

## Candidate Predisposition Variants in Kaposi Sarcoma as Detected by Whole-Genome Sequencing

Sanni J. Rinne,<sup>1,2</sup> Lauri J. Sipilä,<sup>1,2</sup> Päivi Sulo,<sup>1,2</sup> Emmanuelle Jouanguy,<sup>3,4,5</sup> Vivien Béziat,<sup>3,4,5</sup> Laurent Abel,<sup>3,4,5</sup> Jean-Laurent Casanova,<sup>3,4,5,6,7</sup> Nima Parvaneh,<sup>8,9</sup> Kamran Balighi,<sup>10,11</sup> Emma Guttman-Yassky,<sup>12,13</sup> Ronit Sarid,<sup>14</sup> Lauri A. Aaltonen,<sup>1,2</sup> and Mervi Aavikko<sup>1,2</sup>

<sup>1</sup>Applied Tumor Genomics Research Program and <sup>2</sup>Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland; <sup>3</sup>Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale (INSERM) UMR-1163, Paris, France; <sup>4</sup>University Paris Descartes, Imagine Institute, Paris, France; <sup>5</sup>St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, New York; <sup>6</sup>Howard Hughes Medical Institute, New York, New York; <sup>7</sup>Pediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, Paris, France; <sup>8</sup>Division of Allergy and Clinical Immunology, Department of Pediatrics, <sup>9</sup>Research Center for Immunodeficiencies, <sup>10</sup>Department of Dermatology, Razi Hospital, and <sup>11</sup>Autoimmune Bullous Diseases Research Center, Tehran University of Medical Sciences, Tehran, Iran; <sup>12</sup>Department of Dermatology, and <sup>13</sup>Laboratory for Inflammatory Skin Diseases, Icahn School of Medicine at Mount Sinai, New York, New York; <sup>14</sup>The Mina and Everard Goodman Faculty of Life Sciences & Advanced Materials and Nanotechnology Institute, Bar-Ilan University, Ramat-Gan, Israel

Familial clustering of classic Kaposi sarcoma (CKS) is rare with, approximately 100 families reported to date. We studied 2 consanguineous families, 1 Iranian and 1 Israeli, with multiple cases of adult CKS and without overt underlying immunodeficiency. We performed genome-wide linkage analysis and whole-genome sequencing to discover the putative genetic cause for predisposition. A 9-kb homozygous intronic deletion in *RP11-259O2.1* in the Iranian family and 2 homozygous variants, 1 in *SCUBE2* and the other in *CDHR5*, in the Israeli family were identified as possible candidates. The presented variants provide a robust starting point for validation in independent samples.

**Keywords.** classic Kaposi sarcoma; *CDHR5*; genetic linkage; genetic predisposition; *RP11-259O2.1*; *SCUBE2*; whole-genome sequencing.

Kaposi sarcoma (KS) is an inflammatory soft tissue tumor that typically presents as purple lesions of skin and mucosa, but it can also spread to lymph nodes and other organs. It is caused by an infection with Kaposi sarcoma-associated herpesvirus (KSHV), also known as human herpesvirus 8 (HHV-8). The

infection is necessary, albeit not alone sufficient for the onset of the disease. Therefore, other contributing factors are required, but despite a considerable amount of research, the precise etiology of this complex disease remains poorly understood.

Kaposi sarcoma has 4 widely recognized clinico-epidemiological subtypes. They are AIDS-associated Kaposi sarcoma, which affects individuals with HIV infection, the iatrogenic type, which affects patients undergoing immunosuppressive therapy, the African endemic type, which is most common in parts of Central and Eastern Africa, and classic Kaposi sarcoma (CKS), which most often affects elderly males of Mediterranean or Eastern European Jewish backgrounds. The age of onset is usually >65, and only a fraction of the cases are observed in individuals aged <50 [1]. CKS is a slowly progressing disease, where the lesions occur most often on the lower extremities and are prone to enlarge and spread on the skin.

In rare occasions, CKS clusters in families. To date, approximately 100 families have been reported, and the age of onset is often lower than in nonfamilial CKS. Studies on these families have demonstrated, for example, association of certain HLA haplotypes with the disease [2]; however, contradictory results have also been obtained [3]. In addition, different genetic variants of KSHV, KSHV DNA load, and other contributing infections have been studied [4].

We here studied 2 large consanguineous families of Middle Eastern origin, 1 from Iran (hereafter referred to as KapoIran) and the other from Israel (hereafter referred to as KapoIsrael). The seroprevalence of KSHV in Israel is 5%–20% [5], and in Iran it is approximately 2% among the general population [6]. We used genome-wide linkage analysis and whole-genome sequencing to identify the possible genetic causes of CKS predisposition in the 2 families.

## METHODS

### Ethical Considerations

The study was approved by the National Institute for Health and Welfare (THL; 151/5.05.00/2017) and the local ethics committee (HUS; 408/13/03/03/09). A written and informed consent was obtained from the family members before the study.

### Study Families

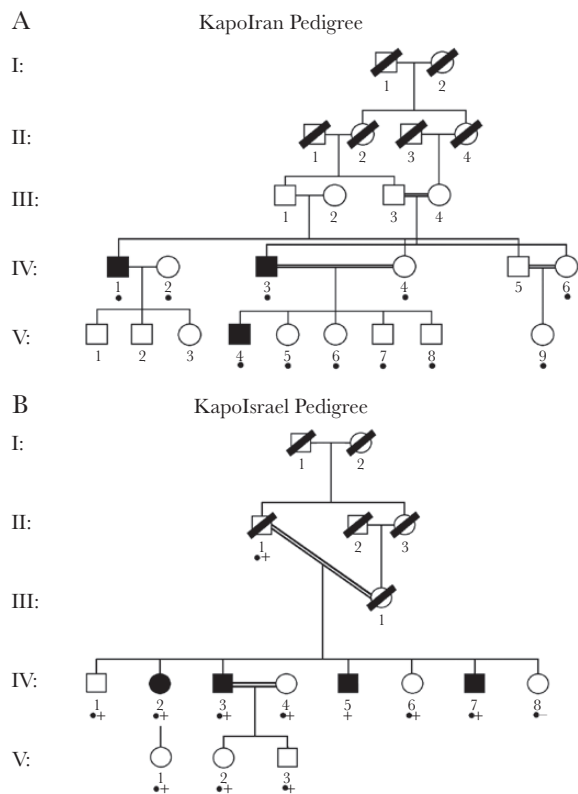
The KapoIran family had 3 affected males in 2 generations (IV-1, IV-3, and V-4) (Figure 1A). Only limited background information on the family and medical histories of the patients was available. The affected were all diagnosed at approximately 40 years of age. To our knowledge, the symptoms were typical for CKS; however, a few years after the CKS diagnosis, IV-1 was also diagnosed with cutaneous T-cell lymphoma. In the

Received 23 May 2019; editorial decision 6 July 2019; accepted 16 July 2019.

Correspondence: M. Aavikko, PhD, Applied Tumor Genomics Research Program, Department of Medical and Clinical Genetics, Medicum, Biomedicum Helsinki, Room B519b2, PO Box 63 (Haartmaninkatu 8), University of Helsinki, FI-00014 Helsinki, Finland (mervi.aavikko@helsinki.fi).

### Open Forum Infectious Diseases®

© The Author(s) 2019. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com DOI: 10.1093/ofid/ofz337



**Figure 1.** Pedigrees of the KapoIran (A) and the KapoIsrael (B) families. The generation number and the individual number in that generation (eg, III-4) identify the family members. Black color (square for male, circle for female) indicates the individuals affected with classic Kaposi sarcoma. Double line marks intrafamilial marriage. Black diagonal line over the symbol marks the deceased. The individuals with a sample available are marked with a black dot under their identification as well as the seropositivity status (+/-) if known. The KapoIsrael pedigree has previously been reported by Guttman-Yassky and colleagues (2004).

KapoIsrael family (Figure 1B), there were 4 affected siblings, 3 males (IV-3, IV-5, and IV-7) and 1 female (IV-2), though no sample was available for 1 of them (IV-5). Their clinical features have been reported previously [4]. The family is of Jewish origin and in addition to the affected individuals, positive KSHV infection statuses have been established in 7 of the 8 unaffected family members (II-1, IV-1, IV-4, IV-6, V-1, V-2, and V-3). The diagnosis ages were 53, 45, 41, and 38 for IV-2, IV-3, IV-5, and IV-7, respectively. DNA samples of 11 individuals in each family were available for the study. Details of the methods are described in the Supplementary Methods.

#### Whole-Genome Sequencing

Two affected family members from both families were whole-genome-sequenced (KapoIran IV-3, KapoIran V-4, KapoIsrael IV-3, and KapoIsrael IV-7). The whole-genome libraries were prepared with the Kapa HyperPrep Kit (Roche, Basel, Switzerland). Paired-end sequencing with 75 base pair reads was performed using Illumina HiSeq4000 (KapoIran) and Illumina HiSeq2000 (KapoIsrael) according to the manufacturer's

instructions. Key measures of the whole-genome sequencing data are presented in Supplementary Table 1.

#### Linkage Analysis

Linkage analyses were conducted with MERLIN, version 1.1.2. The families were analyzed separately, and in both analyses, a recessive inheritance model was assumed using a 0.0001 disease allele frequency and a penetrance of 80%. The available KapoIran family members (Supplementary Methods) were genotyped using Affymetrix (Thermo Fisher Scientific Inc., Santa Clara, CA) Genome-Wide Human SNP Array 6.0 and Affymetrix GeneChip Human Mapping 250K Nsp Array. For the analysis, the KapoIran family was split in 2 (Supplementary Figure 1). In the KapoIsrael family, whole-genome sequencing data were used for genome-wide genotyping of the 2 affected individuals (IV-3 and IV-7). Variants with a minor allele frequency  $\geq 0.05$  in gnomAD (r2.0.1) were genotyped using the UnifiedGenotyper tool from the GATK toolkit (version 3.5) and used for subsequent linkage analysis.

#### Candidate Variant Analysis

The candidate variant analysis focused on genetic alterations shared by the affected individuals in each family. BasePlayer software (version 1.0.2) [7] was used to study the shared homozygous and compound heterozygous single nucleotide variants (SNVs) as well as small insertions and deletions (indels). Structural variants (SVs), including deletions, tandem duplications, inversions, and translocations, were identified with DELLY, version 0.0.9.

The analysis focused on variants locating within the linkage-compatible regions and with a minimum supporting read count of 4 to take into account the relatively low mean coverage of the whole-genome sequencing and to minimize the number of false positives. All variants were analyzed visually and quality-wise to exclude mapping artefacts. SNVs and small indels with a minor allele frequency  $\geq 0.005$  were not considered further. Additional filtering was performed using different in-house control sets, variant databases, and prediction algorithms. The full list of control sets is described in the Supplementary Methods. The remaining candidate variants were validated, and their segregation with the disease was studied by polymerase chain reaction (PCR) and Sanger sequencing (details in the Supplementary Methods).

## RESULTS

We used linkage analysis to map the candidate genomic regions potentially harboring susceptibility variants. In the case of KapoIran, the focus on genomic areas with positive logarithm of the odds (LOD) scores resulted in detection of 2 regions (Supplementary Table 2). Because the linkage analysis in the KapoIsrael family was based on the genotypes of the 2 affected individuals only, we restricted our analysis to the genomic regions with a positive LOD

score and  $\geq 5$  cM in length, as these regions were more likely to harbor the predisposition gene. That resulted in a total of 33 regions (Supplementary Table 2). The longest linkage-compatible regions in the KapoIran and KapoIsrael analyses were a 10.5-Mb (25.5-cM) region in chromosome 5 and a 82-Mb (79-cM) region in chromosome 1, respectively. No overlapping chromosomal regions were identified in the 2 families. As each family had relatively few affected individuals, the maximum LOD scores could not reach formal statistical significance (max. LOD, 1.598 KapoIran and 0.602 KapoIsrael).

To identify the genetic variants located in the candidate regions, we analyzed whole-genome sequences of 2 affected individuals from both families (KapoIran IV-3, KapoIran V-4, KapoIsrael IV-3, and KapoIsrael IV-7). After identifying the shared variants among respective individuals, we established 4 sets of candidate variants: homozygous protein coding, compound heterozygous protein-coding, homozygous noncoding, and homozygous structural variants.

After filtering the data using a minor allele frequency threshold of 0.005, 39 structural variants and 6 homozygous noncoding variants remained in the KapoIran family. No coding variants met the criteria. The visual and quality-based evaluation of the SVs excluded all but 1 of them. The 1 remaining candidate variant was a homozygous deletion of 9005 base pairs between the bases 1 939 257 and 1 948 262 in chromosome 5 (GRCh37) within the intron of long intergenic noncoding RNA gene *RP11-25902.1* (Supplementary Figure 2).

In the KapoIsrael family, there were 10 homozygous protein-coding, 62 homozygous noncoding, and 201 structural candidate variants that located within the linkage-compatible regions. Segregation analysis and SV quality analysis excluded 8 of the protein-coding variants and all SVs. The 2 protein-coding variants that remained were c.1489G > A (p.Pro497Ser) in *SCUBE2* (Signal Peptide, CUB Domain and EGF-like Domain Containing 2) and c.2269C > T (p. Gly757Ser) in *CDHR5* (Cadherin Related Family Member 5) (Supplementary Figure 3). Both were predicted to be either benign or tolerated by Polyphen-2 and SIFT, respectively, although the *SCUBE2* variant was predicted to be damaging by DANN scoring. For more detailed data on the variants, see Supplementary Table 3.

The candidate variants were validated with PCR and Sanger sequencing, and they segregated with CKS in the respective families (Table 1). Finally, we examined the noncoding variants. None of the variants located at any obvious splice site regions were subsequently annotated using the ENCODE data of known regulatory regions. The noncoding variants that located at relevant ENCODE annotated regions are listed in Supplementary Table 4.

## DISCUSSION

The etiology of CKS is still poorly understood. According to population-based studies, only about 1 in 10 000 KSHV-infected

individuals develops classic Kaposi sarcoma [1,8–10] which suggests that other contributing factors exist besides the KSHV infection. The disease is rarely familial, but the existing families provide an excellent opportunity for susceptibility studies. In previous studies, there has been evidence of genetic susceptibility in CKS. Mutations in genes *WAS*, *IFNGR1*, *STIM1*, and *TNFRSF4* (OX40) have been found to cause childhood-onset CKS (reviewed in [11]). In adult-onset CKS, *STAT4* has been suggested as a candidate predisposing gene [12]. All of these genes have functions in adaptive immunity, more specifically in T-cell function.

In this study, we performed a genome-wide candidate variant analysis of 2 families with adult CKS using whole-genome sequencing and linkage analysis. Both families were consanguineous and of Middle Eastern origin and were compatible with a recessive inheritance pattern; however, no overlapping candidate genomic regions were identified, suggesting different roots for the genetic susceptibility.

In the KapoIran family, the most promising susceptibility variant was a homozygous deletion of 9005 base pairs in the intron of long intergenic noncoding RNA (lincRNA) gene *RP11-25902.1*. The deletion was unique to the family, and we did not identify even heterozygous carriers of it in any of the control individuals studied. Long intergenic noncoding RNAs are long noncoding RNAs (lncRNA) located in between protein coding genes. Some lncRNAs have been reported to act in carcinogenesis either as oncogenes promoting cell proliferation or as tumor suppressors [13]. There is also evidence showing that lncRNA dysregulation can contribute to a variety of complex human diseases, including autoimmune diseases such as autoimmune thyroid disease and psoriasis [13]. Additionally, a majority of lncRNAs have been shown to be very tissue-specific [14]. Based on GTEx expression data, *RP11-25902.1* has the highest expression in skin and other epithelial cells (Supplementary Figure 4), which is intriguing considering that CKS usually forms lesions on the skin of the lower limbs.

In the KapoIsrael family, 2 candidate variants emerged. Perhaps the most promising is c.1489G > A (p.Pro497Ser) in *SCUBE2*. *SCUBE2* is a member of a small *SCUBE* gene family, other members being *SCUBE1* and *SCUBE3*. According to some studies, *SCUBE2* has functions in carcinogenesis by being able to inhibit tumor cell growth, invasion, and proliferation [15]. It has also been shown to be expressed in vascular endothelial cells and to interact with vascular endothelial growth factor (VEGF) [16]. Angiogenesis is an important feature in the formation of Kaposi sarcoma lesions, and the tumors overexpress VEGF [17]. Based on GTEx data, *SCUBE2* is most expressed in bladder, prostate, and breast tissues (Supplementary Figure 4).

The other candidate gene *CDHR5*, also known as Mucin and cadherin-like (*MUCDHL*), is a member of the cadherin superfamily. Members of this family encode  $\text{Ca}^{2+}$ -dependent membranous glycoproteins that are important in cell–cell adhesions

**Table 1. Candidate Predisposition Variant Segregation in Kapoiran and Kapolsrael Family; Genotypes of the Iranian and Israeli Family**

Individual	<i>NKD2</i>	<i>RP11-259O2.1</i>	<i>CDHR5</i>	<i>SCUBE2</i>
<b>Kapolran IV-1</b>	HO	HO	...	...
Kapolran IV-2	WT	WT	...	...
<b>Kapolran IV-3</b>	HO	HO	...	...
Kapolran IV-4	HO	HO	...	...
Kapolran IV-6	HE	HE	...	...
<b>Kapolran V-4</b>	HO	HO	...	...
Kapolran V-6	HO	HO	...	...
Kapolran V-7	HO	HO	...	...
Kapolsrael II-1	...	...	HE	HE
Kapolsrael IV-1	...	...	HE	HE
<b>Kapolsrael IV-2</b>	...	...	HO	HO
<b>Kapolsrael IV-3</b>	...	...	HO	HO
Kapolsrael IV-4	...	...	WT	WT
Kapolsrael IV-6	...	...	WT	WT
<b>Kapolsrael IV-7</b>	...	...	HO	HO
Kapolsrael IV-8	...	...	HO	HO
Kapolsrael V-1	...	...	HE	HE
Kapolsrael V-2	...	...	HE	HE
Kapolsrael V-3	...	...	HE	HE

The affected individuals are marked with bold font.

Abbreviations: HO, homozygous; WT, wild-type; HE, heterozygous.

and regulate multiple different processes including cell proliferation and differentiation. *CDHR5* has been shown to interact with  $\beta$ -catenin and act as a growth inhibitor in colon cancer cells [18]. The variant shared by affected family members was predicted benign by all prediction programs that we used. Based on GTEx data, *CDHR5* is most expressed in the intestine and liver (Supplementary Figure 4).

We carried out a thorough genome-wide variant analysis in 2 large Kaposi sarcoma families. We focused on rare genetic alterations that were compatible with monogenic recessive inheritance, though a possibility remains that the predisposition is multifactorial or perhaps simply more common in the general population. We hypothesized that the 2 families may have shared a common genetic predisposition or genetic defect in the same signaling pathway, but this did not seem to be the case, and the possible genetic susceptibility in the families appears to have different genetic roots. We found 3 plausible susceptibility variants that all segregated with CKS in the families. Of these, variants in *SCUBE2* and *RP11-259O2.1* were the most promising due to known functions and expression in the target tissues, respectively. The candidate variants identified in this study provide a platform for validation in extended sample sets. If validated, the findings should provide opportunities for prevention and early detection in susceptible individuals as well as valuable clues to the molecular etiology of CKS.

### Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of

the authors, so questions or comments should be addressed to the corresponding author.

### Acknowledgments

This study could not have been done without the willingness of the families to participate in scientific research and for that, they are thanked. We thank the members of the Aaltonen Group, our international collaborators, and everyone who has been part of collecting the background information and producing previous research results on these 2 families. This study makes use of data generated by The Ashkenazi Genome Consortium.

**Financial support.** This study was supported by grant from the Academy of Finland's Center of Excellence Program 2012–2017, #250345, and 2018–2025, #312041.

**Potential conflicts of interest.** Potential conflicts of interest outside the submitted work: E.G.Y. reports grants from Dermavent, Innovaderm, Janssen Biotech, Novan, and Ralexar, personal fees from Allergan, Amgen, Concert, DBV, Dermira, Eli Lilly, EMD Serono, Escalier, Flx Bio, Kyowa Kirin, Mitsubishi Tanabe, and Sanofi, as well as both grants and personal fees from AbbVie, Asana Biosciences, Celgene, DS Biopharma, Galderma, Glenmark, LEO Pharmaceuticals, Novartis, Pfizer, Regeneron, and Union Therapeutics. J.L.C. reports personal fees from Celgene, Neovacs, ADMA, Nimbus, Vitae Pharmaceuticals, KymeraTX, Sanofi, Asahi Kasei, Pfizer, and Celgene. M.A. reports nonfinancial support from Roche Diagnostics Oy. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

### References

- Iscovich J, Boffetta P, Franceschi S, et al. Classic Kaposi sarcoma: epidemiology and risk factors. *Cancer* **2000**; 88:500–17.
- Masala MV, Carcassi C, Cottoni F, et al. Classic Kaposi's sarcoma in Sardinia: HLA positive and negative associations. *Int J Dermatol* **2005**; 44:743–5.
- Strichman-Almashanu L, Weltfriend S, Gideoni O, et al. No significant association between HLA antigens and classic Kaposi sarcoma: molecular analysis of 49 Jewish patients. *J Clin Immunol* **1995**; 15:205–9.
- Guttman-Yassky E, Cohen A, Kra-Oz Z, et al. Familial clustering of classic Kaposi sarcoma. *J Infect Dis* **2004**; 189:2023–6.

5. Davidovici B, Karakis I, Bourboulia D, et al. Seroepidemiology and molecular epidemiology of Kaposi's sarcoma-associated herpesvirus among Jewish population groups in Israel. *J Natl Cancer Inst* **2001**; 93:194–202.
6. Jalilvand S, Shoja Z, Mokhtari-Azad T, et al. Seroprevalence of human herpesvirus 8 (HHV-8) and incidence of Kaposi's sarcoma in Iran. *Infect Agent Cancer* **2011**; 6:5.
7. Katainen R, Donner I, Cajuso T, et al. Discovery of potential causative mutations in human coding and noncoding genome with the interactive software BasePlayer. *Nat Protoc* **2018**; 13:2580–600.
8. Grulich AE, Beral V, Swerdlow AJ. Kaposi's sarcoma in England and Wales before the AIDS epidemic. *Br J Cancer* **1992**; 66:1135–7.
9. Hjalgrim H, Melbye M, Pukkala E, et al. Epidemiology of Kaposi's sarcoma in the Nordic countries before the AIDS epidemic. *Br J Cancer* **1996**; 74:1499–502.
10. Geddes M, Franceschi S, Barchielli A, et al. Kaposi's sarcoma in Italy before and after the AIDS epidemic. *Br J Cancer* **1994**; 69:333–6.
11. Jackson CC, Dickson MA, Sadjadi M, et al. Kaposi sarcoma of childhood: inborn or acquired immunodeficiency to oncogenic HHV-8. *Pediatr Blood Cancer* **2016**; 63:392–7.
12. Aavikko M, Kaasinen E, Nieminen JK, et al. Whole-genome sequencing identifies STAT4 as a putative susceptibility gene in classic Kaposi sarcoma. *J Infect Dis* **2015**; 211:1842–51.
13. Li J, Xuan Z, Liu C. Long non-coding RNAs and complex human diseases. *Int J Mol Sci* **2013**; 14:18790–808.
14. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **2011**; 25:1915–27.
15. Song Q, Li C, Feng X, et al. Decreased expression of SCUBE2 is associated with progression and prognosis in colorectal cancer. *Oncol Rep* **2015**; 33:1956–64.
16. Lin YC, Chao TY, Yeh CT, et al. Endothelial SCUBE2 interacts with VEGFR2 and regulates VEGF-induced angiogenesis. *Arterioscler Thromb Vasc Biol* **2017**; 37:144–55.
17. Kerr DA, Busarla SVP, Gimbel DC, et al. mTOR, VEGF, PDGFR, and c-kit signaling pathway activation in Kaposi sarcoma. *Hum Pathol* **2017**; 65:157–65.
18. Hinkel I, Duluc I, Martin E, et al. Cdx2 controls expression of the protocadherin Mucdhl, an inhibitor of growth and  $\beta$ -catenin activity in colon cancer cells. *Gastroenterology* **2012**; 142:875–85.e3.