxn1*^{nu}) mice under anesthesia. All injected mice were female and, except for one Nude mouse at 136 days of age, were 46 days old at the time of injection. Tumors were detected eleven days after injection, after which the height and width of the injected leg were measured using calipers to calculate the volume until 60 days post injection for C2C12 controls, and until 26 or 28 days post injection for C2C12-VGLL2NCOA2 injected mice (initial leg volume was 50–100mm^3). At these later timepoints mice were euthanized due to tumor burden (C2C12-VGLL2NCOA2) or due to the end of the experiment (C2C12-pcDNA3.1 control). Tumor or control tissue was snap frozen in liquid nitrogen, and RNA isolated with Trizol for RNA-seq analysis. For immunohistochemistry, tumors were fixed in AFA (70% pure ethanol, 20% 4% formaldehyde solution, 10% glacial acetic acid) and mounted in paraffin blocks for microtome sectioning.

***In situ* hybridization and immunofluorescence—**Fluorescent *in situ* hybridization of *myog (*ACD Bio #895021), *desma (*ACD Bio #1006941-C1), and *myod1 (*ACD Bio #481231) was performed using RNAscope Multiplex Fluorescent Reagent Kit v2. on 8 μm transverse sections of VGLL2-NCOA2 tumor and wild-type zebrafish (ACD Bio #323100).

RNAscope hybridizations were done according to manufacturer's instructions for formalin-fixed paraffin-embedded tissue with the following changes. Protease Plus was applied for 20 min instead of 30 min and Opal 520 (Fisher #NC1601877) was used with TSA dilution solution. After applying HRP blocker, immunofluorescence was performed. Slides were blocked at room temperature for 1 h in block solution (3% donkey serum, 0.1% Triton X-100, in 1X PBS). Slides were then incubated overnight at 4°C in antibody solution (1% donkey serum, 0.05% Tween 20, in 1XPBS) with anti-PCNA antibody at 1:500 dilution (Sigma Aldrich #MABE288). Slides were washed 3x for 10 min each in 1XPBS, then incubated for 2 h at room temperature in antibody solution with Cy3-conjugated donkey anti-mouse antibody at 1:100 (Jackson ImmunoResearch #715165150). Slides were washed 3x for 10 min each, then mounted with Invitrogen Prolong Gold Antifade with DAPI (Fisher #P36935). RNAscope and immunofluorescence samples were imaged using a Nikon Eclipse Ti2-E inverted microscope. Higher magnification images were taken using a 40X Plan Apo λ objective. 16-bit mode was used for DAPI, GFP, and Cy-3 channels. Images were processed with the following settings using Nikon GA3 analysis: Detect Peaks with count 10 was used for DAPI, and Rolling Ball with 10μm radius was used for GFP and Cy3.

Immunohistochemistry—For zebrafish sections, slides were baked in a 60°C oven for 1 h, deparaffinized with xylene/Histo-Clear for two incubations of 10 min, rehydrated in two 3 min washes in 100% Ethanol, 95% Ethanol and then diH_2O. Antigen retrieval was performed in Trilogy reagent for 15 min in a pressure cooker. The slides were incubated with 3% peroxidase H_2O_2 for 30 min, washed with water for 1 min, blocked with 1% BSA/1x PBST for 1 h. Primary antibody incubation was done at 1:100 dilution at 4°C overnight. Primary antibodies used were anti-NCOA2 (Novus BIO #NB100–1756) and anti-ARF6 (Abcam #ab77581). Slides were washed with PBST 4 times for 5 min each, and HRP secondary conjugate (BioRad #1721019) was applied at 1:300 dilution for 1 h at room temperature. Washes were performed as above, and DAB solution (Sigma #D5905) with 0.03% H_2O_2 applied for 5 min. Slides were counterstained with hematoxylin for 5 min and dehydrated in 95% Ethanol, 100% Ethanol, and Xylene/Histo-Clear each for two incubations of 1 min each. Slides were imaged on a Keyence BZ-X710 microscope or a Leica Aperio GT 450. For mouse allograft Desmin, Myogenin, and Ki67 stains, immunohistochemistry was performed using the DAKO Autostainer Link 48 (Agilent Technologies) with the following antibody conditions for each protein target: Desmin, D33 clone (Dako #M0760), pH6 retrieval, 1:200 primary dilution and 30 min incubation; Myogenin, F5D clone (Dako #GA067), pH9 retrieval, 1:100 primary antibody dilution and 60 min incubation. The Ki67 stain was performed using the Bond RX immunostainer and Bond Polymer Refine Detection Kit (Leica Microsystems). The antibody used for Ki67 was SP6 clone (Abcam #16667) with pH9 retrieval, 1:1000 primary antibody dilution and 30 min incubation. Slides were digitized using the Ultra Fast 1.6 RA scanner (Philips, Best, Netherlands). For the mouse allograft Arf6 stain, slides were baked in a 60°C oven for 1 h, deparaffinized with xylene/Histo-Clear for two incubations of 10 min, rehydrated in two 3 min washes in 100% Ethanol, 95% Ethanol and then diH_2O. Antigen retrieval was performed in 10 mM sodium citrate buffer pH 6.0 for 15 min in a pressure cooker. The slides were washed twice with PBS for 3 min then blocked with 3% BSA/1X PBST with 10% goat serum for 1 h. Primary antibody incubation was done at 1:100 dilution at 4°C

overnight. The primary antibody used was anti-ARF6 (Novus BIO #NBP3–15391). Slides were washed with PBST 4 times for 5 min each and incubated with Thermo Scientific Peroxidase Suppressor (Fisher #PI35000) for 30 min and rinsed twice with water for 3 min. ImmPRESS HRP (Vector Labs #MP-7451) was then applied for 30 min at room temperature. PBST washes were performed as above, and DAB solution (Sigma #D5905) with 0.03% H_2O_2 was applied for 3 min. Slides were counterstained with hematoxylin for 5 min and dehydrated in 95% Ethanol, 100% Ethanol, and Xylene/Histo-Clear each for two incubations of 1 min each and were scanned using an Aperio GT 450 (Leica).

Cellular proliferation of C2C12 *in vitro*—Arf6 KO C2C12 cells and WT C2C12 cells, either with or without expression of the VGLL2-NCOA2 fusion, were seeded at a density of 500,000 or 200,000 cells in T175 flasks. Cells were maintained in growth media, supplemented with 1 mg/mL G418 (Fisher). After 2 or 3 days, respectively, cells were incubated with TrypLE Express Enzyme (Fisher #12604013) and then live cells were counted on a Countess Automated Cell Counter (ThermoFisher) with a trypan blue stain.

Myogenic differentiation of C2C12 *in vitro*—C2C12 cells were differentiated by plating 150,000 cells per well in growth media + G418 on 6-well plates coated with 0.1% porcine gelatin/1X PBS. After 48 h, cells were washed in 1X PBS, and differentiation media + G418 was added. Cells were then fused for 6 days in differentiation media (1X DMEM + 10% horse serum + 10μg/mL insulin + 1 mg/mL G418), with fresh media being added every other day. For fusion timepoints, cells were collected and pelleted after incubation in TrypLE Express Enzyme (Fisher #12604013), and then were snap frozen for total RNA isolation.

Soft agar colony formation—Colony formation protocol was adapted from Borowicz et al., 2014.[74] Briefly, in a 6-well plate a bottom layer of 1% noble agar in growth media + 1 mg/mL G418 was plated and allowed to solidify. This was followed by a top layer of 0.6% noble agar + growth media (DMEM + 10% FBS + 1 mg/mL Geneticin/G418) containing either 5000 or 10000 cells per well. The top layer was allowed to solidify, and 100μL of growth media + G418 was added to the top of each well. To prevent drying out, 100μL of growth media + G418 was added every 2–3 days. Colonies were counted manually using a Leica DMIL LED inverted microscope two weeks after plating.

RNA isolation, RT-PCR and qRT- PCR—RNA was isolated from zebrafish tumors, normal tissue, or embryos using the QIAGEN RNeasy micro or mini-kit depending on the tumor size. For skeletal muscle, an additional proteinase K digestion step was included. 500 ng of RNA isolated from zebrafish tumors, normal tissue or n = 30 embryos at 20 h post fertilization was used as input to the RT^2 HT First Strand Kit (Qiagen #330411) to reverse transcribe cDNA. For C2C12 cells, total RNA was isolated from using the RNeasy Mini or Micro Plus Kit (QIAGEN #74104 or #74034). cDNA was reverse transcribed from 1 μg of total RNA with the RT^2 HT First Strand Kit (QIAGEN #330411). Standard PCRs were run using Taq Polymerase (NEB #M0273). qRT-PCR was performed on the BioRad CFX384 using standard cycling conditions and the 2X BioRad Master Mix. Primer sequences are provided in Table S5. Data was analyzed in CFX Maestro Software and are plotted as the

values relative to zero as normalized to two input controls, *rplp13a* and *gapdh* for zebrafish samples, or *Rpl27* and *Gapdh* for mouse samples.

RNA sequencing and analysis—Approximately 2 mg of RNA was used for library preparation of each zebrafish sample. RNA-seq for D738, D739, D742, D777, D799, D800, D801, D807, D808, D813, Muscle 2 and Muscle 3 was done at the Institut Curie. Total RNA was isolated from Nude C2C12-pcDNA3.1 control or C2C12-*VGLL2NCOA2* allografts using Trizol, and approximately 1 mg was used for RNA-seq at the Institut Curie. D809, D811, D812, D814, D815, D874, and additional five mature skeletal muscle samples (D948, D949, D950, D951, D952) were sent to DNALinks for mRNA sequencing with poly-A RNA enrichment (2x75bp run, 70M reads of data generation). All tumors were generated in an AB/TL wild-type background, except for D738 and D742 which are in the *tp53*^{M214K} homozygous mutant background.[52] Additional sample details are in Table S6. Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used for quality and adapter trimming. The qualities of RNA-sequencing libraries were estimated by mapping the reads onto zebrafish reference dataset (Genome Reference Consortium Zebrafish Build 11, GRCz11) using Bowtie (v2.3.4.3)[58]. STAR (v2.7.2b)[58] was employed to align the reads onto the zebrafish genome GRCz11, mouse samples to Mus musculus GRCm38 genome, or human samples to the Homo sapiens GRCh38p4 genome. SAMtools (v1.9)[60] was employed to sort the alignments and HTSeq Python package[60] was employed to count reads per gene. edgeR R Bioconductor package[61–63] was used to normalize read counts and identify differentially expressed (DE) genes. The enrichment of DE genes to pathways and GOs were calculated by Fisher's exact test in R statistical package.[75] Differentially expressed genes were determined using cut-offs of fold changes >1.5 and an FDR of <0.05.

Comparative oncology RNA-seq analysis—Agreement of Differential Expression (AGDEX)[27] analysis was performed using RNA-seq data generated in this study and from previously published sources. The following AGDEX comparisons were made: zebrafish VGLL2-NCOA2 tumors vs mature skeletal muscle and human VGLL2-NCOA2 tumors vs mature skeletal muscle; zebrafish VGLL2-NCOA2 vs CIC-DUX4 tumors and human VGLL2-NCOA2 vs CIC-DUX4 tumors; zebrafish KRAS tumors vs controls and zebrafish VGLL2-NCOA2 tumors vs mature skeletal muscle; zebrafish KRAS tumors vs controls and human VGLL2-NCOA2 tumors vs mature skeletal muscle; mouse allograft VGLL2-NCOA2 tumors vs mature skeletal muscle and human VGLL2-NCOA2 tumors vs mature skeletal muscle. RNA-seq data from human VGLL2-NCOA2 and CIC-DUX4 tumors is previously published in Watson et al., 2018.[4] RNA-seq data from human mature skeletal muscle is from GTEx version 8. RNA-seq data from mouse mature skeletal muscle is previously published and can be accessed through NCBI BioProjects PRJNA625451,[28] PRJNA608179,[29] PRJNA819493, PRJNA813153.[30] RNA-seq data from KRAS driven rhabdomyosarcoma zebrafish tumors and controls is previously published in Ignatius et al., 2018.[48] All other data was generated through this study and can be accessed on GEO under the accession number GSE206039. Gene expression values for both RNA-seq series were assessed using Kallisto 0.46.1[67] using indexes of the Homo sapiens GRCh38p4 genome, Mus musculus GRCm38 genome, or Danio rerio GRCz11 genome. Mapping of the genes between the species was done using BioMart tool from the Ensembl website (https://

www.ensembl.org/biomart/martview). AGDEX[27] analysis was performed with a minimum number of permutations set to 5000. To further assess the statistical significance between species' correlation of the differential expression results, we performed a permutation analysis. For each comparison, we counted the number of genes that changed in the same direction in both datasets (logFC>0 in both or logFC<0 in both). We then used kernel density functions to generate randomized data matching the distribution of the original data and counted how many genes changed in the same direction between the two simulated datasets. This was repeated 100,000 times and we counted the number of iterations where an equal or greater number of genes changed in the same direction. This number was divided by the number of iterations to calculate a p value for each comparison. Where no iteration had as many correlated genes as the real data, the p value was noted as $p < 1x10^{-5}$. UMAP projections of RNA-seq data were generated in R using normalized RNA-seq read counts.

Human sarcoma and skeletal muscle RNA-seq analysis—RNA-seq data from VGLL2-NCOA2 tumors is from Watson et al., 2018.[4] RNA-seq data for n = 264 adult sarcomas is from TCGA and represents dedifferentiated liposarcoma, leiomyosarcoma, undifferentiated pleomorphic sarcoma, myxofibrosarcoma, malignant peripheral nerve sheath tumor, and synovial sarcoma.[49] RNA-seq data from n = 96 Ewing sarcoma tumors is from Crompton et al., 2014,[50] RNA-seq data from n = 87 Osteosarcoma (phs000468) and n = 13 Clear Cell Sarcoma of the Kidney (phs000466) are based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (https://ocg.cancer.gov/programs/target) initiative, phs000468 and phs000466. The data used for this analysis are available at https://portal.gdc.cancer.gov/projects. RNA-seq data from n = 42 fusion-positive and fusion-negative rhabdomyosarcoma is from Chen et al., 2013.[51] RNA-seq data from n = 396 adult skeletal muscle samples are from the Genotype-Tissue Expression Project (GTEx). The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript in Figure 7 were obtained from the GTEx Portal in March 2018, and in Figures 2, 4, S3 and S6 were obtained from the GTEx portal in April 2022. The same computational analysis steps based on hg19 human reference genome data, as detailed in the RNA-seq and analysis section, were applied to process these tumor and muscle datasets side by side to minimize computational batch effects. In addition, we also applied ComBat[65] to remove batch effects. HTSeq Python package[60] was employed to count reads per gene. edgeR R Bioconductor package[61–63] was used to normalize read counts and calculate FPKM values.

PCA of zebrafish tumors and developmental stages—Zebrafish developmental RNA-seq data were from White et al., 2017.[26] Zebrafish tumor and skeletal muscle sample RNA-seq data were generated from this study. To minimize the computational batch effect, FASTQ files from White et al., 2017[26] and this study were processed side by side with the same pipeline, as detailed in the RNA sequencing and analysis section. We then performed the Principal Component Analysis based on princomp package in R.

Gene ontology analysis—In Figure S4, genes with statistically significant expression changes were analyzed using DAVID (https://david.ncifcrf.gov/)[66,67] to identify enriched

Gene Ontology (GO) terms (http://www.geneontology.org/). The direct biological process (top) and molecular function (bottom) ontology terms are listed for genes differentially regulated in zebrafish VGLL2-NCOA2 tumors versus skeletal muscle (FDR<0.1). In Figures 2, S3 and S6, Gene Set Enrichment Analysis was performed using co-differentially expressed genes in zebrafish and human VGLL2-NCOA2 tumors or mouse and human VGLL2-NCOA2 tumors as compared to mature skeletal muscle[76,77] using the clusterProfiler R package.[68,69] We searched the Biological Processes (BP), Molecular Functions (MF) and Cellular Component (CC) subontologies.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of survival curves, tumor incidence and onset curves, Fisher's exact test, and all experimental statistical tests was performed in Prism 8.1.1 (LaJolla, CA). Image analysis was performed in ImageJ. All other calculations were performed in R. Sample sizes and the statistical tests performed are provided in figures or figure legends.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank the UT Southwestern Histopathology core for excellent services and for their expertise, and the Nationwide Children's Hospital (NCH) Animal Resources Core for their exceptional zebrafish husbandry. Additional support was provided by the Nationwide Children's Hospital Microscopy Lab in the Center for Gene Therapy for imaging, the NCH Morphology Core for tissue processing, the NCH CRISPR/Gene Editing Core for generating cell line knockouts, and the NCH High-Performance Computing group for assistance maintaining and using the NCH cluster. G.C.K. is supported by an Alex's Lemonade Stand Foundation A Award, a V Foundation for Cancer Research V Scholar Grant, and an R01CA272872 grant through the National Cancer Institute of the National Institutes of Health. J.F.A. was supported by grant RP120685-P3 from the Cancer Prevention and Research Institute of Texas (CPRIT) and by Curing Kids Cancer. L.X. is supported by Rally Foundation, Cancer Center Support Grant P30 CA142543 and RP180805 from CPRIT. F.T. is supported by the Institut National de la Recherche Médiacale (INSERM). D.R. is supported by grant RP180319 from the Cancer Prevention and Research Institute of Texas (CPRIT) and by the John Lawrence and Patsy Louise Goforth Chair in Pathology. S.W. was supported by a grant from the Fondation Nuovo-Soldati pour la recherche en cancérologie. M.K. is supported by a T32CA269052 Training Program in Basic and Translational Pediatric Oncology Research postdoctoral fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Mitelman F, B J, and Mertens F(2021). Mitelman Database of Chromo-some Aberrations and Gene Fusions in Cancer https://mitelmandatabase.isb-cgc.org.
2. Mitelman F, Johansson B, and Mertens F (2007). The impact of translocations and gene fusions on cancer causation. Nat. Rev. Cancer 7, 233–245. 10.1038/nrc2091. [PubMed: 17361217]
3. Mertens F, Antonescu CR, and Mitelman F (2016). Gene fusions in soft tissue tumors: recurrent and overlapping pathogenetic themes. Genes Chromosomes Cancer 55, 291–310. 10.1002/gcc.22335. [PubMed: 26684580]
4. Watson S, Perrin V, Guillemot D, Reynaud S, Coindre JM, Karanian M, Guinebretiére JM, Freneaux P, Le Loarer F, Bouvet M, et al. (2018). Transcriptomic definition of molecular subgroups of small round cell sarcomas. J. Pathol 245, 29–40. 10.1002/path.5053. [PubMed: 29431183]
5. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, Johann PD, Balasubramanian GP, Segura-Wang M, Brabetz S, et al. (2018). The landscape of genomic

alterations across childhood cancers. Nature 555, 321–327. 10.1038/nature25480. [PubMed: 29489754]

6. Tirode F, Surdez D, Ma X, Parker M, Le Deley MC, Bahrami A, Zhang Z, Lapouble E, Grossetête-Lalami S, Rusch M, et al. (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov 4, 1342–1353. 10.1158/2159-8290.CD-14-0622. [PubMed: 25223734]
7. Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, Ambrogio L, Auclair D, Wang J, Song YK, et al. (2014). Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov 4, 216–231. 10.1158/2159-8290.CD-13-0639. [PubMed: 24436047]
8. Alaggio R, Zhang L, Sung YS, Huang SC, Chen CL, Bisogno G, Zin A, Agaram NP, LaQuaglia MP, Wexler LH, and Antonescu CR (2016). A molecular study of pediatric spindle and sclerosing rhabdomyosarcoma: identification of novel and recurrent VGLL2-related fusions in infantile cases. Am. J. Surg. Pathol 40, 224–235. 10.1097/PAS.0000000000000538. [PubMed: 26501226]
9. Günther S, Mielcarek M, Krüger M, and Braun T (2004). VITO-1 is an essential cofactor of TEF1-dependent muscle-specific gene regulation. Nucleic Acids Res 32, 791–802. 10.1093/nar/gkh248. [PubMed: 14762206]
10. Maeda T, Chapman DL, and Stewart AFR (2002). Mammalian vestigial-like 2, a cofactor of TEF-1 and MEF2 transcription factors that promotes skeletal muscle differentiation. J. Biol. Chem 277, 48889–48898. 10.1074/jbc.M206858200. [PubMed: 12376544]
11. Pobbati AV, and Hong W (2013). Emerging roles of TEAD transcription factors and its coactivators in cancers. Cancer Biol. Ther 14, 390–398. 10.4161/cbt.23788. [PubMed: 23380592]
12. Johnson CW, Hernandez-Lagunas L, Feng W, Melvin VS, Williams T, and Artinger KB (2011). Vgll2a is required for neural crest cell survival during zebrafish craniofacial development. Dev. Biol 357, 269–281. 10.1016/j.ydbio.2011.06.034. [PubMed: 21741961]
13. Wang L, Motoi T, Khanin R, Olshen A, Mertens F, Bridge J, Dal Cin P, Antonescu CR, Singer S, Hameed M, et al. (2012). Identification of a novel, recurrent HEY1-NCOA2 fusion in mesenchymal chondrosarcoma based on a genome-wide screen of exon-level expression data. Genes Chromosomes Cancer 51, 127–139. 10.1002/gcc.20937. [PubMed: 22034177]
14. Sumegi J, Streblow R, Frayer RW, Dal Cin P, Rosenberg A, Meloni-Ehrig A, and Bridge JA (2010). Recurrent t(2;2) and t(2;8) translocations in rhabdomyosarcoma without the canonical PAX-FOXO1 fuse PAX3 to members of the nuclear receptor transcriptional coactivator family. Genes Chromosomes Cancer 49, 224–236. 10.1002/gcc.20731. [PubMed: 19953635]
15. Argani P, Reuter VE, Kapur P, Brown JE, Sung YS, Zhang L, Williamson R, Francis G, Sommerville S, Swanson D, et al. (2018). Novel MEIS1-NCOA2 gene fusions define a distinct primitive spindle cell sarcoma of the kidney. Am. J. Surg. Pathol 42, 1562–1570. 10.1097/PAS.0000000000001140. [PubMed: 30179902]
16. Kao YC, Bennett JA, Suurmeijer AJH, Dickson BC, Swanson D, Wanjari P, Zhang L, Lee JC, and Antonescu CR (2021). Recurrent MEIS1-NCOA2/1 fusions in a subset of low-grade spindle cell sarcomas frequently involving the genitourinary and gynecologic tracts. Mod. Pathol 34, 1203–1212. 10.1038/s41379-021-00744-7. [PubMed: 33574497]
17. Agaram NP, Zhang L, Sung YS, Cavalcanti MS, Torrence D, Wexler L, Francis G, Sommerville S, Swanson D, Dickson BC, et al. (2019). Expanding the spectrum of intraosseous rhabdomyosarcoma: correlation between 2 distinct gene fusions and phenotype. Am. J. Surg. Pathol 43, 695–702. 10.1097/PAS.0000000000001227. [PubMed: 30720533]
18. Mosquera JM, Sboner A, Zhang L, Kitabayashi N, Chen CL, Sung YS, Wexler LH, LaQuaglia MP, Edelman M, Sreekantaiah C, et al. (2013). Recurrent NCOA2 gene rearrangements in congenital/infantile spindle cell rhabdomyosarcoma. Genes Chromosomes Cancer 52, 538–550. 10.1002/gcc.22050. [PubMed: 23463663]
19. Whittle SB, Hicks MJ, Roy A, Vasudevan SA, Reddy K, and Venkatramani R (2019). Congenital spindle cell rhabdomyosarcoma. Pediatr. Blood Cancer 66, e27935. 10.1002/pbc.27935. [PubMed: 31339226]
20. Butel T, Karanian M, Pierron G, Orbach D, Ranchere D, Cozic N, Galmiche L, Coulomb A, Corradini N, Lacour B, et al. (2020). Integrative clinical and biopathology analyses to

understand the clinical heterogeneity of infantile rhabdomyosarcoma: a report from the French MMT committee. Cancer Med 9, 2698–2709. 10.1002/cam4.2713. [PubMed: 32087612]

21. Anderson WJ, and Doyle LA (2021). Updates from the 2020 world Health organization classification of soft tissue and bone tumours. Histopathology 78, 644–657. 10.1111/his.14265. [PubMed: 33438273]
22. . Cyrta J, Gauthier A, Karanian M, Vieira AF, Cardoen L, Jehanno N, Bouvet M, Bouvier C, Komuta M, Le Loarer F, et al. (2021). Infantile rhabdomyosarcomas with VGLL2 rearrangement are not always an indolent disease: a study of 4 aggressive cases with clinical, pathologic, molecular, and radiologic findings. Am. J. Surg. Pathol. 45, 854–867, Publish Ahead of Print. 10.1097/pas.0000000000001702. [PubMed: 33949344]
23. Kendall GC, Watson S, Xu L, LaVigne CA, Murchison W, Rakheja D, Skapek SX, Tirode F, Delattre O, and Amatruda JF (2018). PAX3-FOXO1 transgenic zebrafish models identify HES3 as a mediator of rhabdomyosarcoma tumorigenesis. Elife 7, e33800. 10.7554/eLife.33800. [PubMed: 29869612]
24. Watson S, Kendall GC, Rakheja D, McFaul ME, Draper BW, Tirode F, Delattre O, and Amatruda JF (2019). CIC-DUX4 expression drives the development of small round cell sarcoma in transgenic zebrafish: a new model revealing a role for ETV4 in CIC-mediated sarcomagenesis. Preprint at bioRxiv, 517722. 10.1101/517722.
25. Urasaki A, Morvan G, and Kawakami K (2006). Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. Genetics 174, 639–649. 10.1534/genetics.106.060244. [PubMed: 16959904]
26. White RJ, Collins JE, Sealy IM, Wali N, Dooley CM, Digby Z, Stemple DL, Murphy DN, Billis K, Hourlier T, et al. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. Elife 6, e30860. 10.7554/eLife.30860. [PubMed: 29144233]
27. Pounds S, Gao CL, Johnson RA, Wright KD, Poppleton H, Finkelstein D, Leary SES, and Gilbertson RJ (2011). A procedure to statistically evaluate agreement of differential expression for cross-species genomics. Bioinformatics 27, 2098–2103. 10.1093/bioinformatics/btr362. [PubMed: 21697127]
28. Tanner MK, Tang Z, and Thornton CA (2021). Targeted splice sequencing reveals RNA toxicity and therapeutic response in myotonic dystrophy. Nucleic Acids Res. 49, 2240–2254. 10.1093/nar/gkab022. [PubMed: 33503262]
29. Chakraborty N, Waning DL, Gautam A, Hoke A, Sowe B, Youssef D, Butler S, Savaglio M, Childress PJ, Kumar R, et al. (2020). Gene-metabolite Network linked to inhibited bioenergetics in association with spaceflight-induced loss of male mouse quadriceps muscle. J. Bone Miner. Res 35, 2049–2057. 10.1002/jbmr.4102. [PubMed: 32511780]
30. Mao F, Wang E, Xu J, Lu J, Yan G, Fu L, Jiao Y, Wu L, Liu T, and Li Y (2022). Transcriptome analysis of multiple metabolic tissues in high-salt diet-fed mice. Front. Endocrinol 13, 887843. 10.3389/fendo.2022.887843.
31. Donaldson JG (2003). Multiple roles for Arf6: sorting, structuring, and signaling at the plasma membrane. J. Biol. Chem 278, 41573–41576. 10.1074/jbc.R300026200. [PubMed: 12912991]
32. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498–503. 10.1038/nature12111. [PubMed: 23594743]
33. McConnell AM, Noonan HR, and Zon LI (2021). Reeling in the zebrafish cancer models. Annu. Rev. Cancer Biol 5, 331–350. 10.1146/annurev-cancerbio-051320-014135.
34. Amatruda JF (2021). Modeling the developmental origins of pediatric cancer to improve patient outcomes. Dis. Model. Mech 14, dmm048930. 10.1242/dmm.048930. [PubMed: 33619212]
35. Mielcarek M, Günther S, Krüger M, and Braun T (2002). VITO-1, a novel vestigial related protein is predominantly expressed in the skeletal muscle lineage. Mech. Dev 119, S269–S274. 10.1016/s0925-4773(03)00127-8. [PubMed: 14516696]

36. Honda M, Hidaka K, Fukada SI, Sugawa R, Shirai M, Ikawa M, and Morisaki T (2017). Vestigial-like 2 contributes to normal muscle fiber type distribution in mice. Sci. Rep 7, 7168. 10.1038/s41598-017-07149-0. [PubMed: 28769032]
37. Honda M, Tsuchimochi H, Hitachi K, and Ohno S (2019). Transcriptional cofactor Vgll2 is required for functional adaptations of skeletal muscle induced by chronic overload. J. Cell. Physiol 234, 15809–15824. 10.1002/jcp.28239. [PubMed: 30724341]
38. Hashimoto S, Onodera Y, Hashimoto A, Tanaka M, Hamaguchi M, Yamada A, and Sabe H (2004). Requirement for Arf6 in breast cancer invasive activities. Proc. Natl. Acad. Sci. USA 101, 6647–6652. 10.1073/pnas.0401753101. [PubMed: 15087504]
39. Stepicheva NA, Dumas M, Kobi P, Donaldson JG, and Song JL (2017). The small GTPase Arf6 regulates sea urchin morphogenesis. Differentiation 95, 31–43. 10.1016/j.diff.2017.01.003. [PubMed: 28188999]
40. Suzuki T, Kanai Y, Hara T, Sasaki J, Sasaki T, Kohara M, Maehama T, Taya C, Shitara H, Yonekawa H, et al. (2006). Crucial role of the small GTPase ARF6 in hepatic cord formation during liver development. Mol. Cell Biol 26, 6149–6156. 10.1128/MCB.00298-06. [PubMed: 16880525]
41. Chen EH, Pryce BA, Tzeng JA, Gonzalez GA, and Olson EN (2003). Control of myoblast fusion by a guanine nucleotide exchange factor, loner, and its effector ARF6. Cell 114, 751–762. 10.1016/s0092-8674(03)00720-7. [PubMed: 14505574]
42. Morishige M, Hashimoto S, Ogawa E, Toda Y, Kotani H, Hirose M, Wei S, Hashimoto A, Yamada A, Yano H, et al. (2008). GEP100 links epidermal growth factor receptor signalling to Arf6 activation to induce breast cancer invasion. Nat. Cell Biol 10, 85–92. 10.1038/ncb1672. [PubMed: 18084281]
43. Yoo JH, Shi DS, Grossmann AH, Sorensen LK, Tong Z, Mleynek TM, Rogers A, Zhu W, Richards JR, Winter JM, et al. (2016). ARF6 is an actionable node that orchestrates oncogenic GNAQ signaling in uveal melanoma. Cancer Cell 29, 889–904. 10.1016/j.ccell.2016.04.015. [PubMed: 27265506]
44. Yoo JH, Brady SW, Acosta-Alvarez L, Rogers A, Peng J, Sorensen LK, Wolff RK, Mleynek T, Shin D, Rich CP, et al. (2019). The small GTPase ARF6 activates PI3K in melanoma to induce a prometastatic state. Cancer Res. 79, 2892–2908. 10.1158/0008-5472.CAN-18-3026. [PubMed: 31048499]
45. Oka S, Uramoto H, Shimokawa H, Yamada S, and Tanaka F (2014). Epidermal growth factor receptor-GEP100-Arf6 axis affects the prognosis of lung adenocarcinoma. Oncology 86, 263–270. 10.1159/000360089. [PubMed: 24902879]
46. Hashimoto S, Mikami S, Sugino H, Yoshikawa A, Hashimoto A, Onodera Y, Furukawa S, Handa H, Oikawa T, Okada Y, et al. (2016). Lysophosphatidic acid activates Arf6 to promote the mesenchymal malignancy of renal cancer. Nat. Commun 7, 10656. 10.1038/ncomms10656. [PubMed: 26854204]
47. Yamauchi Y, Miura Y, and Kanaho Y (2017). Machineries regulating the activity of the small GTPase Arf6 in cancer cells are potential targets for developing innovative anti-cancer drugs. Adv. Biol. Regul 63, 115–121. 10.1016/j.jbior.2016.10.004. [PubMed: 27776975]
48. Ignatius MS, Hayes MN, Moore FE, Tang Q, Garcia SP, Blackburn PR, Baxi K, Wang L, Jin A, Ramakrishnan A, et al. (2018). tp53 deficiency causes a wide tumor spectrum and increases embryonal rhabdomyosarcoma metastasis in zebrafish. Elife 7, e37202. 10.7554/eLife.37202. [PubMed: 30192230]
49. . Cancer Genome Atlas Research Network Electronic address elizabethde-micco@sinaihealthsystemca; Cancer Genome Atlas Research Network (2017). Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell 171, 950–965.e28. 10.1016/j.cell.2017.10.014. [PubMed: 29100075]
50. Crompton BD, Stewart C, Taylor-Weiner A, Alexe G, Kurek KC, Calicchio ML, Kiezun A, Carter SL, Shukla SA, Mehta SS, et al. (2014). The genomic landscape of pediatric Ewing sarcoma. Cancer Discov 4, 1326–1341. 10.1158/2159-8290.CD-13-1037. [PubMed: 25186949]
51. Chen X, Stewart E, Shelat AA, Qu C, Bahrami A, Hatley M, Wu G, Bradley C, McEvoy J, Pappo A, et al. (2013). Targeting oxidative stress in embryonal rhabdomyosarcoma. Cancer Cell 24, 710–724. 10.1016/j.ccr.2013.11.002. [PubMed: 24332040]

52. Berghmans S, Murphey RD, Wienholds E, Neuberg D, Kutok JL, Fletcher CDM, Morris JP, Liu TX, Schulte-Merker S, Kanki JP, et al. (2005). tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors. Proc. Natl. Acad. Sci. USA 102, 407–412. 10.1073/pnas.0406252102. [PubMed: 15630097]
53. Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, Campbell DS, Parant JM, Yost HJ, Kanki JP, and Chien CB (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev. Dyn 236, 3088–3099. 10.1002/dvdy.21343. [PubMed: 17937395]
54. Mosimann C, Kaufman CK, Li P, Pugach EK, Tamplin OJ, and Zon LI (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. Development 138, 169–177. 10.1242/dev.059345. [PubMed: 21138979]
55. Berger J, and Currie PD (2013). 503unc, a small and muscle-specific zebrafish promoter. Genesis 51, 443–447. 10.1002/dvg.22385. [PubMed: 23444339]
56. Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, and Mishina M (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev. Cell 7, 133–144. 10.1016/j.devcel.2004.06.005. [PubMed: 15239961]
57. Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. 10.1038/nmeth.1923. [PubMed: 22388286]
58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. 10.1093/bioinformatics/bts635. [PubMed: 23104886]
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. 10.1093/bioinformatics/btp352. [PubMed: 19505943]
60. Anders S, Pyl PT, and Huber W (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. 10.1093/bioinformatics/btu638. [PubMed: 25260700]
61. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5, R80. 10.1186/gb-2004-5-10-r80. [PubMed: 15461798]
62. McCarthy DJ, Chen Y, and Smyth GK (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40, 4288–4297. 10.1093/nar/gks042. [PubMed: 22287627]
63. Robinson MD, McCarthy DJ, and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. 10.1093/bio-informatics/btp616. [PubMed: 19910308]
64. Bray NL, Pimentel H, Melsted P, and Pachter L (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol 34, 525–527. 10.1038/nbt.3519. [PubMed: 27043002]
65. Zhang Y, Parmigiani G, and Johnson WE (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom. Bioinform 2, lqaa078. 10.1093/nargab/lqaa078. [PubMed: 33015620]
66. Huang DW, Sherman BT, and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc 4, 44–57. 10.1038/nprot.2008.211. [PubMed: 19131956]
67. Huang DW, Sherman BT, and Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1–13. 10.1093/nar/gkn923. [PubMed: 19033363]
68. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2, 100141. 10.1016/j.xinn.2021.100141. [PubMed: 34557778]
69. Yu G, Wang LG, Han Y, and He QY (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287. 10.1089/omi.2011.0118. [PubMed: 22455463]
70. Kendall GC, and Amatruda JF (2016). Zebrafish as a model for the study of solid malignancies. Methods Mol. Biol 1451, 121–142. 10.1007/978-1-4939-3771-4_9.

71. Provost E, Rhee J, and Leach SD (2007). Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. Genesis 45, 625–629. 10.1002/dvg.20338. [PubMed: 17941043]

72. Manoli M, and Driever W (2012). Fluorescence-activated cell sorting (FACS) of fluorescently tagged cells from zebrafish larvae for RNA isolation. Cold Spring Harb. Protoc 2012, pdb.prot069633. 10.1101/pdb.prot069633. [PubMed: 22854565]

73. Sunkel BD, Wang M, LaHaye S, Kelly BJ, Fitch JR, Barr FG, White P, and Stanton BZ (2021). Evidence of pioneer factor activity of an oncogenic fusion transcription factor. iScience 24, 102867. 10.1016/j.isci.2021.102867. [PubMed: 34386729]

74. Borowicz S, Van Scoyk M, Avasarala S, Karuppusamy Rathinam MK, Tauler J, Bikkavilli RK, and Winn RA (2014). The soft agar colony formation assay. J. Vis. Exp, e51998. 10.3791/51998. [PubMed: 25408172]

75. Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45, D353–D361. 10.1093/nar/gkw1092. [PubMed: 27899662]

76. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. 10.1073/pnas.0506580102. [PubMed: 16199517]

77. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet 34, 267–273. 10.1038/ng1180. [PubMed: 12808457]

Highlights

- VGLL2-NCOA2 is transforming in transgenic zebrafish and mouse allograft models
- Tumors resemble immature skeletal muscle and express developmental genes
- Zebrafish, mouse, and human tumors overexpress a small GTPase, ARF6
- ARF6 and VGLL2-NCOA2 genetically cooperate for pro-tumorigenic activity

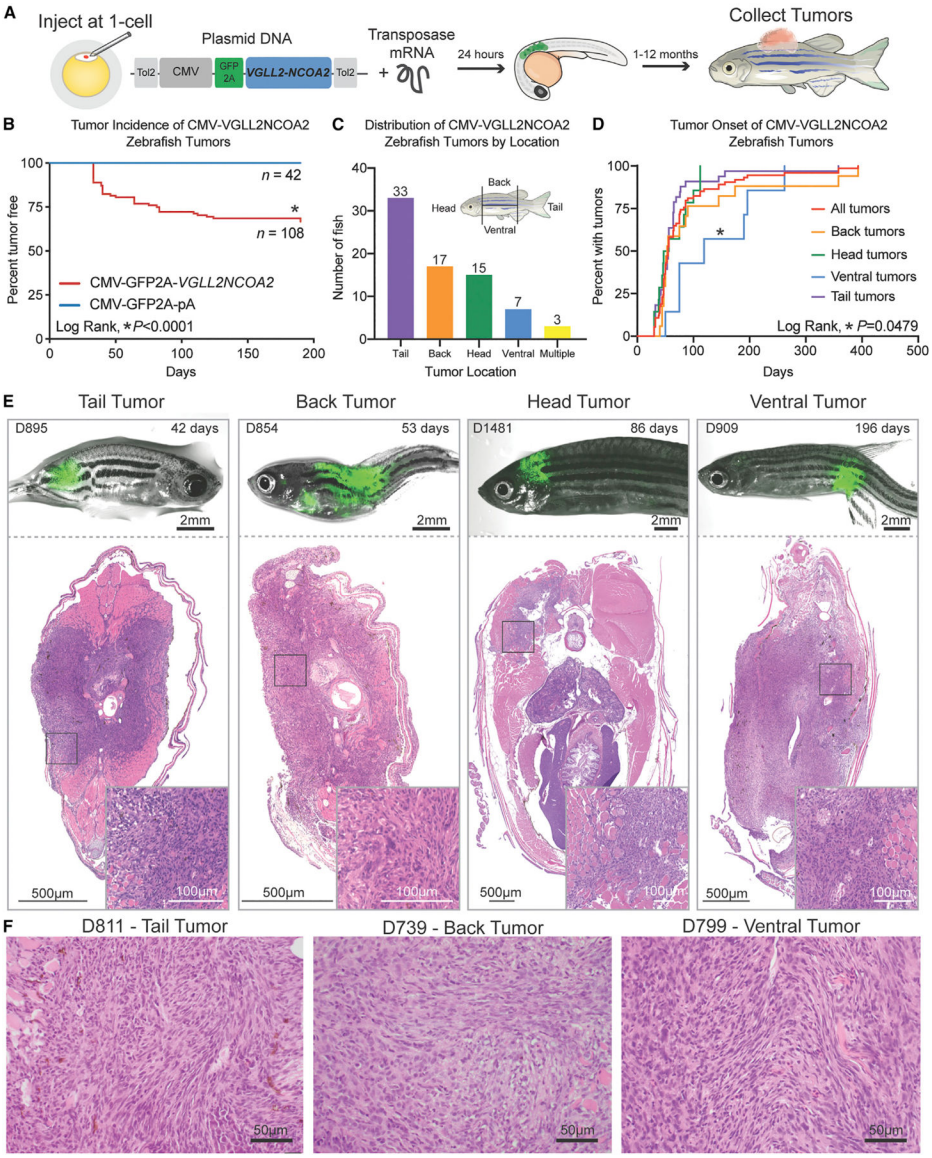


Figure 1. A zebrafish model of human VGLL2-NCOA2-driven tumorigenesis

(A) Zebrafish were injected at the one-cell stage with plasmid DNA carrying the human fusion gene *VGLL2-NCOA2* driven by the CMV promoter and tagged with GFP2A, and Tol2 transposase mRNA to stably integrate the construct. The GFP+ embryos were tracked from 1 to 12 months for tumor formation.

(B) Tumor incidence curve of one independent experiment showing n = 108 adult fish injected with CMV-GFP2A-VGLL2NOCA2 compared with n = 42 adult sibling fish injected with CMV-GFP2A-pA and log rank (Mantel-Cox) test performed with $p < 0.0001$. All fish that survived past 30 days of age are included. This experiment was repeated in three independent cohorts, with similar tumor incidence curves (data not shown).

(C) Distribution of CMV-VGLL2NCOA2 tumors classified by location as depicted on the inset fish schematic.

(D) Tumor onset curve for fish with a CMV-VGLL2NCOA2-driven tumor, stratified by location on the fish as defined in (C). Log rank (Mantel-Cox) analysis showed significant difference between the tumor onset of ventral tumors and all other tumors ($p = 0.0479$).
(E) Representative CMV-VGLL2NCOA2-injected fish with tumors classified by location. For each fish, brightfield and GFP fluorescent images are overlaid and shown above the H&E stain of a transverse section through the GFP+ area. The top left text is the unique fish identification number, and the top right is the age of the fish when the tumor was resected. Scale bars for GFP and brightfield overlay, 2 mm. Scale bars for H&E and H&E inset, 500 and 100 mm, respectively.
(F) High-magnification images (100× objective) of H&E stain of transverse sections through representative tumors from the tail, back, and ventral regions. Scale bars, 50 μm.

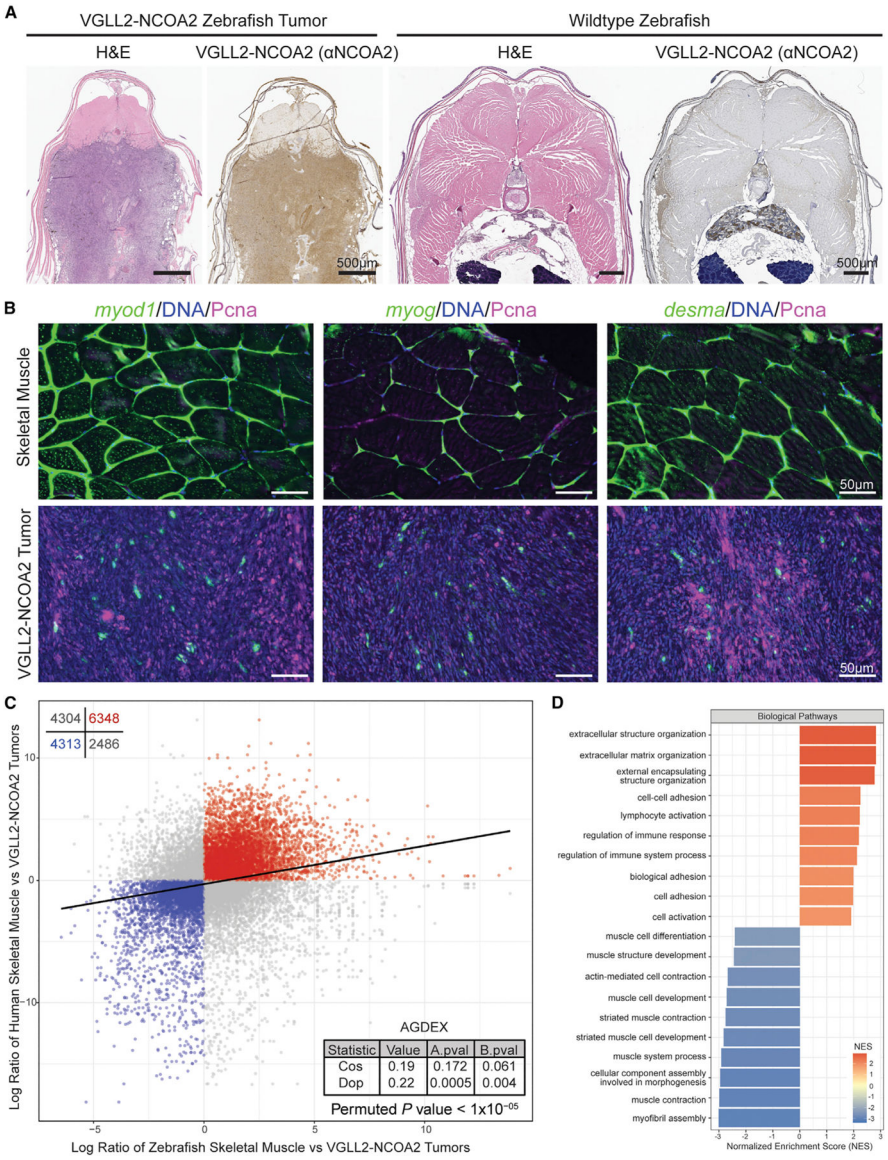


Figure 2. Zebrafish VGLL2-NCOA2 tumors recapitulate the human disease

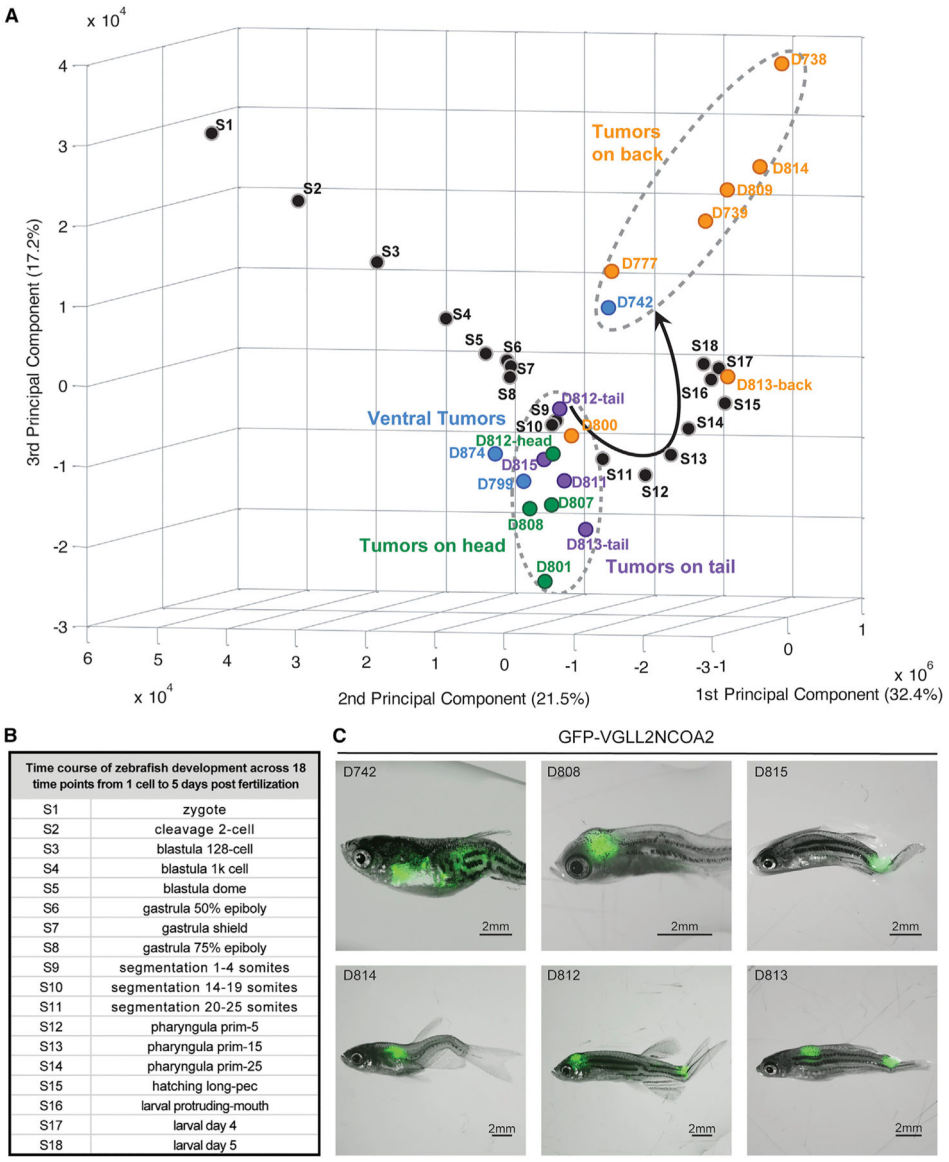
(A) Transverse sections through the GFP+ tumors were stained with H&E, and the subsequent step section stained with a human anti-NCOA2 antibody to detect the VGLL2-NCOA2 fusion. All n = 9 zebrafish tumors stained were positive. A transverse section was taken through a control wild-type zebrafish, and shown are the H&E and subsequent anti-NCOA2 antibody stain. Scale bars, 500 μm.

(B) Representative RNAscope of transverse sections from wild-type zebrafish (portraying normal skeletal muscle expression) or zebrafish with VGLL2-NCOA2 tumors showing expression of *myod1*, *myog*, or *desma* (green), Pcna (purple), or DNA (blue). Scale bars, 50 μm. Five wild-type or tumor samples were assessed in total, with results presented in Table S3.

(C) RNA-seq data of zebrafish VGLL2-NCOA2 tumors (n = 18) and mature skeletal muscle (n = 7) were compared with their human counterpart (VGLL2-NCOA2 tumors n = 5 from

Watson et al.[4] and mature skeletal muscle n = 803 from GTEx version 8) in an AGDEX analysis. Plotted in blue and red are genes that are shared as differentially expressed between the zebrafish and human VGLL2-NCOA2 tumors. The statistics of the AGDEX analysis are shown in the bottom right, as well as the p value generated by permutation analysis of randomly sampled data.

(D) Gene set enrichment analysis associated with the n = 6,348 genes upregulated or n = 4,313 downregulated in (C). Plotted is the normalized enrichment score for each sub-ontology.



Time course of zebrafish development across 18 time points from 1 cell to 5 days post fertilization	
S1	zygote
S2	cleavage 2-cell
S3	blastula 128-cell
S4	blastula 1k cell
S5	blastula dome
S6	gastrula 50% epiboly
S7	gastrula shield
S8	gastrula 75% epiboly
S9	segmentation 1-4 somites
S10	segmentation 14-19 somites
S11	segmentation 20-25 somites
S12	pharyngula prim-5
S13	pharyngula prim-15
S14	pharyngula prim-25
S15	hatching long-pec
S16	larval protruding-mouth
S17	larval day 4
S18	larval day 5

Figure 3. Gene expression profiles of VGLL2-NCOA2 zebrafish tumors cluster with developmental time points

(A) Principal-component analysis (PCA) of RNA-seq data of zebrafish development time points from White et al. [26] and VGLL2-NCOA2 zebrafish tumors from this study. All FASTQ files were processed by the same computational pipeline to minimize computational batch effects. Colors indicate location of the tumor on the fish: back tumors in orange, ventral tumors in blue, head tumors in green, and tail tumors in purple. In the PCA, principal component 1 (PC1) describes the most variance but is presented to appreciate differences in PC2, which best discriminates tumor cohorts.

(B) Embryonic stages labeled as per the Zebrafish Information Network (ZFIN).

(C) Brightfield and GFP overlaid images of a subset of the VGLL2-NCOA2 tumors used in (A). Scale bars, 2 mm.

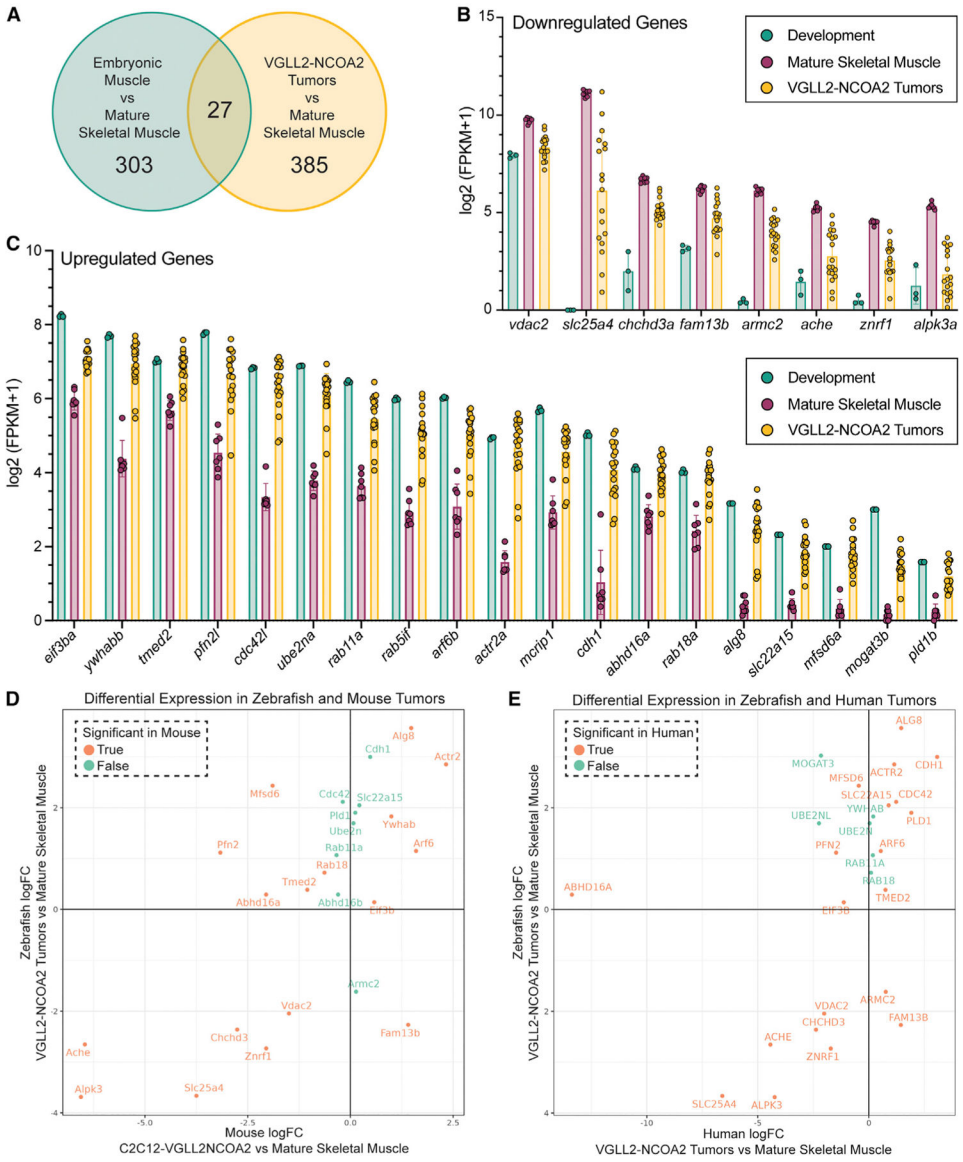


Figure 4. VGLL2-NCOA2 reactivates developmental genes in zebrafish, mouse allograft, and human tumors

RNA-seq was performed on n = 18 zebrafish VGLL2-NCOA2 tumors, n = 7 adult zebrafish mature skeletal muscle samples, and n = 3 pooled samples of larval zebrafish from segmentation time points at 1–4 somites, 14–19 somites, and 20–25 somites (10.33, 16, and 19 h post fertilization at 28°C, respectively).

(A) Venn diagram depicting the genes differentially regulated in the developmental time points and tumors compared with mature skeletal muscle, with n = 27 genes differentially regulated in both tumors and development. A false discovery rate (FDR) of 0.01 was used for genes to be included in this analysis.

(B) Plot of FPKM values from n = 8 genes downregulated in developmental time points and tumors compared with mature skeletal muscle. The error bars represent the mean ± SD.

(C) Plot of FPKM values from n = 19 genes upregulated in developmental time points and tumors compared with mature skeletal muscle. The error bars represent the mean ± SD.

(D) Comparison of the expression of these 27 developmental genes in a mouse VGLL2-NCOA2-driven allograft model. Plotted is the fold change of these 27 genes in VGLL2-NCOA2 zebrafish tumors (n = 18) compared with mature skeletal muscle (n = 7), and in C2C12-VGLL2NCOA2 mouse allografts (n = 8) compared with mature skeletal muscle (n = 20; NCBI BioProjects PRJNA625451,[28] PRJNA608179,[29] PRJNA819493, PRJNA813153[30]). If the gene is red, it is statistically significant, and, if it is blue, it is not statistically significant in the mouse context.

(E) Comparison of the expression of these 27 developmental genes in human VGLL2-NCOA2 tumors. Plotted is the fold change of these 27 genes in VGLL2-NCOA2 zebrafish tumors (n = 18) compared with mature skeletal muscle (n = 7), and in human VGLL2-NCOA2 tumors (n = 5; Watson et al.[4]) compared with mature skeletal muscle (n = 803; GTEx version 8). If the gene is red, it is statistically significant, and, if it is blue, it is not statistically significant in the human context. An FDR of 0.1 was used for genes to be included in the analysis in (D) and (E).

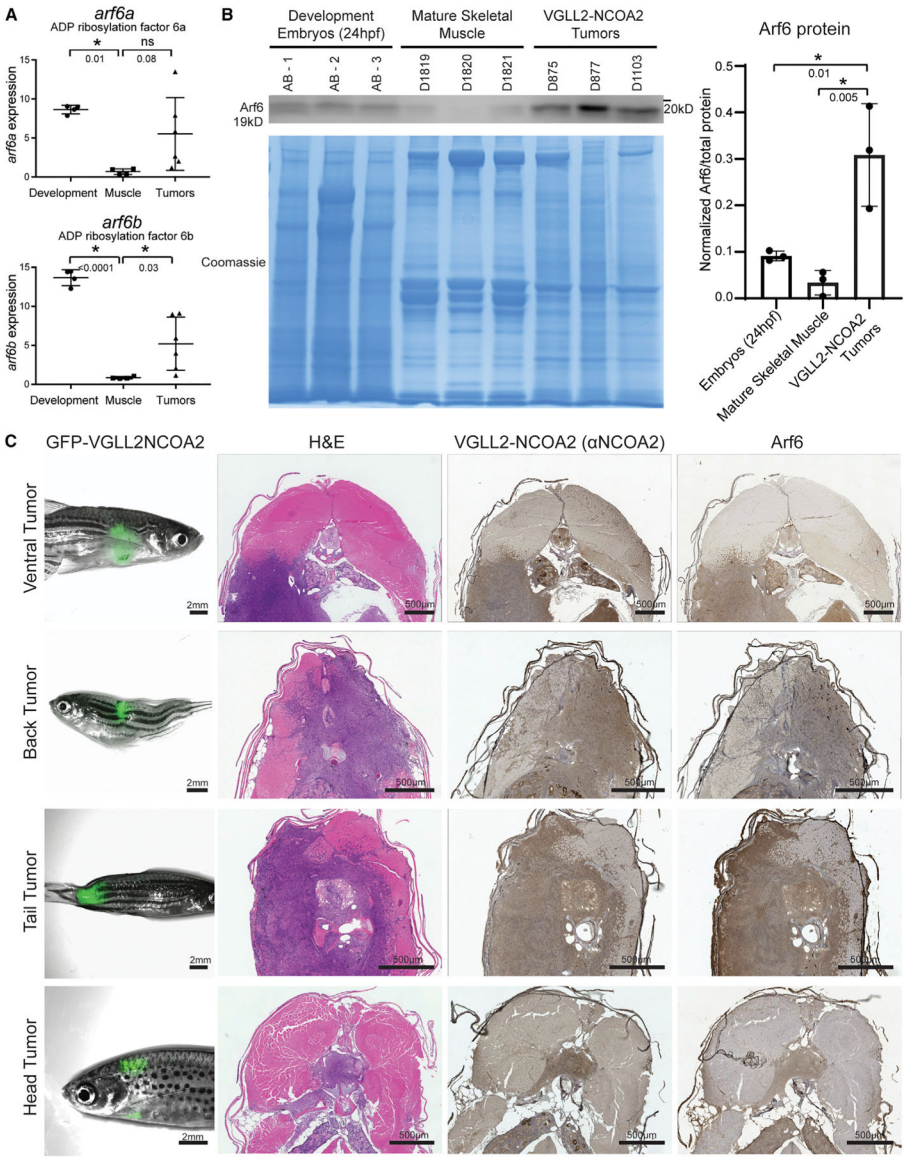


Figure 5. Zebrafish VGLL2-NCOA2 tumors express Arf6 protein

(A) qRT-PCR of *arf6a* and *arf6b* levels using an independent cohort of n = 6 zebrafish tumors, n = 4 zebrafish mature skeletal muscle samples, and n = 4 pools of zebrafish embryos at a developmental time point of 20 h post fertilization. Each data point is an individual tumor or normal tissue sample. The error bars represent the mean ± SD. The p values were calculated using a one-way ANOVA followed by Tukey's multiple comparisons post hoc test.

(B) Western blot for Arf6 protein. For each group, protein lysate was loaded from pools of developing wild-type embryos (n = 12 embryos per pool) or 25 μg of lysate from mature skeletal muscle or VGLL2-NCOA2 tumors. Loading consistency is verified by Coomassie staining. Quantification is performed by normalizing Arf6 to total protein as determined by Coomassie. Each plotted data point is a biological replicate. The error bars represent the

mean ± SD. The p values were calculated using a one-way ANOVA followed by Tukey's multiple comparisons post hoc test.

(C) Representative images of VGLL2-NCOA2 zebrafish tumors shown as brightfield overlaid with GFP fluorescence (scale bars, 2 mm), with serial transverse sections stained with H&E, an anti-NCOA2 antibody, and an anti-ARF6 antibody (scale bars, 500 μm). All n = 8 zebrafish tumors stained for Arf6 were positive.

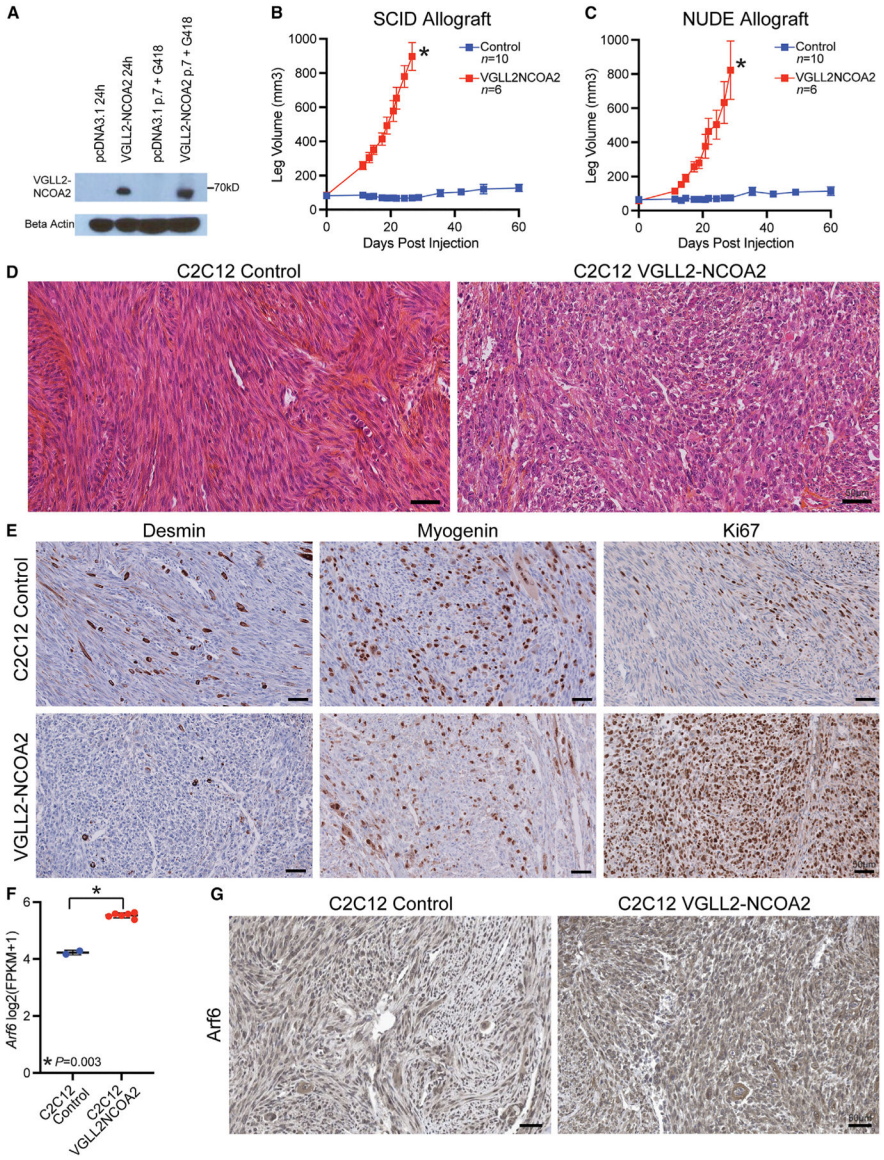


Figure 6. VGLL2-NCOA2 is transforming in mouse myoblast allograft models

(A) C2C12 mouse myoblasts transfected with human VGLL2-NCOA2 express the VGLL2-NCOA2 fusion protein 24 h post transfection and stably after passaging and selection in growth medium with G418 compared with a C2C12-pcDNA3.1 empty control.

(B) Allograft leg volume of C2C12-pcDNA3.1 control and C2C12-VGLL2NCOA2 after intramuscular injection into the leg of SCID mice. C2C12-pcDNA3.1 control was injected in n = 10 allografts and C2C12-VGLL2NCOA2 was injected in n = 6 allografts. The error bars represent the mean leg volume ± SEM. Error bars are not shown if it is within the boundaries of the symbol. Leg volume at time points were compared using a Mann-Whitney U test corrected for multiple comparisons using the Benjamini, Krieger, and Yekutieli method. *p < 0.0005. Every time point after zero is statistically significant.

(C) Allograft leg volume of C2C12-pcDNA3.1 control and C2C12-VGLL2NCOA2 after intramuscular injection into the leg of Swiss Nude mice. C2C12-pcDNA3.1 was injected

in n = 10 allografts and C2C12-VGLL2NCOA2 was injected in n = 6 allografts. The error bars represent the mean leg volume ± SEM. Error bars are not shown if they are within the boundaries of the symbol. Tumor volume at time points were compared using a Mann-Whitney U test corrected for multiple comparisons using the Benjamini, Krieger, and Yekutieli method. *$p < 0.0005$. Every time point after zero is statistically significant.

(D) Representative hematoxylin-eosin saffron (HES) staining from a C2C12-pcDNA3.1 control allograft and a C2C12-VGLL2NCOA2 tumor allograft. Scale bars, 50 μm.

(E) Representative immunohistochemistry for Desmin, Myogenin, and Ki67 for C2C12-pcDNA3.1 control allograft and a C2C12-VGLL2NCOA2 tumor allograft. Scale bars, 50 μm.

(F) *Arf6* mRNA levels from RNA-seq of allografts from Swiss Nude mice, including two C2C12-pcDNA3.1 controls and six C2C12-VGLL2NCOA2 tumors. Error bars represent mean ± SD. *Arf6* mRNA expression was compared using a Welch's t test.

(G) Representative immunohistochemistry for Arf6 protein levels in C2C12-pcDNA3.1 controls and C2C12-VGLL2NCOA2 tumors. Eleven C2C12-pcDNA3.1 controls and six C2C12-VGLL2NCOA2 tumor samples were analyzed. Results are presented in Table S4. Scale bars, 50 μm.

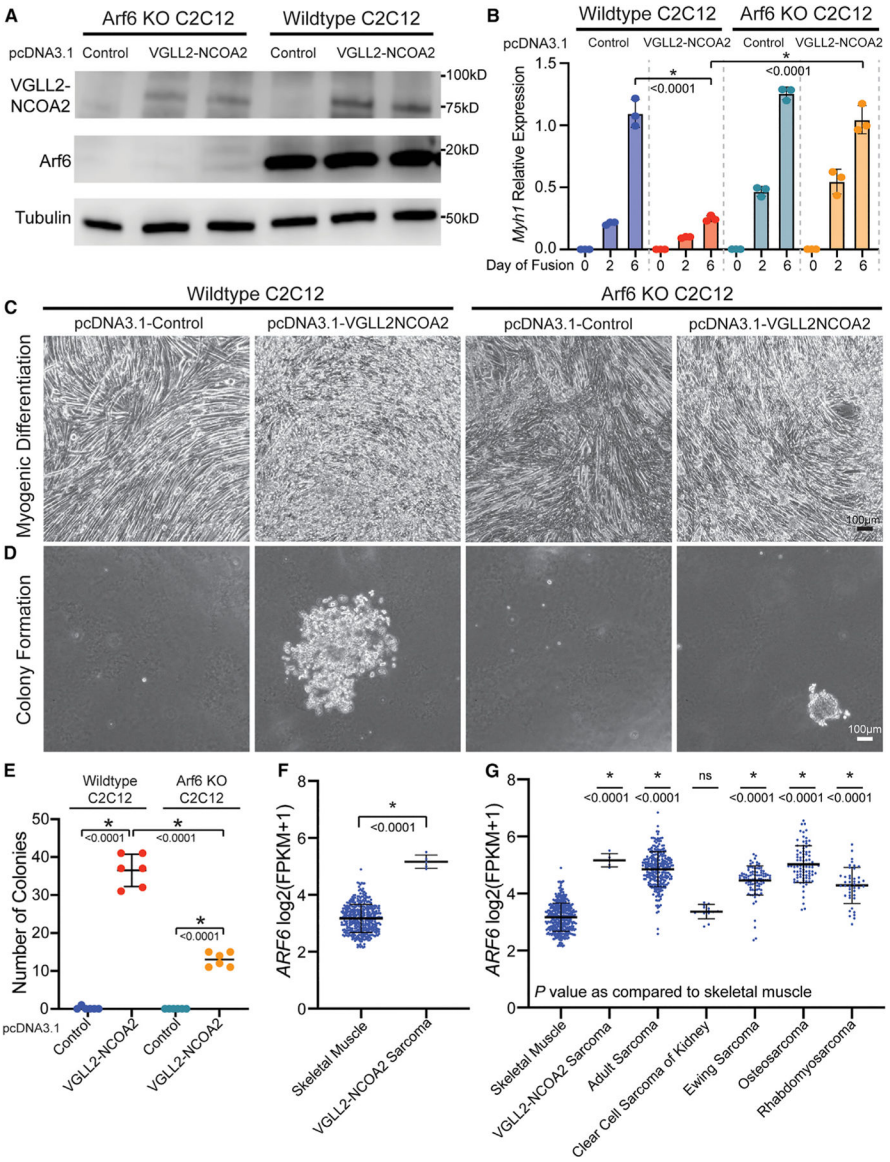


Figure 7. Genetic cooperation of VGLL2-NCOA2 and ARF6 for sarcomagenesis

(A) Monoclonal C2C12 populations with stable Arf6 knockout were generated with CRISPR-Cas9 and guides targeting *Arf6*. Wild-type or Arf6 knockout cells were transfected with pCDNA3.1 empty control or pcDNA3.1 expressing VGLL2-NCOA2 under control of the CMV promoter. Shown is a western blot of generated C2C12 cell lines indicating VGLL2-NCOA2 fusion expression, Arf6 expression, and Tubulin as a loading control.

(B) qRT-PCR for *Myh1* expression relative to loading controls, *Gapdh* and *Rpl27*, for days 0, 2, and 6 of myogenic fusion. Each data point is a biological replicate. The error bars represent the mean ± SD. The p values were calculated using a one-way ANOVA followed by Tukey's multiple comparisons post hoc test. This was repeated once.

(C) Representative brightfield images of myotube morphology taken on day 6 of fusion for each treatment group. Scale bar, 100 μm.

(D) Representative brightfield images of colonies in soft agar taken on day 14 post plating for each treatment group. Scale bar, 100 μm.

(E) Quantification of the number of colonies per well on a six-well plate. Each data point is a biological replicate. The error bars represent the mean ± SD. The p values were calculated using a one-way ANOVA followed by Tukey's multiple comparisons post hoc test. This was repeated twice.

(F) *ARF6* expression determined by RNA-seq in human mature skeletal muscle (n = 396; GTEx) and VGLL2-NCOA2 tumors from Watson et al.[4] (n = 5; one is a primary and recurrent from the same patient). Each data point is an individual tumor or normal tissue sample. The error bars represent the mean ± SD. The p values were calculated using the Mann-Whitney U test.

(G) *ARF6* expression from (F) and compared with additional sarcoma tumor samples, including adult sarcoma from The Cancer Genome Atlas (TCGA) (n = 264), clear cell sarcoma of kidney (n = 13), Ewing sarcoma (n = 95), osteosarcoma (n = 87), and rhabdomyosarcoma (n = 42). Each data point is an individual tumor or normal tissue sample. The error bars represent the mean ± SD. The p values were calculated using the Mann-Whitney U test and are a comparison of each tumor cohort with skeletal muscle.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-KAT13C/NCOA2	Abcam	Cat# ab10508; RRID:AB_297249
Monoclonal mouse anti-beta actin, clone AC-74	Sigma-Aldrich	Cat# A5316; RRID:AB_476743
Goat Anti-Rabbit IgG (H + L)-HRP Conjugate	Bio-Rad	Cat# 1706515; RRID:AB_11125142
Goat Anti-Mouse IgG (H + L)-HRP Conjugate	Bio-Rad	Cat# 1706516; RRID:AB_11125547
α-Tubulin (DM1A) Mouse mAb	Cell Signaling	Cat# 3873; RRID:AB_1904178
Rabbit polyclonal anti-ARF6 antibody	Abcam	Cat# ab226389
Histone H3K27ac antibody (pAb)	Active Motif	Cat# 39133; RRID:AB_2561016
Opal 520	Fisher Scientific	Cat# NC1601877
Anti-PCNA Antibody, clone PC10	Sigma-Aldrich	Cat# MABE288; RRID:AB_11203836
Cy3 AffiniPure Donkey Anti-Mouse IgG (H + L)	Jackson ImmunoResearch	Cat# 715165150; RRID:AB_2340813
Polyclonal rabbit anti-NCOA2 antibody	Novus BIO	Cat# NB100–1756; RRID:AB_2150934
ARF6 antibody	Abcam	Cat# ab77581; RRID:AB_2058475
Monoclonal Mouse Anti-Desmin, Clone D33	Agilent Dako	Cat# M0760; RRID:AB_2335684
FLEX Monoclonal Mouse anti-Myogenin, Clone F5D	Agilent Dako	Cat# GA067; RRID:AB_2250893
Recombinant Anti-Ki67 antibody (SP6)	Abcam	Cat# ab16667; RRID:AB_302459
ARF6 Antibody (9E7T9)	Novus BIO	Cat# NBP3–15391
Chemicals, peptides, and recombinant proteins		
Cytiva HyClone High Glucose DMEM	Fisher Scientific	Cat# SH3002201
Gibco DMEM, high glucose, pyruvate	Fisher Scientific	Cat# 11–995-065
Alt-R S.p. HiFi Cas9 Nuclease V3	IDT	Cat# 1081061
Alt-R Cas9 Electroporation Enhancer	IDT	Cat# 1075916
Gibco Geneticin Selective Antibiotic (G418 Sulfate)	ThermoFisher Scientific	Cat# 10131035
GenJet In Vitro DNA Transfection Reagent for C2C12 Cells	SignaGen Laboratories	Cat# SL100489-C2C12
Roche cOmplete Protease Inhibitor Cocktail	Sigma-Aldrich	Cat# 11836145001
Halt Protease and Phosphatase Inhibitor Cocktail	ThermoFisher Scientific	Cat# 78446
RIPA Buffer	Sigma-Aldrich	Cat# R0278
SuperSignal West Pico PLUS Chemiluminescent Substrate	ThermoFisher Scientific	Cat# 34577
T-PER Tissue Protein Extraction Reagent	ThermoFisher Scientific	Cat# 78510
QC Colloidal Coomassie	Bio-Rad	Cat# 1610803
iTaq Universal SYBR Green Supermix	Bio-Rad	Cat# 1725122
Invitrogen Prolong Gold Antifade with DAPI	Fisher Scientific	Cat# P36935
3,3′-Diaminobenzidine tetrahydrochloride (DAB)	Sigma-Aldrich	Cat# D5905
Thermo Scientific Peroxidase Suppressor	Fisher Scientific	Cat# PI35000
Gibco TrypLE Express Enzyme (1X), no phenol red	Fisher Scientific	Cat# 12–604-013

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Noble agar, molecular biology and tissue culture grade	Fisher Scientific	Cat# 50–488-478
Taq DNA Polymerase with Standard Taq Buffer	NEB	Cat# M0273
Gibco DMEM, powder, high glucose, pyruvate	Fisher Scientific	Cat# 12–800-017
Nitro blue Tetrazolium Chloride	ThermoFisher Scientific	Cat# N6495
Critical commercial assays		
RNeasy Micro Kit	QIAGEN	Cat# 74004
4D-Nucleofector Kit	Lonza	Cat# 1032
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69504
Cell Line Nucleofector Kit V	Lonza	Cat# VCA-1003
Pierce BCA Protein Assay Kit	ThermoFisher	Cat# 23225
MinElute PCR Purification Kit	Scientific QIAGEN	Cat# 28004
RNAscope Multiplex Fluorescent Reagent Kit v2	Advanced Cell Diagnostics	Cat# 323100
ImmPRESS HRP Goat Anti-Rabbit IgG Polymer Detection Kit, Peroxidase	Vector Laboratories	Cat# MP-7451
RNeasy Mini Kit	QIAGEN	Cat# 74104
RNeasy Micro Plus Kit	QIAGEN	Cat# 74034
RT^2 HT First Strand Kit	QIAGEN	Cat# 330411
Deposited data		
Raw data and analysis	This paper	GEO: GSE206039
Zebrafish reference genome build 11, GRCz11	Genome Reference Consortium	RefSeq: GCF_000002035.6
Mouse reference genome build 38, GRCm38	Genome Reference Consortium	RefSeq: GCF_000001635.20
Human reference genome build 38 patch 4, GRCh38p4	Genome Reference Consortium	RefSeq: GCF_000001405.30
RNA-seq data from human VGLL2-NCOA2 and CIC-DUX4 tumors	Watson et al., 2018[4]	EGA: EGAS00001002189
RNA-seq data from human mature skeletal muscle	Broad Institute	GTEx Analysis Release V8
RNA-seq of mouse mature skeletal muscle	Tanner et al., 2021[28]	BioProject: PRJNA625451
RNA-seq of mouse mature skeletal muscle	Chakraborty et al., 2020[29]	BioProject: PRJNA608179
RNA-seq of mouse mature skeletal muscle		BioProject: PRJNA819493
RNA-seq of mouse mature skeletal muscle	Mao et al., 2022[30]	BioProject: PRJNA813153; GEO: GSE197972
RNA-seq data from KRAS driven rhabdomyosarcoma zebrafish tumors and controls	Ignatius et al., 2018[48]	GEO: GSE109581
RNA-seq data for adult sarcomas	Cancer Genome Atlas Research, 2017[49]	https://portal.gdc.cancer.gov/projects
RNA-seq data from Ewing sarcoma tumors	Crompton et al., 2014[50]	dbGaP: phs000804.v1.p1
RNA-seq data from osteosarcoma tumors	TARGET Initiative	dbGaP: phs000468
RNA-seq data from clear cell sarcoma of the kidney	TARGET Initiative	dbGaP: phs000466
RNA-seq data from fusion-positive and fusion-negative rhabdomyosarcoma	Chen et al., 2013[51]	EGA: EGAD00001000878
Zebrafish developmental RNA-seq data	White et al., 2017[26]	ENA: ERP014517
Experimental models: Cell lines		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
C2C12 mouse myoblasts	ATCC	Cat# CRL-1722
C2C12-pcDNA3.1 empty	This paper	N/A
C2C12-VGLL2NCOA2	This paper	N/A
Arf6 KO C2C12	This paper	N/A
Arf6 KO C2C12-pcDNA3.1 empty	This paper	N/A
Arf6 KO C2C12-VGLL2NCOA2	This paper	N/A
Experimental models: Organisms/strains		
Zebrafish: AB	ZIRC	Cat# ZL1
Zebrafish: WIK	ZIRC	Cat# ZL84
Zebrafish: TL	ZIRC	Cat# ZL86
Zebrafish: *tp53*^{M214K}	Laboratory of Tom Look (Berghmans et al., 2005)[52]	ZIRC: ZL1057
Mouse: CB17/Icr-*Prkdc*^{scid}/IcrIcoCrl	Charles River	Cat# 236
Mouse: Crl:NU(Ico)-*Foxn1*^{nu}	Charles River	Cat# 620
Oligonucleotides		
Arf6 gRNA (mouse): TCTTCGGGAACAAGGAAATG	Synthego	N/A
Mouse Negative Control Primer Set 2	Active Motif	Cat# 71012
RNAscope Probe-Dr-myog	Advanced Cell Diagnostics	Cat# 895021
RNAscope Probe-Dr-desma-C1	Advanced Cell Diagnostics	Cat# 1006941-C1
RNAscope Probe-DR-myod1	Advanced Cell Diagnostics	Cat# 481231
Primers for RT-PCR, cloning, CRISPR/Cas9, ChIP-qPCR, qRT-PCR	See Table S5	N/A
Recombinant DNA		
pDestTol2pA2 destination vector	Kwan et al., 2007[53]	N/A
p3E_SV40 late polyA	Kwan et al., 2007[53]	N/A
p5E_beta actin	Kwan et al., 2007[53]	N/A
p5E_cmv	Kwan et al., 2007[53]	N/A
p5E_mcs	Kwan et al., 2007[53]	N/A
pENTR5'_ubi	Laboratory of Leonard Zon (Mosimann et al., 2011)[54]	Addgene: 27320
p5E_unc503	Laboratory of Peter Currie (Berger and Currie, 2013)[55]	Addgene: 64020
pmE_beta globin splice acceptor	Kawakami et al., 2004[56]	N/A
pmE_GFP-viral2A	Kendall et al., 2018[23]	N/A
pCS2FA-transposase	Laboratory of Koichi Kawakami (Urasaki et al., 2006)[25]	N/A
p3E_VGLL2NCOA2	This paper	N/A
CMV-GFP2A-VGLL2NCOA2	This paper	N/A
BetaActin-GFP2A-VGLL2NCOA2	This paper	N/A
ubi-GFP2A-VGLL2NCOA2	This paper	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
unc503-GFP2A-VGLL2NCOA2	This paper	N/A
MCS-beta-globin-SpliceAcceptor-GFP2A-VGLL2NCOA2	This paper	N/A
Invitrogen pcDNA3.1(+) Mammalian Expression Vector	Fisher Scientific	Cat# V79020
pcDNA3.1(+) – VGLL2NCOA2	This paper	N/A
Software and algorithms		
Code utilized in data analysis	This paper	https://github.com/MVesuviusC/2022_VGLL2_NCOA2_paper
Benchling	N/A	https://www.benchling.com/
ICE Analysis	Synthego	https://ice.synthego.com/#/
Trim Galore	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Bowtie 2 v2.3.4.3	Langmead and Salzberg, 2012[57]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
STAR v2.7.2b	Dobin et al., 2013[58]	https://github.com/alexdobin/STAR/releases
SAMtools v1.9	Li et al., 2009[59]	https://github.com/samtools/
Python	Python	https://www.python.org/downloads/
HTSeq	Anders et al., 2015[60]	https://pypi.org/project/HTSeq/
R	The R Foundation	https://www.r-project.org/
Bioconductor	Gentleman et al., 2004[61]	https://www.bioconductor.org/install/
edgeR	McCarthy et al., 2012[62]; Robinson et al., 2010[63]	http://bioconductor.org/packages/release/bioc/html/edgeR.html
AGDEX	Pounds et al., 2011[27]	https://www.stjude.org/research/departments/biostatistics/software/agdex.html
Kallisto 0.46.1	Bray et al., 2016[64]	http://pachterlab.github.io/kallisto/download
Biomart	Ensembl	https://www.ensembl.org/biomart/martview
ComBat	Zhang et al., 2020[65]	https://www.bioconductor.org/packages/release/bioc/html/sva.html
DAVID	Huang da et al., 2009a,b[66,67]	https://david.ncifcrf.gov/
clusterProfiler R package	Wu et al., 2021[68]; Yu et al., 2012[69]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Prism 8.1.1	GraphPad	https://www.graphpad.com/
ImageJ	NIH	https://imagej.nih.gov/ij/download.html