scientific reports

Check for updates

OPEN

Defining the transcriptome of *PIK3CA*-altered cells in a human capillary malformation using single cell long-read sequencing

Michelle A. Wedemeyer[1,2,3,4], Tianli Ding[1], Elizabeth A. R. Garfinkle[1], Jesse J. Westfall[1], Jaye B. Navarro[1], Maria Elena Hernandez Gonzalez[1], Elizabeth A. Varga[1], Patricia Witman[2,5], Elaine R. Mardis[1,2,4], Catherine E. Cottrell[1,2,6], Anthony R. Miller[1] & Katherine E. Miller[1,2]✉

***PIK3CA*-related overgrowth spectrum (PROS) disorders are caused by somatic mosaic variants that result in constitutive activation of the phosphatidylinositol-3-kinase/AKT/mTOR pathway. Promising responses to molecularly targeted therapy have been reported, although identification of an appropriate agent can be hampered by the mosaic nature and corresponding low variant allele frequency of the causal variant. Moreover, our understanding of the molecular consequences of these variants—for example how they affect gene expression profiles—remains limited. Here we describe in vitro expansion of a human capillary malformation followed by molecular characterization using exome sequencing, single cell gene expression, and targeted long-read single cell RNA-sequencing in a patient with clinical features consistent with Megalencephaly-Capillary Malformation Syndrome (MCAP, a PROS condition). These approaches identified a targetable *PIK3CA* variant with expression restricted to PAX3+ fibroblast and undifferentiated keratinocyte populations. This study highlights the innovative combination of next-generation single cell sequencing methods to better understand unique transcriptomic profiles and cell types associated with MCAP, revealing molecular intricacies of this genetic syndrome.**

Somatic variation occurring within RAS-MAPK, PI3K-AKT, and related pathways are increasingly recognized as a cause of a wide spectrum of vascular malformations and overgrowth syndromes including lymphatic malformations, arteriovenous malformations, cavernous malformations, megalencephaly-capillary malformation syndrome (MCAP), *PIK3CA*-related overgrowth spectrum (PROS) disorders, megalencephaly-polydactyly-polymicrogyria-hydrocephalus syndrome (MPPH), and others[1–10]. The clinical spectrum of disease ranges from mild, localized vascular, or lymphatic malformations to debilitating overgrowth syndromes accompanied by developmental delay, seizures, hydrocephalus, symptomatic Chiari I malformations, and body asymmetry[1,3–6,9–16]. Although clinical improvement has been reported with the use of molecularly targeted therapies such as alpelisib, clinical molecular testing may be non-diagnostic due to the mosaic nature and low variant allele fraction/frequency (VAF) of the variant in available tissues for study[17–21]. Strategies to overcome challenges in detecting pathogenic variants with low VAF have included in vitro expansion of affected tissues prior to sequencing, high depth sequencing with targeted gene panels, or droplet-based PCR gene panels to study DNA extracted from biofluids such as cerebrospinal fluid, cyst fluid, or venous blood draining affected tissues[3,19,22–25].

The introduction of a genetic change during embryogenesis can result in the presence of two or more genetically distinct cell lines in an individual, a phenomenon known as mosaicism. Phenotypic consequences of somatic mosaicism are impacted by the nature and timing of the genetic change including the distribution of affected tissues and extent of disease. Some clinical laboratory genetic tests include *PIK3CA* among larger constitutional gene panels focused on macrocephaly or overgrowth syndromes for individuals in whom a differential diagnosis

[1]The Steve and Cindy Rasmussen Institute for Genomic Medicine, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, USA. [2]Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA. [3]Department of Neurosurgery, Nationwide Children's Hospital, Columbus, OH, USA. [4]Department of Neurosurgery, The Ohio State University College of Medicine, Columbus, OH, USA. [5]Division of Dermatology, Nationwide Children's Hospital, Columbus, OH, USA. [6]Department of Pathology, The Ohio State University College of Medicine, Columbus, OH, USA. ✉email: Katherine.Miller@nationwidechildrens.org

may be broad[3]. *PIK3CA* testing for the diagnosis of PROS usually requires specimens from an affected region, such as a skin biopsy or a surgical sample from an overgrown tissue, or vascular or lymphatic malformation. Previous reports have shown that postzygotic *PIK3CA* mutations are generally not detectable in blood samples of affected individuals, though exceptions have occurred in individuals with MCAP syndrome[26]. Improved detection sensitivity has also been reported with enrichment for suspected causative cell types such as CD31+ endothelial cells; however, sampling of involved tissue can be impractical such as in the setting of a predominantly central nervous system phenotype[19,27]. Despite the sometimes significant clinical phenotype and morbidities seen with PROS and the patient reported in this study, the VAF of the causal variant is frequently at or below the level of detection for many standard clinical sequencing assays leading to falsely negative clinical testing. The implementation of targeted panel-based sequencing with read depths up to 500× and sensitivity for VAFs as low as 0.15% have permitted the identification of ever lower frequency mosaicism and categorization of the spectrum of phenotypes associated with specific mutations[3,4,19]. Additionally, while cell-free DNA techniques are emerging for detection of mosaic variation, generally access of disease-involved tissue requires invasive sampling through biopsy or skin punch. The need for invasive testing can create barriers to care due to the desire of families to avoid invasive testing and the need for involving the desire of families to participate in an invasive procedure and the need for prior authorization from payors.

Despite these genomic insights, it remains unclear how low frequency mosaicism involving a small fraction of cells leads to the dramatic spectrum of PROS phenotypes. Techniques to analyze diseased tissues and probe the impact of somatic mosaic variation have evolved in recent years. Single cell RNA sequencing (scRNA-seq) is a powerful tool to analyze transcriptional signatures in individual cells, but current technology, by design, produces short-read sequencing biased towards the 3′ end of transcripts, which has limited value for genotyping variants distal from the polyA tail[28–30]. Integration of long-read sequencing data can be helpful to capture full-length transcripts, however, current long-read technologies have insufficient read output, scalability, and limited depth of coverage for the confident assignment of genotypes to thousands of transcriptionally profiled single cells. MAS-ISO-seq (Multiplexed ArrayS ISOform SEQuencing) was recently developed to permit high-throughput long-read transcriptome sequencing, helping to overcome this barrier[31]. This approach uses PCR to combine up to 15 cDNA molecules into concatenated molecules, which are then sequenced using the Pacific Biosciences (PacBio) circular consensus sequencing approach. Subsequent "de-concatenation" is achieved in silico by leveraging established primer sequences and known 10× single cell barcodes to facilitate the assignment of transcripts to single cells. A comparable library preparation kit called 'Kinnex' corresponding to this method was commercially released by PacBio.

Here, we report the use of MAS-ISO-seq to study the transcriptome in *PIK3CA*-altered fibroblasts from a capillary malformation in a patient with suspected MCAP. Despite initially negative results obtained from sequencing DNA extracted from peripheral blood mononuclear cells to identify an underlying MCAP-associated variant, high-depth whole exome sequencing of cultured cells identified a constitutively activating single nucleotide variant (SNV) in *PIK3CA*: NM_006218.4:c.3139C>T;p.His1047Tyr; this variant is known to confer susceptibility to the *PIK3CA* inhibitor alpelisib and to mTOR inhibitors and confirmed the diagnosis of MCAP. Subsequent single cell RNA-sequencing of cultured cells using two next-generation sequencing (NGS) approaches (short-read sequencing of the 3′ ends of transcripts and long-read targeted sequencing of *PIK3CA*) identified a PAX3+/SOX11+ fibroblast-like population enriched for the variant. Cell clusters enriched for *PIK3CA*-mutant cells (*PIK3CA*^{mut}) exhibited a distinct gene expression profile compared to cell clusters with fewer *PIK3CA*^{mut} cells, providing insight into how a variant restricted to a small population of cells might lead to the dramatic clinical phenotype seen in patients with MCAP. These results illustrate the rapidly advancing potential for the application of single cell sequencing technologies to probe the mechanisms by which previously undetectable somatic mosaicism may lead to debilitating clinical syndromes.

Results

Case description

A male of Northern European ancestry born at 36 weeks was noted at birth to have right facial asymmetry and widespread port wine stains. He was otherwise well-appearing and did not exhibit seizure-like activity. Birth history was significant for in utero exposure to methamphetamines and hepatitis C. Although decreased facial movement was noted on the enlarged right side, the remainder of his neurologic exam was within normal limits. At birth, widespread port wine stains, a form of capillary malformation, were noted throughout his trunk and bilateral lower extremities.

Clinical diagnostic testing

Given concern for a PROS disorder such as Klippel–Trenaunay syndrome or Megalencephaly-Capillary Malformation syndrome (MCAP), the patient was referred to clinical geneticist in the Division of Genetic & Genomic Medicine. An MRI of the brain revealed asymmetric enlargement of the right cerebral hemisphere and cortical dysplasia (Fig. 1a,b). Lower extremity radiographs showed a leg length discrepancy of > 1 cm (Fig. 1c). Due to the constellation of clinical findings, blood was sent for *PIK3CA* sequence analysis and deletion/duplication testing on day of life 5, but no pathogenic variants were identified. An abdominal ultrasound was negative for embryonal tumors. Formal language testing at 2 years 4 months of age identified a mixed receptive-expressive language disorder with an overall language ability at the 7th percentile for his age group. Subsequent developmental testing found him to be below average for his age group in the domains of physical, social-emotional, cognitive, communication, and general development with average adaptive skills.

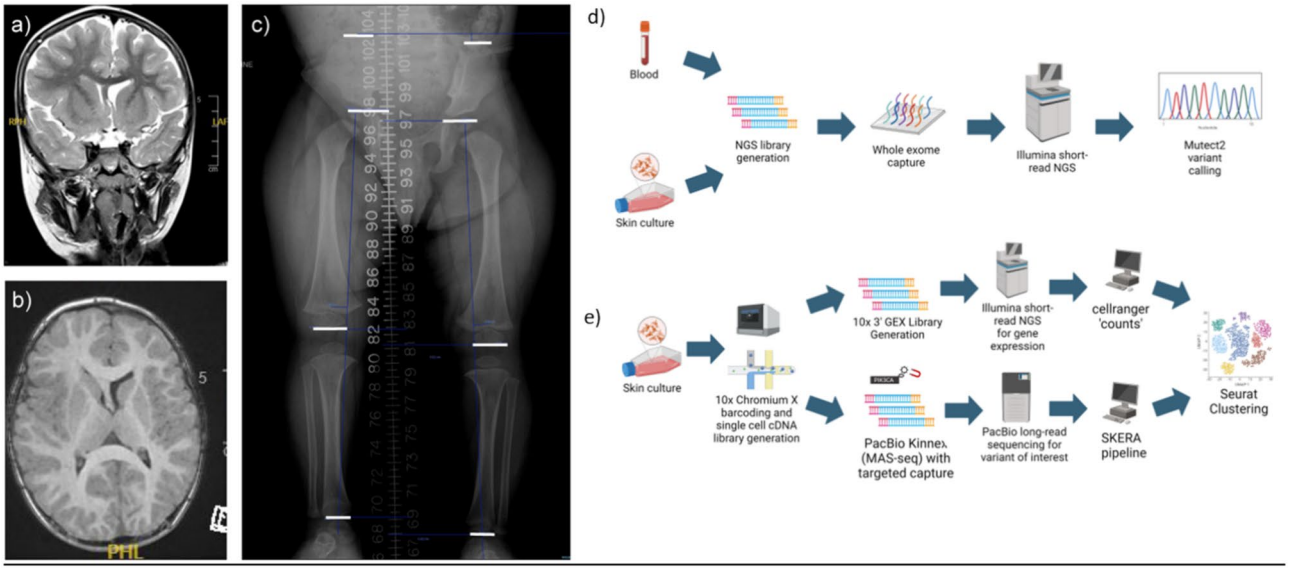

Fig. 1. Clinical presentation and diagnostic approach. (**a,b**) T2 coronal (**a**) and T1 axial (**b**) MRI obtained age 2 shows enlarged right cerebral cortex with effacement of the right ventricle and cortical dysplasia. (**c**) AP lower extremity radiograph obtained at age 2 shows right greater than left leg length discrepancy. (**d**) Diagnostic testing included whole exome sequencing from skin cell culture. (**e**) Single cell gene expression was performed, in parallel with long-read sequencing, targeted for *PIK3CA* transcripts, to enable genotyping of single cells for the pathogenic *PIK3CA* variant.

High-depth exome sequencing identifies a pathogenic *PIK3CA* variant

Given continued clinical suspicion for a PROS disorder, a skin biopsy of the right trunk was obtained at 2 years 2 months of age for in vitro expansion of skin affected by port wine stains. Targeted sequencing of DNA from cultured fibroblasts performed at an outside clinical lab for *PIK3CA* copy number and sequence variation (mean depth 144×) was negative. The patient was subsequently enrolled in a translational research protocol and high-depth exome sequencing was performed on DNA extracted from cultured cells (Fig. 1d) (mean depth of 244× coverage) which identified a missense variant in *PIK3CA* (NM_006218.4:c.3139C>T;p.His1047Tyr). Allelic depth at the variant position was 386× with a variant allele frequency/fraction (VAF) of 11.9% for the *PIK3CA*:p.His1047Tyr alteration. Clinical Sanger sequencing of PCR products obtained from the same DNA extract confirmed the presence of this activating variant (Fig. S1), which is known to confer susceptibility to the PIK3CA inhibitor alpelisib and mTOR inhibitors (tacrolimus, everolimus, sirolimus). This variant was classified as a Tier I (Level A) variant indicating confirmed pathogenicity[32].

PIK3CA mutation status is associated with increased fraction of G1 phase cells

To better understand the effect of *PIK3CA* mutational status on the transcriptome of individual cells, we performed short- and long-read single cell sequencing of cultured cells (Fig. 1d,e). Single cell short-read RNA-sequencing yielded 12,378 cells that passed QC filtering and were included in downstream analysis. A total of 10 clusters were identified from the single cell data (Fig. 2a). Although short-read RNA-sequencing techniques are capable of identifying variants biased towards the 3′ end, few reads will be of sufficient length to capture variants at longer distances from the 3′ end. Overall, 1.2% (173/14,092) of single cells profiled by short-read sequencing had coverage to capture the *PIK3CA*:c.3139C>T locus, permitting genotyping of these cells using short-read methodology alone, although the small number of cells does not permit differential gene expression. By incorporating long-read sequencing of the transcripts via MAS-ISO-seq long-read RNA-sequencing of the *PIK3CA* transcript to the barcoded single cell transcriptomes, we captured the locus of the known *PIK3CA:c.3139C>T* variant in 15.3% (1894/12,378) of cells profiled. Thus, the use of the multiplexed MAS-ISO-seq methodology permitted > 12-fold improvement over short-read technologies in the capacity to genotype single cells (Fig. S2).

Gene expression quantification through short- and long-read single cell sequencing revealed variation driven by *PIK3CA* mutational status and cell cycle status in the first principal component, while the second principal component was driven primarily by cell cycle status (Fig. 2b–d). Despite cell cycle regression, the top five biological processes contributing to the PCA space were all associated with the cell cycle including Cell Cycle Process, Mitotic Cell Cycle, Mitotic Cell cycle process, Chromosome Organization, and Cell Division (Table S1). *PIK3CA* mutation status was assigned to clusters with > 100 genotyped cells. Clusters with *PIK3CA* mutations in > 10% of genotyped cells were designated $PIK3CA^{mut}$ while those with < 10% mutational burden were designated $PIK3CA^{wt}$ (Fig. 2c). A greater proportion of $PIK3CA^{mut}$ cells were assigned to the G1 growth phase of the cell cycle (Fig. 2d).

$PIK3CA^{mut}$ cells express neural crest lineage markers

Although significant homogeneity was noted overall, clusters enriched for $PIK3CA^{mut}$ cells clustered together along the first principal component (Fig. 2c). Using automated mapping to a publicly available reference skin

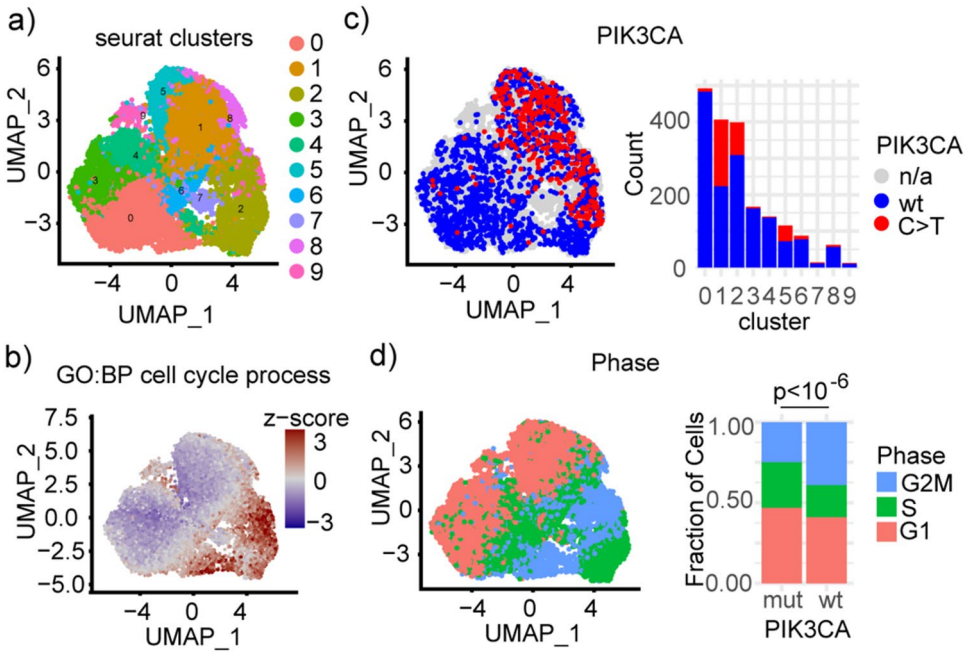

Fig. 2. *PIK3CA*^{mut} cells are transcriptionally distinct and mitotically active. (**a**) Unsupervised clustering using principal component analysis followed by dimensionality reduction using uniform manifold approximation and projection (UMAP) yielded 10 clusters. (**b**) A UMAP plot of enrichment for the GO:BP cell cycle pathway shows a significant contribution to both the first and second principal components. (**c**) Cells enriched for *PIK3CA*^{mut} cells cluster together on a UMAP plot. Barplot (right) showing count of genotyped cells in each cluster. (**d**) UMAP (left) plot shows cells clustering by phase and barplot (right) shows an increased fraction of *PIK3CA*^{mut} cells in the G1 growth phase of the cell cycle.

set[33], our clusters were classified as follows: hip fibroblasts (10,747/12,378; 86.8% of all cells), undifferentiated keratinocytes (882/12,378; 7.1%), palm/sole fibroblasts 1 (372/12,378; 4%), and pericytes (377/12,378; 3%) (Fig. 3a). The reference skin set showed that compared to the bHLH transcription factor *HES1,* which is expressed across multiple cell type clusters, *PAX3* is normally expressed only in Schwann cells and melanocytes (Fig. 3b). However, genotyping of the cultured skin cells (Fig. 3c) revealed *PIK3CA*^{wt} clusters were highly enriched for the *NOTCH3* receptor and *HES1* (Fig. 3d), while *PIK3CA*^{mut} clusters (fibroblasts and undifferentiated keratinocytes) were highly enriched for the neural crest lineage transcription factors *PAX3* and *SOX11* (Fig. 3e). This finding shows that *PIK3CA*^{mut} fibroblasts and undifferentiated keratinocytes express typically melanocyte and neura crest fate specific markers.

PIK3CA^{wt} cells express notch pathway signaling genes

The most highly differentially expressed genes in *PIK3CA*^{wt} cells (versus mutated cells) included genes involved in the insulin-like growth factor pathway (*IGFBP7*, *IGFBP4*), chondroitin sulfate metabolism (*DCN*), non-coding RNAs (*AC011246.1*, *PAX8-AS1*), effectors of canonical notch signaling (*HES1*) and WNT signaling (*WNT5A*). Conversely, the mostly highly expressed genes in *PIK3CA*^{mut} cells were involved in inflammation (*CXCL14*), cytoskeleton (*TUBB2B*), chromatin remodeling (*BC11B*), cell–cell interactions (*NCAM1*), and regulation of transcription (*PAX3, NFIB*) (Fig. 3f, Table S2). An analysis of ligand receptor signaling patterns using CellChat revealed a *PIK3CA*^{wt} fibroblast population as an effector of Notch signaling via JAG1-NOTCH3 and JAG1-NOTC1 signaling (Fig. 4a,b). Gene set enrichment analysis of *PIK3CA*^{wt} cells revealed a predominance of genes involved in blood vessel formation and wound healing while *PIK3CA*^{mut} cells were enriched for genes involved in cell division (Fig. 4c,d).

PIK3CA mutation status drives chemokine signaling

Analysis of the communication probabilities of 1939 receptor-ligand signaling pairs in 223 pathways using CellChat identified three outgoing signaling (sender) patterns representing undifferentiated keratinocytes (Pattern 1, black), *PIK3CA*^{wt} fibroblasts 1–3 and pericytes (pattern 2, blue), and *PIK3CA*^{mut} fibroblasts 1–3 and palm/sole fibroblasts (pattern 3, red) (Fig. 4a). While sender pattern 1 was dominated by epidermal growth factor (EGF), immune (*CD226*, *IL6*), and endothelin (EDN) pathways, sender pattern 2 was dominated by ephrin (EPHA), periostin, activin, and BMP pathways. The most significant pathway contribution to the relatively silent *PIK3CA*^{mut} sender pattern 3 was the cell adhesion molecule pathway (CADM). Incoming receiver patterns were not associated with *PIK3CA*^{mut} status (Figs. 4c, 5a,b).

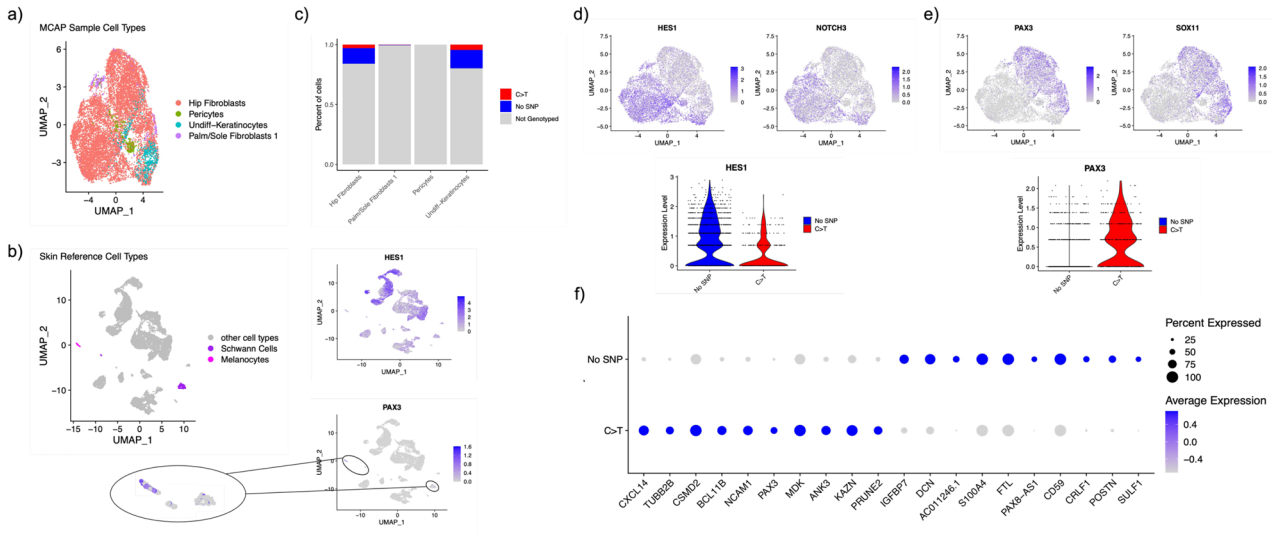

Fig. 3. *PIK3CA*^{mut} clusters express neural crest markers. (**a**) UMAP plot of cultured skin cells mapped to a normal human skin cell reference. (**b**) UMAP plot of normal human skin cell reference with melanocytes (pink) and Schwann cells (purple) highlighted. Feature plots (right) showing the wide expression of the bHLH transcription factor *HES1* across multiple cell clusters and the limited expression of neural crest transcription factor *PAX3* in melanocytes and Schwann cells. (**c**) Barplot displaying percent of genotyped cells in each cell type. (**d**) Feature plot (top) showing expression of the Notch-regulated bHLH transcription factor *HES1* and the *NOTCH3* receptor in the *PIK3CA*^{wt} enriched clusters and violin plot (lower) showing expression of HES1 transcript in *PIK3CA*^{mut} versus *PIK3CA*^{wt} clusters. (**e**) Feature plot (top) showing expression of the neural crest transcription factors *PAX3* and *SOX11* in the *PIK3CA*^{mut} enriched clusters and violin plot (lower) showing expression of *HES1* transcript in *PIK3CA*^{mut} versus *PIK3CA*^{wt} clusters. (**f**) Dotplot showing expression of top 10 expressed markers in *PI3KCA*^{wt} versus *PIK3CA*^{mut} cells.

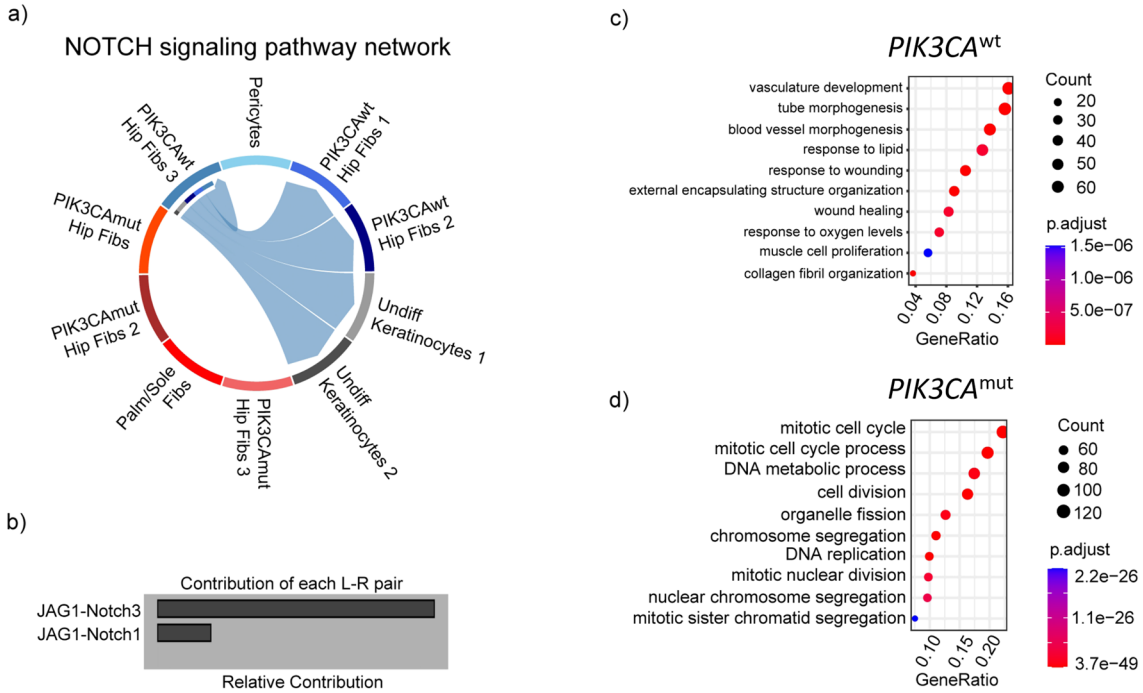

Fig. 4. Notch signaling is absent in *PIK3CA*^{mut} populations. (**a**) Chord diagram showing Notch-mediated signaling between *PIK3CA*^{wt} but not *PIK3CA*^{mut} fibroblasts and undifferentiated keratinocytes. (**b**) Barplot showing the relative contribution of the JAG1-NOTCH3 versus JAG1-NOTCH1 receptor signaling to overall Notch pathway signaling. (**c,d**) Dotplot of the top 10 enriched Gene Ontology Biological Processes in *PIK3CA*^{wt} (**c**) and *PIK3CA*^{mut} (**d**) cells.

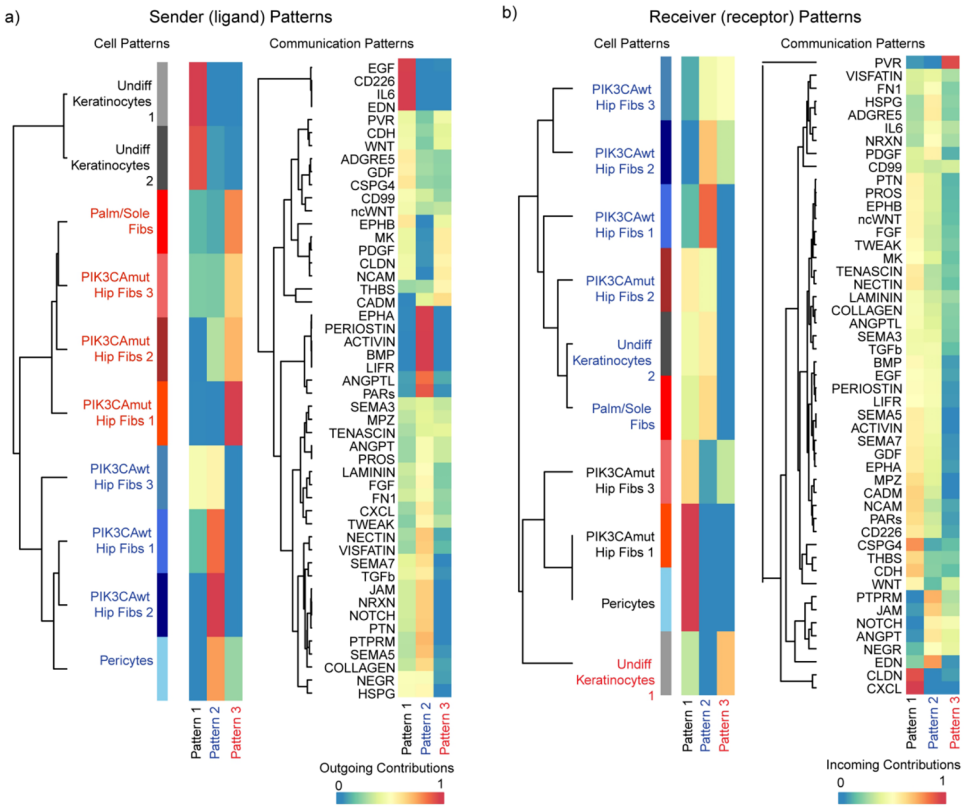

Fig. 5. *PIK3CA*^{mut} clusters exhibit distinct receptor/ligand signaling behavior. (**a**) Heatmaps of sender patterns of clusters show three predominant cell patterns (left) corresponding to ligand-mediated communications (right) from keratinocytes, *PIK3CA*^{wt}, and *PIK3CA*^{mut} clusters respectively. (**b**) Heatmaps of receiver cell patterns (left) and receptor-mediated communication patterns (right) showing the relative contribution of cell clusters to the microenvironment.

Discussion

The detection of disease-associated somatic mosaic variants in patients with clinical signs suggestive of PROS and related conditions is challenging due to the low VAF. These variants can escape detection in clinical sequencing when limited tissue samples are tested or if assay sensitivity is inadequate. Consequently, implementing molecularly targeted therapy to address potentially life-threatening symptoms becomes hindered. In our study, we addressed this challenge by employing clinical cell culture expansion of a skin punch biopsy to facilitate the diagnosis of a molecularly targetable *PIK3CA* variant. Bulk exome sequencing of DNA derived from cultured cells enabled enrichment for a dividing cell population and the subsequent identification of this variant via standard clinical whole exome sequencing followed by confirmatory Sanger sequencing[1]. The use of cell culture to selectively amplify dividing cells from affected tissues is a useful adjunct to direct sequencing of affected tissues and may be readily translated into the workflow of many clinical laboratories. Additionally, we utilized a novel single-cell assay to perform gene expression profiling coupled with single cell genotyping via targeted long-read RNA-sequencing from the same cell, facilitating mechanistic insights into disease pathogenesis.

This innovative single cell approach revealed the *PIK3CA* variant was enriched in a fibroblast-like cell population characterized by the expression of neural crest markers *PAX3* and *SOX11* and an absence of NOTCH signaling regulation, which is known to maintain epithelial self-renewal and promote establishment of the keratinocyte lineage[34,35]. In addition to their contribution to the PROS overgrowth spectrum diseases, activating mutations in *PIK3CA* are among the three most commonly identified genetic alterations in cancers[36,37]. *PIK3CA* codes for the catalytic subunit of the enzyme phosphatidylinositol 3-kinase (PI3K) which activates migration, survival, cell cycle, and growth pathways[38]. Mutations are clustered in either the helical domain (E545K, E542K) or the kinase domain (H1047R), as seen in our case[37,39,40]. Several lines of evidence indicate that activating mutations in the kinase domain, including the H1047R mutation, lead to cell growth and hypertrophy via p70S6K and mTOR-dependent mechanisms without activating cell cycle pathways[38–43]. Our observation that *PIK3CA*^{mut} clusters are enriched for cells in the G1 growth phase of the cell cycle is consistent with the clinical finding in this patient of hemibody hypertrophy without development of cancer. Although acquisition of a *PIK3CA* mutation is typically a late event in cancer, clinical trials in multiple cancers have noted that PIK3CA kinase domain mutations confer increased sensitivity to treatment with mTOR inbitors compared to helicase mutations suggesting that the activating role of kinase mutations primarily affects growth pathways[37,39]. Overall, these findings support obervations that the role of PIK3CA in cell division and migration is decoupled from the role of PIK3CA in cell growth and hypertrophy[39,40,42,43].

Expression of *PAX3* in a healthy human skin reference appeared to be restricted to neural crest lineages such as melanocytes and schwann cells[44]. Although *PAX3* is an established marker for skin melanocytes, normal nevi, and malignant melanoma, murine studies have identified *PAX3* as a direct regulator of Notch effectors such as *HES1*[45,46]. As the *PIK3CA*^{mut} PAX3+ populations identified in these in vitro studies do not express other melanocyte markers such as *MLANA*, their in vivo correlate remains to be identified. The finding that clusters enriched for the *PIK3CA* mutation are not subject to NOTCH1/3 regulation is of particular relevance to this patient's clinical condition. NOTCH3, and to a lesser extent NOTCH1, have a well established role in the promotion of angiogenic remodeling of the fetal primary capillary plexus to form arteries[47–52]. Furthermore, *NOTCH1* and *NOTCH3* deficient mouse models exhibit a loss of pericyte-induced stabilization of developing blood vessels that leads to arteriovenous malformations and *NOTCH3* mutations lead to the clinical stroke syndrome CADASIL[48,53,54]. Given the known role of Notch signaling in the maturation of blood vessels, specifically in the involution of small capillaries during the formation of larger vessels, the absence of Notch signaling in cell clusters enriched for activating *PIK3CA* mutations reflects the raises the possibility that an expanding *PIK3CA*^{mut} population leads to widespread disruption of NOTCH1/3-dependent maturation of the fetal capillary plexus into mature arteries and persistence of fetal capillary networks after birth.

Our findings support two potential hypotheses regarding the etiology of the PROS disorders. In one, specific progenitor cell populations are prone to prolonged expansion upon exposure to mutant PIK3CA while in another, the presence of mutant PIK3CA induces a blockade to differentiation. Arteriovenous malformations, a related somatically driven vascular overgrowth anomaly, appear to be driven by activating *KRAS* mutations that are restricted to CD31+ endothelial cells, lending support to the hypothesis that a mutation in a single cell type may lead to an overgrowth syndrome affecting the surrounding tissue[27]. Similarly, our finding that the *PIK3CA*^{mut} cell population exhibits a distinct outgoing signaling profile supports the hypothesis that these cells may exert a unique trophic effect on the surrounding tissue. Indeed, mutant clusters exhibited increased outgoing PDGF signaling which has long been recognized as a driver of angiogenesis[55]. Although the low VAF of pathogenic variants identified in disease-involved tissue specimens from patients is suggestive of a unique molecular event restricted to a specific population of cells, further research is warranted to determine whether the distinct transcriptional signature is a direct result of fate decisions influenced by the constitutively activated *PIK3CA* or rather, related to a post-zygotic mutation event restricted to a select population of fate-committed cells[1,4,15,19,22].

Our research not only enhances our biological understanding of vascular malformations but also introduces a powerful long-read NGS workflow. This innovative approach facilitates the precise assignment of genomic alterations to individual cells within mosaic diseased tissue, opening up new avenues for comprehensive molecular analyses. Although our study utilized a hybridization capture approach to specifically enrich *PIK3CA* transcripts, the described method is suitable for detecting other variants of interest by enriching for different gene set(s). Moving forward, application of multiplexed single cell RNA-sequencing with targeted long-read based genotyping to diseased tissues such as skin, resected vascular malformations, and tonsils resected at the time of surgery for Chiari malformations may be expected to provide further insights into the mechanisms by which small populations of mutated cells influence the surrounding microenvironment to produce human disease.

Methods

Human subjects

Written informed consent was obtained from the patient's parents in this study under a research protocol approved by the Institutional Review Board (IRB) at Nationwide Children's Hospital (IRB17-00206). All research presented in this study was performed in accordance with relevant guidelines and regulations as set forth by the IRB at Nationwide Children's Hospital. The patient is a male of Northern European ancestry who was born at term and presented on the first day of life. He was 3 years old at the time of biopsy.

Cell culture

Skin biopsy tissue from the right trunk was morcellated, digested in collagenase at 37 °C for 1 h then pelleted in a centrifuge at 900 RPM × 10 min. The pellet was resuspended in alpha-MEM with 20% FBS, 1.5% L-glutamine, 1% penicillin streptomycin, and 1 ml fungizone, plated in a T25, and incubated T25 at 37 °C and 5% CO2. Flasks were passaged with trypsin–EDTA.

Exome sequencing

Exome sequencing was performed as a clinical test. Briefly, libraries for enhanced exome sequencing were prepared using 100 ng of DNA isolated from cultured cells using the NEB Ultra II FS Kit (New England Biolabs) followed by target enrichment with IDT xGen Lockdown v2.0 human exome reagent (catalog number 10005153) with xGenCNV Backbone Panel and Cancer spike-in (Integrated DNA Technologies, Coralville, IA). Libraries were sequenced on an Illumina NovaSeq6000 (Illumina, Inc., San Diego, CA) to generate 150 bp paired-end reads. Output data were aligned and analyzed using the Churchill workflow which uses a balanced regional parallelization strategy to perform variant discovery[56].

Clinical sanger sequencing

To validate the *PIK3CA* finding, Sanger sequencing was performed on PCR products amplified from the extracted genomic DNA originally used for exome sequencing. Forward and reverse sequencing reactions were performed with the Big Dye v3.1 terminator mix (ThermoFisher Scientific, Waltham, MA). Sequencing was performed on the Applied Biosystems 3730 instrument. Primer sequences are as follows: PIK3CA_Ex21_F (3′-GTAAAACGACGGCCAGCTGAGCAAGAGGCTTTGGAG-5′); PIK3CA_Ex21_R (3′-CAGGAAACAGCTATGACCAGAGTGAGCTTTCATTTTCTCA-5′).

10× genomics 3′-based single cell RNA-sequencing library preparation

To generate libraries for a 3'-based single cell RNA-sequencing, a single confluent T25 flask of cells cultured from a skin biopsy of affected tissue at passage three were harvested using tryspin-EDTA. Purified cells were then processed for library preparation according to the manufacturer protocol for Chromium Next GEM Single-Cell 3'-Reagent Kit v.3.1. Libraries were sequenced on an Illumina NovaSeq 6000 instrument to generate paired-end sequencing data with a minimum of 50,000 reads per cell.

Long-read single cell RNA-sequencing library preparation

To identify coding variants from single cells prepared using the 10× Genomics 3′ kit, 75 ng of pre-fragmented cDNA from the 10× Genomics workflow was used as input into the PacBio Kinnex single-cell RNA kit (PN 102-166-600) with the following modifications: cDNA remaining post-TSO artifact removal was enriched for *PIK3CA* transcripts using a custom probe panel and the xGen Hybridization and Wash Kit protocol (Integrated DNA Technologies, #1080577). Enriched cDNA was then used as input into Kinnex PCR for subsequent array formation according to the manufacturer's recommendations. Sequencing of the final SMRTbell library was performed using the Sequel II Binding Kit 3.2, the Sequel II Sequencing Plate 2.0, and a single 8 M SMRT Cell with an on plate loading concentration of 100 pM. Data collection included a 2-h pre-extension followed by a 30 h movie.

Analysis of long-read single cell RNA-sequencing data

Preliminary analysis included the use of the PacBio application "Read Segmentation and Single-Cell Iso-Seq" to perform SKERA (https://skera.how/) read splitting or de-concatenation of the array into original 10× Genomic cDNA molecules based on MAS barcodes. A total of 14,999,323 segmented reads were generated with an average length of 637 bp. The analysis application provides aligned BAMs, which were then split using samtools, to generate a new BAM per single cell. Bcftools mpileup was then used to extract reads for our *PIK3CA* variant of interest. Genotyping calls for the variant were then added as a metadata column to the single cell Seurat object.

Analysis of 3′-based single cell RNA-sequencing

Data preprocessing, including read alignment to the GRCh38 reference transcriptome, filtering, barcode counting, and unique molecular identifier counting, were performed using 10× Genomics CellRanger v.6.0 software following the default parameters for the 'count' pipeline. The resulting count files were input into Seurat version 4 using R version 4.2.1 and analyzed as follows[57,58]. Only cells with mitochondrial RNA < 10%; features > 500; RNA counts > 1000 and < 100,000 were included for downstream analysis. Initial analysis revealed a prominent role of cell cycle genes in the first principal component, so cell cycle phase assignment was performed using the CellCycleScoring function and the cell-cycle difference was calculated as the difference between the S-phase and the G2M phase scores. Normalization, identification of variable features, and data scaling were performed using SCTransform. To eliminate biased variance, cell cycle phases and mitochondrial reads were regressed out. Principal component analysis and dimensionality reduction using RunPCA and RunUMAP. Both reverse and forward PCA were performed to permit mapping to a cell reference. Clustering and differential gene expression were performed using FindNeighbors and FindAllMarkers. A publicly available skin reference dataset generated from the dermis and epidermis of human hip, palm, and sole (GSE20232) was downloaded and processed as above to generate a reference datasest of normal human skin. An anchor dataset was generated using FindTransferAnchors and cultured cells were mapped to the reference using TransferData. The R Bioconductor packages msigdbr, clusterProfiler, and fgsea were utilized to perform gene set enrichment analysis of the clusters and PCA component spaces to identified enriched Biological Processes (GO:BP) from the GeneOntology database. The R Bioconductor package CellChat was utilized to analyze receptor-ligand interactions between clusters[59].

Statistics and reproducibility

The FindAllMarkers feature in Seurat was used to perform differential gene expression with the default Wilcoxon Rank Sum Test. A min.pct of 0.25 and a log fold change threshold of 0.25 was used. An adjusted p-value of < 0.05 was used to identify significantly enriched Biological Processes (GO:BP) from the GeneOntology database. Pearson's Chi-squared test was performed to determine significance between cell clusters and cell cycle phases of $PIK3CA^{wt}$ and $PIK3CA^{mut}$. Enrichment scores of receptor-ligand interactions between clusters were calculated using the CellChat R Bioconductor package which combines cell–cell communication analysis with differential gene expression analysis. Due to the singular nature of the clinical skin biopsy sample, exome, Sanger, 3′ single cell, and long-read single cell sequencing was performed as an N of 1. Additionally, since each cell is assumed unique in single cell sequencing, it is not conventional to have biological replicates.

Data availability

Single cell data is available for browsing at https://www.nationwidechildrens.org/specialties/institute-for-genomic-medicine/ped-cell-browser. Raw and processed data for single cell RNA-sequencing has been deposited to NCBI Gene Expression Omnibus (GSE253153) and for exome sequencing has been deposited to dbGaP under accession number phs001820.v1.p1.

Code availability

All R code is provided as a supplemental file 'PIK3CA_R_code.doc'.

Received: 7 March 2024; Accepted: 4 September 2024
Published online: 25 October 2024

References

1. Cottrell, C. E. *et al.* Somatic PIK3R1 variation as a cause of vascular malformations and overgrowth. *Genet. Med.* **23**(10), 1882–1888 (2021).
2. Davis, S. *et al.* Growth hormone deficiency in megalencephaly-capillary malformation syndrome: An association with activating mutations in PIK3CA. *Am. J. Med. Genet. A* **182**(1), 162–168 (2020).
3. McNulty, S. N. *et al.* Diagnostic utility of next-generation sequencing for disorders of somatic mosaicism: A five-year cumulative cohort. *Am. J. Hum. Genet.* **105**(4), 734–746 (2019).
4. Park, H. J. *et al.* Detailed analysis of phenotypes and genotypes in megalencephaly-capillary malformation-polymicrogyria syndrome caused by somatic mosaicism of PIK3CA mutations. *Orphanet. J. Rare Dis.* **15**(1), 205 (2020).
5. Riviere, J. B. *et al.* De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat. Genet.* **44**(8), 934–940 (2012).
6. Zenner, K. *et al.* Somatic activating BRAF variants cause isolated lymphatic malformations. *HGG Adv.* **3**(2), 100101 (2022).
7. Peyre, M. *et al.* Somatic PIK3CA mutations in sporadic cerebral cavernous malformations. *N. Engl. J. Med.* **385**(11), 996–1004 (2021).
8. Weng, J. *et al.* Somatic MAP3K3 mutation defines a subclass of cerebral cavernous malformation. *Am. J. Hum. Genet.* **108**(5), 942–950 (2021).
9. Nikolaev, S. I., Fish, J. E. & Radovanovic, I. Somatic activating KRAS mutations in arteriovenous malformations of the brain. *N. Engl. J. Med.* **378**(16), 1561–1562 (2018).
10. Priemer, D. S. *et al.* Activating KRAS mutations in arteriovenous malformations of the brain: Frequency and clinicopathologic correlation. *Hum. Pathol.* **89**, 33–39 (2019).
11. Deleu, T., Jansen, K. & Calenbergh, F. V. Brain overgrowth associated with megalencephaly-capillary malformation syndrome causing progressive Chiari and syringomyelia. *Surg. Neurol. Int.* **13**, 211 (2022).
12. Loughan, A. R., Harrell, M., Perna, R., Allen, A. & Suddarth, B. Megalencephaly-capillary malformation polymicrogyria: A review and complex pediatric case report. *Appl. Neuropsychol. Child.* **6**(4), 369–377 (2017).
13. Mirzaa, G. M. *et al.* Megalencephaly-capillary malformation (MCAP) and megalencephaly-polydactyly-polymicrogyria-hydrocephalus (MPPH) syndromes: Two closely related disorders of brain overgrowth and abnormal brain and body morphogenesis. *Am. J. Med. Genet. A* **158A**(2), 269–291 (2012).
14. Mirzaa, G. M. & Dobyns, W. B. The "megalencephaly-capillary malformation" (MCAP) syndrome: The nomenclature of a highly recognizable multiple congenital anomaly syndrome. *Am. J. Med. Genet. A* **161A**(8), 2115–2116 (2013).
15. Sarma, K., Nayak, M. K., Mishra, B. & Gaikwad, S. B. Megalencephaly-capillary malformation-polymicrogyria syndrome (MCAP): A rare dynamic genetic disorder. *Cureus* **14**(5), e25123 (2022).
16. St John, L. J. & Rao, N. Autism spectrum disorder in a child with megalencephaly-capillary malformation-polymicrogyria syndrome (MCAP). *BMJ Case Rep.* **14**(12), e247034 (2021).
17. Edwards, E. A. *et al.* Monitoring arteriovenous malformation response to genotype-targeted therapy. *Pediatrics* **146**, 3 (2020).
18. Lekwuttikarn, R., Lim, Y. H., Admani, S., Choate, K. A. & Teng, J. M. C. Genotype-guided medical treatment of an arteriovenous malformation in a child. *JAMA Dermatol.* **155**(2), 256–257 (2019).
19. Li, D. *et al.* Genomic profiling informs diagnoses and treatment in vascular anomalies. *Nat. Med.* **29**(6), 1530–1539 (2023).
20. Venot, Q. *et al.* Targeted therapy in patients with PIK3CA-related overgrowth syndrome. *Nature* **558**(7711), 540–546 (2018).
21. Venot, Q. *et al.* Author Correction: Targeted therapy in patients with PIK3CA-related overgrowth syndrome. *Nature* **568**(7752), E6 (2019).
22. Chen, W. L. *et al.* The utility of cerebrospinal fluid-derived cell-free DNA in molecular diagnostics for the PIK3CA-related megalencephaly-capillary malformation (MCAP) syndrome: A case report. *Cold Spring Harb. Mol. Case Stud.* **8**, 3 (2022).
23. Zenner, K. *et al.* Cell-free DNA as a diagnostic analyte for molecular diagnosis of vascular malformations. *Genet. Med.* **23**(1), 123–130 (2021).
24. Palmieri, M. *et al.* Cell-free DNA next-generation sequencing liquid biopsy as a new revolutionary approach for arteriovenous malformation. *JVS Vasc. Sci.* **1**, 176–180 (2020).
25. Chang, F. *et al.* Molecular diagnosis of mosaic overgrowth syndromes using a custom-designed next-generation sequencing panel. *J. Mol. Diagn.* **19**(4), 613–624 (2017).
26. Kuentz, P. *et al.* Molecular diagnosis of PIK3CA-related overgrowth spectrum (PROS) in 162 patients and recommendations for genetic testing. *Genet. Med.* **19**(9), 989–997 (2017).
27. Morita, H. & Komuro, I. Somatic activating KRAS mutations in arteriovenous malformations of the brain. *N. Engl. J. Med.* **378**(16), 1561 (2018).
28. Nam, A. S. *et al.* Somatic mutations and cell identity linked by genotyping of transcriptomes. *Nature* **571**(7765), 355–360 (2019).
29. Giustacchini, A. *et al.* Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.* **23**(6), 692–702 (2017).
30. Cheow, L. F. *et al.* Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat. Methods* **13**(10), 833–836 (2016).
31. Al'Khafaji, A. M. *et al.* High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat. Biotechnol.* **42**, 582 (2023).
32. Li, M. M. *et al.* Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J. Mol. Diagn.* **19**(1), 4–23 (2017).
33. Wang, S. *et al.* Single cell transcriptomics of human epidermis identifies basal stem cell transition states. *Nat. Commun.* **11**(1), 4239 (2020).
34. Nguyen, B. C. *et al.* Cross-regulation between Notch and p63 in keratinocyte commitment to differentiation. *Genes Dev.* **20**(8), 1028–1042 (2006).
35. Zhou, B. *et al.* Notch signaling pathway: Architecture, disease, and therapeutics. *Signal Transduct. Target Ther.* **7**(1), 95 (2022).
36. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**(5670), 554 (2004).
37. Janku, F. *et al.* PIK3CA mutations in advanced cancers: Characteristics and outcomes. *Oncotarget* **3**(12), 1566–1575 (2012).
38. Garcia, Z., Kumar, A., Marques, M., Cortes, I. & Carrera, A. C. Phosphoinositide 3-kinase controls early and late events in mammalian cell division. *EMBO J.* **25**(4), 655–661 (2006).
39. Zhao, L. & Vogt, P. K. Helical domain and kinase domain mutations in p110alpha of phosphatidylinositol 3-kinase induce gain of function by different mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **105**(7), 2652–2657 (2008).
40. Zhao, L. & Vogt, P. K. Hot-spot mutations in p110alpha of phosphatidylinositol 3-kinase (pI3K): Differential interactions with the regulatory subunit p85 and with RAS. *Cell Cycle* **9**(3), 596–600 (2010).
41. Shioi, T. *et al.* The conserved phosphoinositide 3-kinase pathway determines heart size in mice. *EMBO J.* **19**(11), 2537–2548 (2000).
42. Fingar, D. C., Salama, S., Tsou, C., Harlow, E. & Blenis, J. Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. *Genes Dev.* **16**(12), 1472–1487 (2002).
43. Saucedo, L. J. & Edgar, B. A. Why size matters: Altering cell size. *Curr. Opin. Genet. Dev.* **12**(5), 565–571 (2002).
44. Wiedemann, J. *et al.* Differential cell composition and split epidermal differentiation in human palm, sole, and hip skin. *Cell Rep.* **42**(1), 111994 (2023).

45. Medic, S. & Ziman, M. PAX3 expression in normal skin melanocytes and melanocytic lesions (naevi and melanomas). *PLoS ONE* **5**(4), e9977 (2010).
46. Nakazaki, H. *et al.* Key basic helix-loop-helix transcription factor genes Hes1 and Ngn2 are regulated by Pax3 during mouse embryonic development. *Dev. Biol.* **316**(2), 510–523 (2008).
47. Luo, W. *et al.* Arterialization requires the timely suppression of cell growth. *Nature* **589**(7842), 437–441 (2021).
48. Kofler, N. M., Cuervo, H., Uh, M. K., Murtomaki, A. & Kitajewski, J. Combined deficiency of Notch1 and Notch3 causes pericyte dysfunction, models CADASIL, and results in arteriovenous malformations. *Sci. Rep.* **5**, 16449 (2015).
49. Hasan, S. S. *et al.* Endothelial Notch signalling limits angiogenesis via control of artery formation. *Nat. Cell Biol.* **19**(8), 928–940 (2017).
50. Conway, E. M., Collen, D. & Carmeliet, P. Molecular mechanisms of blood vessel growth. *Cardiovasc. Res.* **49**(3), 507–521 (2001).
51. Adams, R. H. & Alitalo, K. Molecular regulation of angiogenesis and lymphangiogenesis. *Nat. Rev. Mol. Cell Biol.* **8**(6), 464–478 (2007).
52. Kofler, N. M. *et al.* Notch signaling in developmental and tumor angiogenesis. *Genes Cancer* **2**(12), 1106–1116 (2011).
53. Tefft, J. B. *et al.* Notch1 and Notch3 coordinate for pericyte-induced stabilization of vasculature. *Am. J. Physiol. Cell Physiol.* **322**(2), C185–C196 (2022).
54. Domenga, V. *et al.* Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev.* **18**(22), 2730–2735 (2004).
55. Battegay, E. J., Rupp, J., Iruela-Arispe, L., Sage, E. H. & Pech, M. PDGF-BB modulates endothelial proliferation and angiogenesis in vitro via PDGF beta-receptors. *J. Cell Biol.* **125**(4), 917–928 (1994).
56. Kelly, B. J. *et al.* Churchill: An ultra-fast, deterministic, highly scalable and balanced parallelization strategy for the discovery of human genetic variation in clinical and population-scale genomics. *Genome Biol.* **16**(1), 6 (2015).
57. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**(5), 411–420 (2018).
58. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**(7), 1888–1902 (2019).
59. Jin, S. *et al.* Inference and analysis of cell–cell communication using cell chat. *Nat. Commun.* **12**(1), 1088 (2021).

Acknowledgements

We are grateful for the patient and family who participated in this study. The schematics in Fig. 1d,e were generated using biorender.com.

Author contributions

Conceptualization: CEC, ARM, KEM. Methodology: CEC, ARM, KEM. Software: MAW, JJW, KEM. Formal analysis: MAW, JJW. Investigation: MAW, JJW, JBN, MEHG. Resources: EAV, PW, ERM, CEC, PW. Data curation: MAW, JJW, TD, EARG, PW. Visualization: MAW, EARG. Writing—Original draft: MAW. Writing—Review & editing: All authors. Supervision: ARM, KEM. Project administration: CEC, ARM, KEM. Funding acquisition: CEC, ERM.

Funding

This work was supported by the Nationwide Foundation Pediatric Innovation Fund.

Competing interests

The authors declare no competing interests.

Informed consent

Written informed consent was obtained from the patient's parents under a research protocol approved by the Institutional Review Board at Nationwide Children's Hospital (IRB17-00206).

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-72167-8.

Correspondence and requests for materials should be addressed to K.E.M.