

Germline Heterozygous Variants in *SEC23B* Are Associated with Cowden Syndrome and Enriched in Apparently Sporadic Thyroid Cancer

Lamis Yehia,^{1,2,3} Farshad Niazi,^{1,2} Ying Ni,^{1,2,4,11} Joanne Ngeow,^{1,2,5,6} Madhav Sankunny,^{1,2} Zhigang Liu,^{1,2} Wei Wei,^{1,2,7} Jessica L. Mester,^{1,2} Ruth A. Keri,^{8,9} Bin Zhang,^{1,2} and Charis Eng^{1,2,10,11,12,*}

Cancer-predisposing genes associated with inherited cancer syndromes help explain mechanisms of sporadic carcinogenesis and often inform normal development. Cowden syndrome (CS) is an autosomal-dominant disorder characterized by high lifetime risks of epithelial cancers, such that ~50% of affected individuals are wild-type for known cancer-predisposing genes. Using whole-exome and Sanger sequencing of a multi-generation CS family affected by thyroid and other cancers, we identified a pathogenic missense heterozygous *SEC23B* variant (c.1781T>G [p.Val594Gly]) that segregates with the phenotype. We also found germline heterozygous *SEC23B* variants in 3/96 (3%) unrelated mutation-negative CS probands with thyroid cancer and in The Cancer Genome Atlas (TCGA), representing apparently sporadic cancers. We note that the TCGA thyroid cancer dataset is enriched with unique germline deleterious *SEC23B* variants associated with a significantly younger age of onset. *SEC23B* encodes Sec23 homolog B (*S. cerevisiae*), a component of coat protein complex II (COPII), which transports proteins from the endoplasmic reticulum (ER) to the Golgi apparatus. Interestingly, germline homozygous or compound-heterozygous *SEC23B* mutations cause an unrelated disorder, congenital dyserythropoietic anemia type II, and *SEC23B*-deficient mice suffer from secretory organ degeneration due to ER-stress-associated apoptosis. By characterizing the p.Val594Gly variant in a normal thyroid cell line, we show that it is a functional alteration that results in ER-stress-mediated cell-colony formation and survival, growth, and invasion, which reflect aspects of a cancer phenotype. Our findings suggest a different role for *SEC23B*, whereby germline heterozygous variants associate with cancer predisposition potentially mediated by ER stress “addiction.”

Introduction

Cowden syndrome (CS [MIM: 158350]) is an underdiagnosed difficult-to-recognize autosomal-dominant disorder characterized by multiple hamartomas and an increased predisposition to breast, thyroid, endometrial, and other cancers. Other, non-neoplastic phenotypes include macrocephaly and mucocutaneous findings.^{1,2} The tumor-suppressor gene phosphatase and tensin homolog (*PTEN* [MIM: 601728]) was identified as the first susceptibility gene for CS.^{3,4} Earlier studies identified germline *PTEN* mutations in 85% of CS individuals^{1,5} meeting the strict International Cowden Consortium (ICC) operational diagnostic criteria, which select for familial cases and exaggerated phenotypes accrued from tertiary medical centers.^{6,7} Subsequent prospective community-based accrual of >3,000 probands over 12 years revealed that ~25% of individuals who met the diagnostic criteria harbored germline pathogenic *PTEN* mutations.⁸ Relatedly, we note that many individuals show clinical features reminiscent of CS but do not meet the ICC criteria. We refer to this phenotype as CS-like (CSL),

and these affected individuals show extensive phenotypic heterogeneity, and only 5% have germline *PTEN* mutations.¹

Relevant to clinical practice, we generated a predictive *PTEN* risk calculator⁸ that evaluates the pretest probability that an individual harbors a deleterious germline *PTEN* mutation on the basis of key clinical features.⁸ The resultant score is known as the *PTEN* Cleveland Clinic score, or CC score. A CC score of 10, corresponding to a pretest probability of >3% and a sensitivity of 90%, is recommended as a threshold for referring adult CS or CSL individuals for *PTEN* mutational testing and further genetic assessment. However, many CS and CSL individuals evaluated in clinic have high CC scores (range 10–54) and no underlying germline *PTEN* mutations or large deletions, indicating high phenotypic burden independent of *PTEN* and, hence, a high likelihood that other susceptibility genes exist. Indeed, we have so far reported that within this *PTEN*-mutation-negative subset of CS and CSL individuals, up to 8% harbor germline variants in mitochondrial complex II subunit genes *SDHB* (MIM: 185470), *SDHC* (MIM: 602413), and *SDHD* (MIM: 602690),^{9,10} up

¹Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH 44195, USA; ²Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA; ³Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA; ⁴Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA; ⁵Division of Medical Oncology, National Cancer Centre, Singapore 169610, Singapore; ⁶Oncology Academic Clinical Program, Duke-NUS Graduate Medical School, Singapore 169857, Singapore; ⁷Department of Pediatrics, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin 300020, China; ⁸Department of Pharmacology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA; ⁹Division of General Medical Sciences-Oncology, Case Western Reserve University, Cleveland, OH 44106, USA; ¹⁰Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH 44195, USA; ¹¹Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA; ¹²CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106, USA

*Correspondence: engc@ccf.org

<http://dx.doi.org/10.1016/j.ajhg.2015.10.001>. ©2015 by The American Society of Human Genetics. All rights reserved.

to 30% have promoter hypermethylation in tumor-suppressor gene *KLLN* (MIM: 612105),¹¹ and 11% have germline *AKT1* (MIM: 164730) or *PIK3CA* (MIM: 171834) gain-of-function mutations.¹²

Epithelial thyroid carcinomas (both papillary and follicular histologies) are a major component of CS.¹³ We have shown in a prospective series of 2,723 CS and CSL individuals that germline mutations in a subset of the established susceptibility genes (*PTEN*, *SDHB*, *SDHC*, *SDHD*, and *KLLN*) increase the risk of epithelial thyroid cancer by at least 45-fold in comparison to the risk in the US general population.¹⁴ Moreover, included in our series of CS individuals are *PTEN*-mutation-positive probands diagnosed under the age of 18 years (the earliest age at diagnosis was 7 years), reflecting a strong genetic component in the etiology of thyroid cancer. In support of these findings, epidemiological studies have reported that thyroid cancers show a high index of “familiality” among all solid tumors.^{15,16} However, the hunt for candidate genes associated with familial thyroid cancer continues today, and there is an increased appreciation for the heterogeneity of the disease and the contribution of hereditary predisposition and the micro- and macro-environment to tumor pathobiology.¹⁷ In this context, not only would additional CS-relevant genes serve as candidates for apparently sporadic predisposition to epithelial thyroid cancer, but their identification would also improve molecular diagnosis, especially predictive testing, cancer risk assessment, genetic counseling, and clinical management in CS and CSL. Therefore, we sought to identify previously undescribed cancer-predisposing gene(s) in CS and CSL via an approach combining exome sequencing and family studies and, subsequently, to characterize the functional consequence.

Material and Methods

Research Participants and Clinical Data

We scanned our clinical database for accrued CS- and CSL-affected families. Eligible families consisted of at least one unaffected and two affected family members. Eligible probands tested negative for germline mutations in *PTEN*, *SDHB*, *SDHC*, *SDHD*, *AKT1*, and *PIK3CA* and met at least the relaxed ICC operational diagnostic criteria (Table S1). Relaxed criteria are defined as full criteria minus one, and such individuals are referred to as having CSL.⁷

For all selected individuals, we reviewed the CC score, a semi-quantitative score that is based on weighting clinical features and that estimates the pretest probability of finding a germline *PTEN* mutation (see Web Resources).⁸ Given that all individuals are wild-type for *PTEN*, we used the CC score as a surrogate of phenotypic burden. We hence targeted probands with a high CC score and eligible family members with available blood samples for genetic testing. We reviewed medical records, including pathology reports, for each research participant and extracted family history from clinical notes associated with cancer genetics and/or genetic-counseling visits, where applicable and with the individuals' consent. Cleveland Clinic Institu-

tional Review Board approval (protocol 8458) and informed consent from all research participants were obtained for this study.

Exome Sequencing and Bioinformatic Analysis

We subjected germline genomic DNA extracted from peripheral-blood leukocytes of the eligible proband to exome sequencing. Exome enrichment was performed with the TruSeq SBS v.3 Kit (Illumina), and subsequent 100 bp paired-end sequencing was performed with an Illumina HiSeq 2000 platform. Sequencing was performed at an Illumina Sequencing Service Center. Raw sequencing reads were mapped to the human reference haploid genome sequence (UCSC Genome Browser hg19) with the Burrows-Wheeler Aligner (BWA v.0.6.1).¹⁸ Insertion or deletion (indel) realignment, base- and quality-score recalibrations, and removal of PCR duplicates from the resultant binary alignment map (BAM) files were performed with the Genome Analysis Toolkit (GATK),^{19,20} Sequence Alignment/Map (SAMtools), and Picard.²¹ Variant discovery and genotype calling of single-nucleotide variations (SNVs) and short (<50 bp) indels were performed with the GATK Haplotype Caller.

Variant Filtration and Annotation

To prioritize causal variants, we applied the ANNOVAR Variants Reduction pipeline.²² First, we discarded synonymous variants and intronic variants more than 2 bp from exon boundaries. We retained variants in conserved genomic regions on the basis of a 46-species alignment and removed variants existing in segmental duplication regions. We excluded variants with a minor allele frequency (MAF) ≥ 0.0005 in the 1000 Genomes Project (April 2012 release) or the NHLBI Exome Sequencing Project (ESP) Exome Variant Server (ESP6500 dataset) and variants reported in the dbSNP137 non-flagged database (excluding clinically associated SNPs).

To predict the potential impact of missense variants on protein function, we used the program Consensus Deleteriousness (Condel v.1.5),²³ which combines multiple prediction software algorithms (SIFT,²⁴ PolyPhen-2,²⁵ and MutationAssessor²⁶) into a single weighted score. We inspected all resultant variants through the Integrative Genomics Viewer (IGV).^{27,28} We also used an in-house database containing variants from 11 individuals with sporadic polyposis and 13 exomes from individuals with unrelated phenotypes (connective-tissue disorders); these were sequenced with our cohort of research participants as an additional internal control filter. We excluded variants that appeared more than once in these in-house exome databases.

Mutation Validation

Mutations in candidate genes of interest were validated by PCR-based region-specific mutation analysis through Sanger sequencing. In brief, gene-specific primers encompassing exonic regions harboring the particular variations were designed. Sequencing was performed in the forward and reverse directions with Applied Biosystems ABI 3730xl DNA Analyzers at the Genomics Core of the Lerner Research Institute of the Cleveland Clinic and the Eurofins Genomics DNA-sequencing facility (Louisville). Resultant chromatograms were analyzed with Mutation Surveyor DNA Variant Analysis Software (SoftGenetics), and mutations were reported according to the Human Genome Variation Society (HGVS) guidelines. Incidental findings were reported to our

genetic counselors to be communicated to the respective affected individuals and physicians of record according to the ACMG guidelines.²⁹ Confirmed variants were sequenced in all recruited family members regardless of disease status.

Variant Interpretation and In Silico Pathogenicity Predictions

Variants retained after prioritization were considered potentially pathogenic. Importantly, prioritized variants had to be shared in all affected members of the pedigree. Further prioritization was performed on the basis of the relation of the genes to the clinical phenotypes (where applicable), predicted loci of the mutations affecting functional protein domains, and in silico assessment of the mutations' effect on protein integrity through a combination of mutation-prediction algorithms, namely Condel,²³ SIFT,²⁴ PolyPhen-2,²⁵ MutationTaster,³⁰ MutPred,³¹ and Combined Annotation Dependent Depletion (CADD).³²

Prioritized candidate genes resulting from pedigree analysis were then screened in an independent series of 96 CS and/or CSL probands who were negative for germline mutations in *PTEN*, *SDHB*, *SDHC*, and *SDHD* and who had overlapping clinical phenotypes with those observed in the pedigree under study (mainly thyroid, breast, and endometrial cancers). Gene-specific primers were designed for all exons, and Sanger sequencing was performed as stated above.

The accession numbers for the *C16orf72*, *PTPN2*, and *SEC23B* sequences reported in this paper are GenBank: NM_014117.2, NM_080422.2, and NM_001172745.1, respectively.

TCGA Germline Variants Dataset Analysis

We analyzed The Cancer Genome Atlas (TCGA) datasets on papillary thyroid carcinoma (THCA, $n = 494$) and a subset of breast invasive carcinoma (BRCA, $n = 222$) and uterine corpus endometrial carcinoma (UCEC, $n = 156$). We analyzed all tumors with sequenced adjacent normal tissue or blood for analysis of germline mutations. Germline mutations were extracted and analyzed from raw sequencing data. In brief, genomic loci corresponding to candidate genes were extracted from the aligned BAM files, and a $MAF \leq 0.01$ was used for filtering variants through dbSNP137, 1000 Genomes, and ESP6500. Mutation predictions were obtained from SIFT,²⁴ PolyPhen-2,²⁵ and MutationTaster.³⁰ Importantly, we visualized and manually curated all mutations through the IGV^{27,28} to confirm that they indeed exist in the tumor and adjacent normal tissue and/or blood and are hence in the germline. Corresponding clinical data were obtained from the TCGA Data Portal Matrix.

Predicted Structural Effects of Variants

We predicted the structural effects of variants by using the HOPE server, which integrates a series of standard algorithms to generate predictions.³³ Because the *Homo sapiens* SEC23B crystal structure has not been resolved, the crystal structure of the human SEC23A-SEC24A heterodimer, complexed with the SNARE protein SEC22B³⁴ (PDB: 2NUP), was identified as a modeling template for predicting structural effects.

Cell Lines and Culture Conditions

Immortalized lymphoblastoid cell lines (LCLs) derived from affected individuals or normal control individuals were generated by the Genomic Medicine Biorepository of the Genomic Medicine Institute of the Cleveland Clinic according to standard procedures

and were subsequently maintained in RPMI-1640 supplemented with 20% fetal bovine serum (FBS) and 1% penicillin and streptomycin. HEK293T cells (originally purchased from the American Type Culture Collection) in 2011 and obtained in 2014 from the Cleveland Clinic Lerner Research Institute Cell Culture Core) were cultured in DMEM supplemented with 10% FBS and 1% penicillin and streptomycin. The Nthy-ori 3-1 human thyroid follicular epithelium cell line (catalog no. EC90011609, lot no. 09C008, passage no. 16, purchased in 2014 from Sigma) was cultured in RPMI-1640 supplemented with 2 mM glutamine and 10% FBS. All cell lines were maintained at 37°C and 5% CO₂ culture conditions and tested negative upon routine mycoplasma testing with the MycoAlert Mycoplasma Detection Kit (Lonza) at the C.E. lab (luminescence ratios < 0.9). Cell lines used have not been listed as cells known to be misidentified according to the International Cell Line Authentication Committee.³⁵ The Nthy-ori 3-1 cell line was authenticated through short-tandem-repeat PCR (AmpFLSTR SGM Plus PCR Amplification Kit, Life Technologies) by the European Collection of Cell Cultures (original source of the cell line, test date April 14, 2009).

RNA Extraction and qRT-PCR

RNA was extracted from LCLs derived from affected and unaffected individuals with the RNeasy Mini Kit (QIAGEN), purified with Turbo DNase treatment (Life Technologies), and reverse transcribed with Superscript III Reverse Transcriptase (Life Technologies). Primers were designed for gene transcripts of interest, and cDNA was quantified with SYBR Green (Life Technologies). Results were analyzed by the standard $\Delta\Delta CT$ method.

Immunoblotting

Protein was extracted from whole-cell lysates with the Mammalian Protein Extraction Reagent M-PER (Thermo Scientific Pierce) supplemented with a cocktail of protease and phosphatase inhibitors (Sigma-Aldrich) and was quantified with the BCA Protein Assay Kit (Thermo Scientific Pierce). Lysates were separated by SDS-PAGE and transferred onto nitrocellulose membranes. We probed for anti-SEC23B rabbit polyclonal (Abcam ab151258) at 1:1,000, for anti-SEC23A rabbit polyclonal generated against synthetic peptides (CQKFGYHKDDPNSFRFSET and DNAKYVKKGTRKHFEA)³⁶ at 1:2,000, and for anti-GAPDH rabbit monoclonal (Cell Signaling 2118) at 1:20,000 dilutions. Blots were scanned digitally and quantified with the Odyssey Infrared Imaging System (Li-Cor Biosciences).

Plasmids, Mutagenesis, and Cell-Line Transfection and Transduction

GFP- and FLAG-tagged wild-type *SEC23B* plasmids (pMSCV and pEGFP-C3 vectors, respectively) were used.³⁶ Wild-type plasmids were mutagenized for the missense mutation of interest with the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). All expression constructs were validated by Sanger sequencing prior to transfection or transduction. For transient overexpression, cell lines were transfected with Lipofectamine 3000 reagent (Invitrogen) according to the manufacturer's instructions. Fresh media replaced the transfection media 5 hr after transfection. Cells were grown for at least 24 hr before harvesting. To generate stable cell lines, we kept retrovirally transduced cells (a pool, given that no individual clones were isolated) under 1 $\mu\text{g/ml}$ puromycin selection for >30 days prior to downstream interrogation. We used four different pools of cells at passage no. 28 after

selection. All experiments were conducted with cells at passage nos. between 29 and 42.

Immunofluorescence

Cells were seeded on coverslips and were fixed with 4% paraformaldehyde for 10 min at room temperature. Coverslips were mounted with ProLong Gold Antifade mountant with DAPI (Invitrogen). Slides were visualized and images were obtained with a TCS SP5 confocal microscope (Leica).

Immunoprecipitation

Cells were pelleted after transfection and lysed with M-PER (Thermo Scientific Pierce) supplemented with a cocktail of protease and phosphatase inhibitors (Sigma-Aldrich). Protein lysates were collected by centrifugation at 13,000 RPM for 15 min at 4°C and pre-cleared by incubation with protein A/G PLUS-Agarose immunoprecipitation reagent (Santa Cruz sc-2003) for 30 min at 4°C on a rotator. Pre-cleared protein lysates were quantified with the BCA Protein Assay Kit (Thermo Scientific Pierce), and 1 mg/ml lysates were prepared. We used anti-SAR1A (Santa Cruz sc-130463) and anti-DDK (Origene 4C5) antibodies for pull-down and immunoblotting at the recommended dilutions.

Doubling Time and Cell Viability

Cells were trypsinized, homogenized, and counted with the Countess Automated Cell Counter (Invitrogen). We used Trypan blue to account for dead cells and to assess cell viability. We also used the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay (Sigma) to assess cell viability according to the manufacturer's instructions.

In Vitro Migration and Invasion Assays

Transiently transfected cells were grown for 48 hr, and a migration assay was performed according to Liang et al.,³⁷ with the minor variation of creating scratches on the cell monolayer with a p10 pipet tip. Migration of the stable cell lines was assessed with 6.5 mm transwells and 8.0 µm pore polycarbonate membrane inserts (Corning). Invasion was assessed with BioCoat Matrigel invasion chambers and 8.0 µm polyethylene terephthalate (PET) membrane (Corning). Cells were serum starved for 24 hr and harvested with the HyClone HyQTase Cell Detachment Reagent (GE Healthcare). For both assays, 5×10^4 cells were seeded in the upper chambers in serum-starved media. The lower chambers were filled with complete growth media supplemented with 10% FBS; however, for back-ground correction sets, both chambers were filled with serum-starved media. Migration and invasion experiments proceeded for 24 hr. Cells that migrated or invaded were fixed with 4% paraformaldehyde for 5 min, permeabilized with 100% methanol for 20 min, and stained with Giemsa for 15 min all at room temperature. Cells were photographed under 10× magnification and counted in quadruplicate fields of view in triplicate membranes.

SDS-PAGE Zymography

Cells were serum starved for 24 hr, and conditioned media were collected and concentrated with spin columns with a 10 kDa molecular-weight cutoff (Biovision). Protein concentrations were determined with the BCA Protein Assay Kit (Thermo Scientific Pierce). Equal volumes of sample and zymogram sample buffer (Bio-Rad) were mixed and loaded on 10% gelatin zymogram gels (Bio-Rad). After electrophoresis, gels were washed with renaturation buffer (Bio-Rad), incubated for 18 hr at 37°C in developing

buffer (Bio-Rad), and stained for 2 hr with Coomassie Brilliant Blue R-250 staining solution (Bio-Rad). Enzymatic activity was visualized as a colorless band against the Coomassie blue background. Bands were quantified with Image J software (NIH), and values were normalized to total protein.

ER-Stress Induction and Assessment of Cell Viability and Colony Formation

Cells were seeded and allowed to grow overnight and then treated with either 1 µM thapsigargin or 0.1 µg/ml tunicamycin (Sigma-Aldrich). Cells were checked every day for colony formation and cell number, and viability was assessed through Trypan blue staining. At the respective drug-treatment endpoints, cells were fixed with 70% ethanol for 5 min at room temperature and stained with Giemsa (Sigma-Aldrich). Colonies (defined as clusters of ≥ 50 cells) were counted with the use of a microscope (3× magnification).³⁸

Statistical Analyses

Sample-size power was calculated with OpenEpi software (see [Web Resources](#)); a population size of 1,000,000 individuals and a hypothesized percent frequency of outcome factor (*SEC23B*-variant-positive status) of $2\% \pm 5\%$ were used. The sample size (n) was predicted for various confidence levels (80%–99.99%).

Standardized incidence rate (SIR) for thyroid cancer is a summary ratio that allows comparison of incidence rates of thyroid cancer between affected (variant-positive) individuals and the general population. The expected number of thyroid cancers was obtained with the use of age-specific Surveillance Epidemiology and End Results (SEER) incidence rates from 2008 to 2012. There were 19 categories based on two genders and 19 age groups. Incidence was assumed to be zero for categories where statistics were not provided. Years of observation time per person (person-years of observation) were calculated from the date of birth to the date of cancer diagnosis for individuals who developed thyroid cancer or to the date of most recent information for individuals without cancer. We calculated the expected number of cancers by multiplying SEER incidence rates in each of the 19 categories by person-years of observation in each category. OpenEpi software was used for calculating SIR. The 95% confidence interval (CI) and corresponding p value were calculated with the mid-p exact test. A $SIR > 1.0$ implies that the incidence rate is greater for the population of interest than for the standard population. For analyses between affected population groups, we used 2×2 tables to calculate the odds ratio (OR) and 95% CI and used the mid-p exact test to calculate corresponding p values.

Experimental data between wild-type and mutant cell lines are given as means \pm SEM, and n corresponds to the number of experiments performed. Student's t test was used for significance testing as indicated in the figure legends. All statistical tests were two sided, and p values ≤ 0.05 were deemed significant.

Results

Exome Sequencing and Targeted Genotyping Identifies *SEC23B* as a Candidate Gene in a CS-Affected Family with Predominant Thyroid Cancer

We identified a multi-generation CS-affected family (family 616) presenting with different types of CS-related cancers, particularly thyroid cancer, across four generations.

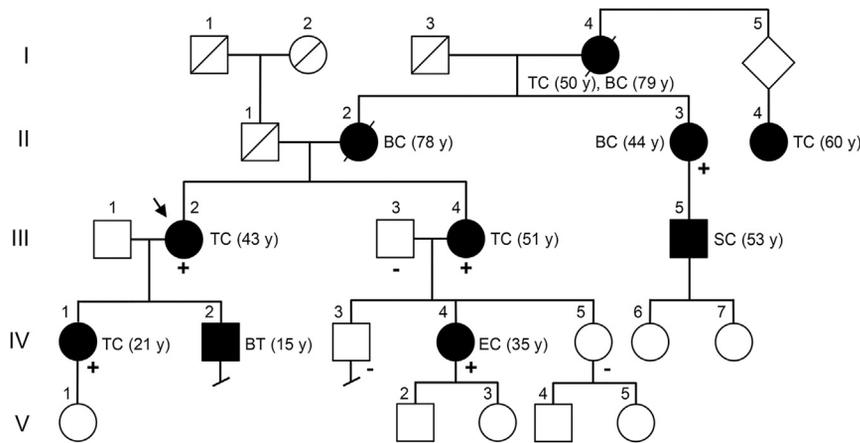


Figure 1. Pedigree of CS-Affected Family 616

The pedigree of CS-affected family 616 is shown with the presence of wild-type (–) or c.1781T>G (p.Val594Gly) mutant (+) *SEC23B* alleles. Symbols are as follows: square, male; circle, female; black fill, affected; slash, deceased. The proband (III-2) for whom whole-exome sequencing was performed is indicated by an arrow. Roman numerals represent generation number, and Arabic numerals represent individuals within each generation. Affected members show various types of CS-related cancers, including thyroid (II-4, III-2, III-4, and IV-1), breast (II-2 and II-3), endometrial (IV-4), skin (III-5), and both thyroid and breast (I-4) cancers. Age at diagnosis is indicated between parentheses. Abbreviations are as follows: TC, thyroid cancer; BT, benign thyroid; BC, breast cancer; EC, endometrial cancer; SC, skin cancer.

The family pedigree suggests autosomal-dominant inheritance with complete penetrance (Figure 1). The proband (III-2) is a 60-year-old female who presented at our CS clinic with follicular variant papillary thyroid cancer (FvPTC; diagnosed at age 43), goiter, breast fibroadenoma and papilloma, typical breast ductal hyperplasia, fibro-

cystic breast disease, trichilemmoma, papillomatous papules of the mucosa, gastrointestinal polyps, and uterine fibroids. The individual had a high CC score of 16, but importantly, analyses for *PTEN* intragenic mutations and deletions revealed a wild-type sequence. This suggested a high phenotypic burden independent of *PTEN*, and

Table 1. Characteristics of the Three Candidate Variants Identified in Affected Members of Family 616

	Genes		
	<i>SEC23B</i>	<i>C16orf72</i>	<i>PTPN2</i>
Genomic position ^a	chr20: 18,529,290	chr16: 9,186,804	chr18: 12,794,321
DNA variant	c.1781T>G	c.253T>C	c.1204G>A
Protein alteration	p.Val594Gly	p.Ser85Pro	p.Ala402Thr
Predicted Effect of Missense Variants			
Condel ^b	deleterious	deleterious	deleterious
MutationTaster	disease causing	disease causing	polymorphism
MutPred g score (>0.5)	0.785	0.525	0.578
CADD score ^c	22.2	23.2	9.28
Evolutionary Conservation and Allele Frequency			
PhyloP ^d	4.572	2.882	1.282
PhastCons ^e	1	1	0.994
NHLBI ESP ^f MAF	G = 0, T = 13,006	C = 0, T = 12,994	T = 0, C = 13,006
Tissue-Specific Protein Levels^g			
Thyroid	+	–	+
Breast	+	–	+
Endometrium	+	–	+

^aGenomic positions correspond to the UCSC hg19 reference assembly.

^bCondel combines multiple prediction software algorithms (SIFT, PolyPhen-2, and MutationAssessor).

^cCADD-scaled C score ≥ 20 indicates variants within the top 1% of likely deleterious variants across the human genome.

^dPhyloP scores range between –14 and +6; conserved sites have positive scores.

^ePhastCons scores range between 0 and 1; values closer to 1 have a higher probability of nucleotide conservation.

^fNHLBI Exome Sequencing Project (ESP) Exome Variant Server (see [Web Resources](#)) v.0.0.28 (accessed May 9, 2015).

^gData analyzed from the Human Protein Atlas for immunohistochemically stained normal and cancer tissues.

Table 2. Genotype and Clinical Phenotype of Known Mutation-Negative CS and/or CSL Individuals with Germline *SEC23B* Variants

	Probands		
	CCF02565	CCF05664	CCF04372
Age	59 years	69 years	51 years
Sex	female	female	female
CC score	7	11	14
Clinical features	FvPTC (age 52 years), macrocephaly	PTC (age 45 years), ductal carcinoma in situ, atypical ductal hyperplasia, fibrocystic breast disease, hemangioma, macrocephaly	PTC (age 45 years), goiter, Hashimoto thyroiditis, macrocephaly
<i>SEC23B</i> variant	c.490G>T (p.Val164Leu), rs36023150	c.490G>T (p.Val164Leu), rs36023150	c.1512T>C (p.=), near splice, rs138198461
NHLBI ESP MAF	T = 97, G = 12,909 (0.0075)	T = 97, G = 12,909 (0.0075)	C = 52, T = 12,954 (0.0040)
SIFT	damaging (0.01)	damaging (0.01)	–
PolyPhen-2	possibly damaging (0.87)	possibly damaging (0.87)	–
MutationTaster	disease causing	disease causing	disease causing (acceptor lost)
PhyloP ^a	5.615	5.615	–
PhastCons ^b	1	1	–

Abbreviations are as follows: FvPTC, follicular variant papillary thyroid cancer; PTC, papillary thyroid cancer.

^aPhyloP scores range between –14 and +6; conserved sites have positive scores.

^bPhastCons scores range between 0 and 1; values closer to 1 have a higher probability of nucleotide conservation.

subsequent testing also showed no germline mutations or deletions in *SDHB*, *SDHC*, or *SDHD* and the absence of *KLLN* hypermethylation. We therefore recruited the proband for our whole-exome sequencing efforts to identify previously unidentified candidate genes in mutation-negative CS individuals with high CC scores. Exome sequencing of germline genomic DNA derived from III-2 achieved a mean coverage of 96.3×, and >90% of the targeted regions were covered by at least ten reads. Filtering the gene variants through ANNOVAR²² and Condel²³ and subsequent manual curation through the IGV^{27,28} retained 44 potentially deleterious gene variants, of which 39 were validated by Sanger sequencing (Table S2).

To identify candidate cancer-predisposing CS-related genes segregating in the proband's family, we applied Sanger sequencing for the 39 validated gene variants in seven other family members with available DNA (II-3, III-3, III-4, IV-1, IV-3, IV-4, and IV-5). Four (II-3, III-4, IV-1, and IV-4) met CS diagnostic criteria and presented with different cancer types, whereas the remaining three (III-3, IV-3, and IV-5) had no apparent phenotypes. Upon variant filtration (Table S2), all affected members were found to share three genes with heterozygous missense variants: *C16orf72* (c.253T>C [p.Ser85Pro]), *PTPN2* (c.1204G>A [p.Ala402Thr]), and *SEC23B* (c.1781T>G [p.Val594Gly]). Notably, no variants were noted in *PTEN*, *SDHB*, *SDHC*, *SDHD*, *KLLN*, *PIK3CA*, or *AKT1*. The three variants in *SEC23B*, *C16orf72*, and *PTPN2* have not been reported in public databases (dbSNP137, 1000 Genomes, or ESP6500). We examined in silico the predicted functional impact of these prioritized variants and noted a higher index of pathogenicity for the *SEC23B* variant ac-

ording to multiple mutation prediction databases and multi-species evolutionary amino acid residue conservation (Table 1). We next inspected the protein status of the prioritized genes in tissues of interest through the Human Protein Atlas.^{39,40} We found that only *PTPN2* and *SEC23B* are expressed in thyroid, breast, and/or endometrial normal and cancer tissues (organs affected in family 616). However, given that the *PTPN2* variant was predicted as the least damaging and annotated as a “polymorphism” by MutationTaster, the combined in silico data indicate that the *SEC23B* variant is the most critical germline variant in family 616 (Figure S1).

Germline Heterozygous *SEC23B* Variants Exist in Unrelated CS Probands with Thyroid Cancer

We next sought to investigate whether germline variants in *SEC23B* and the other two prioritized genes exist in other CS individuals. We selected 96 CS and CSL probands with either a diagnosis of thyroid cancer or signs of other cancers observed in family 616, namely breast and endometrial carcinomas (Table S3). This sample size gave us >95% power to detect a 2% prevalence of variants. We directly sequenced all exons of *SEC23B*, *PTPN2*, and *C16orf72*. We found deleterious germline heterozygous *SEC23B* variants in three probands (3.1%). Papillary thyroid cancer (PTC) was found in all three individuals, one of whom showed the follicular variant on histology. Interestingly, two individuals had the same missense *SEC23B* variant (c.490G>T [p.Val164Leu]) affecting a highly evolutionarily conserved residue within the *SEC23B* trunk domain,⁴¹ and all three variants we found are rare in the general population and predicted to be damaging (Table 2). In support of our

Table 3. Characteristics of Germline *SEC23B* Variants in TCGA Thyroid, Breast, and Endometrial Cancer Datasets

Cancer Datasets	Variants	NHLBI ESP ^a MAF	SIFT	PolyPhen-2	MutationTaster	Conservation	
						Phylo ^b	PhastCons ^c
THCA (n = 494)	c.1661G>A (p.Arg554Gln)	0	damaging (0.02)	probably damaging (0.998)	disease causing	5.365	1
	c.1598T>G (p.Val533Gly)	0	damaging (0)	possibly damaging (0.621)	disease causing	4.497	1
	c.167A>G (p.Tyr56Cys)	0	damaging (0)	possibly damaging (0.752)	disease causing	4.798	1
	c.301A>G (p.Ile101Val)	0	damaging (0.04)	possibly damaging (0.643)	disease causing	4.760	1
	c.689+1G>C	0	–	–	disease causing	5.047	1
	c.2101C>T (p.Arg701Cys)	0	damaging (0)	probably damaging (1.000)	disease causing	2.648	0.988
	c.1636C>T (p.Arg546Trp)	T = 1, C = 13,005 (0.000077)	damaging (0)	probably damaging (1.000)	disease causing	1.987	1
	c.2031G>A (p.Met677Ile)	A = 2, G = 13,004 (0.000154)	damaging (0.02)	benign (0.014)	disease causing	5.584	1
	c.40C>T (p.Arg14Trp)	T = 3, C = 13,003 (0.000231)	damaging (0.01)	probably damaging (0.999)	disease causing	2.132	1
	c.1484G>A (p.Arg495His)	A = 95, G = 12,911 (0.007358)	tolerated (0.12)	probably damaging (0.961)	disease causing	5.579	1
	c.490G>T (p.Val164Leu)	T = 97, G = 12,909 (0.0075)	damaging (0.01)	possibly damaging (0.870)	disease causing	5.615	1
BRCA (n = 222)	c.884C>A (p.Pro295His)	0	damaging (0)	probably damaging (0.983)	disease causing	5.778	1
	c.985G>T (p.Ala329Ser)	T = 4, G = 13,002 (0.000307)	damaging (0.04)	possibly damaging (0.601)	disease causing	5.591	1
	c.74C>A (p.Pro25His)	A = 5, C = 13,001 (0.000385)	damaging (0)	probably damaging (1.000)	disease causing	5.701	1
	c.490G>T (p.Val164Leu)	T = 97, G = 12,909 (0.0075)	damaging (0.01)	possibly damaging (0.870)	disease causing	5.615	1
UCEC (n = 156)	c.389T>C (p.Ile130Thr)	0	damaging (0.05)	benign (0.006)	disease causing	4.638	1
	c.649C>T (p.Arg217*)	T = 1, C = 13,005 (0.000077)	–	–	disease causing	–0.056	0.031
	c.490G>T (p.Val164Leu)	T = 97, G = 12,909 (0.0075)	damaging (0.01)	possibly damaging (0.870)	disease causing	5.615	1

^aNHLBI Exome Sequencing Project (ESP) Exome Variant Server v.0.0.28 (accessed May 9, 2015).

^bPhyloP scores range between –14 and +6; conserved sites have positive scores.

^cPhastCons scores range between 0 and 1; values closer to 1 have a higher probability of nucleotide conservation.

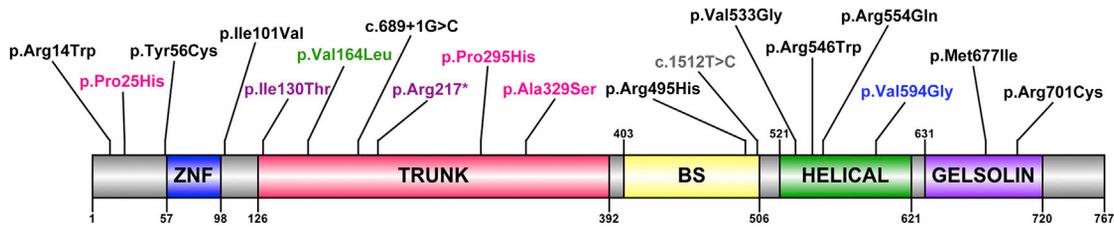


Figure 2. Germline *SEC23B* Mutation Spectra in CS and Apparently Sporadic CS-Related Cancers from TCGA

A schematic of *SEC23B* and its functional domains is shown; Arabic numerals correspond to amino acid number. The amino acid positions corresponding to germline heterozygous variants are indicated for variants identified in CS-affected family 616 (blue), an unrelated CS proband (gray), TCGA apparently sporadic thyroid cancer (black), breast cancer (pink), endometrial cancer (purple), and a shared variant observed in all three cancer types and in CS (green). Abbreviations are as follows: ZNF, zinc finger domain; BS, beta-sandwich.

in silico gene prioritization, no germline variants were found in *C16orf72* or *PTPN2* in the 96 unrelated CS and CSL probands.

Germline Heterozygous *SEC23B* Variants Are Enriched in Apparently Sporadic Thyroid Cancer and Associate with Early Age of Onset

To determine whether *SEC23B* is associated with apparently sporadic cancers, we next analyzed the spectrum and frequencies of germline variants in three TCGA CS-component cancers (also observed in family 616): papillary thyroid ($n = 494$), breast invasive ($n = 222$), and uterine corpus endometrioid ($n = 156$) carcinomas. Indeed, we found germline heterozygous variants in all three studied cancer types, and the highest frequency (4%) was observed in thyroid cancer. Furthermore, we observed an overrepresentation of unique *SEC23B* variants in thyroid cancer (Table 3). All variants we found are predicted to be deleterious and have various protein structural effects (Table S4). For affirmation, we also inspected *C16orf72* and *PTPN2* and observed the absence of mutations or much lower mutation frequencies in the three cancer types (Table S5). Although we used a small sample size, we calculated the age-adjusted SIR of thyroid cancer in *SEC23B*-variant-positive TCGA individuals ($n = 19$) and found that it is significantly higher than in the US general population (SIR = 242.6; 95% CI = 150.4–371.8; $p = 1.0 \times 10^{-11}$). Relatedly, the median age at diagnosis of thyroid cancer is younger for *SEC23B*-variant-positive individuals, especially for those with unique variants, than for the entire TCGA thyroid cancer cohort (36 versus 46 years, $p = 0.025$). One such individual who was diagnosed with PTC at age 21 had a parent with thyroid cancer. Lastly, three unrelated individuals harbored the same *SEC23B* variant (c.1598T>G [p.Val533Gly]) affecting the helical *SEC23B* domain, where the p.Val594Gly variant from family 616 also occurs. The spectrum of germline *SEC23B* variants in CS and TCGA CS-related cancers is depicted in Figure 2.⁴²

Our findings indicate that both unique and previously reported ($0.00008 \leq \text{MAF} \leq 0.008$) germline *SEC23B* variants exist in CS individuals and in the studied individuals with sporadic cancer. This prompted us to investigate the frequency of heterozygous *SEC23B* variants in individuals

with no reported cancer diagnoses. We analyzed the NHLBI exomes representing 2,203 African-American and 4,300 European-American unrelated individuals, totaling 13,006 chromosomes. We found deleterious *SEC23B* variants in 155 (1.2%; $\text{MAF} \leq 0.008$) or 58 (0.45%; $\text{MAF} \leq 0.0004$) chromosomes (Table S6). In comparison, the studied TCGA populations with apparently sporadic cancer show deleterious *SEC23B* variants in 29 (1.7%; $\text{MAF} \leq 0.008$) or 16 (0.93%; $\text{MAF} \leq 0.0004$) out of 1,744 chromosomes. Interestingly, this comparison yields a significantly higher OR for thyroid cancer in TCGA populations than in NHLBI populations, especially for variants with a $\text{MAF} \leq 0.0004$ (OR = 2.5; 95% CI = 1.3–4.7; $p = 0.0116$) (Table S7).

Functional Characterization of p.Val594Gly in Family 616 Reveals a Tumorigenic Cellular, Molecular, and Biochemical Phenotype

SEC23B is a component of COPII vesicles for anterograde ER-to-Golgi transport.^{43,44} Recessive loss-of-function mutations cause congenital dyserythropoietic anemia type II (CDAIL),^{45,46} resulting in decreased *SEC23B* levels.^{45,47,48} Available clinical laboratory results from a subset of the available heterozygous *SEC23B*-variant-positive CS individuals were not suggestive of a classic CDAIL phenotype, and reports signed off as “normal” for most of these individuals (Table S8). Also, lymphoblastoid cells derived from family 616 showed no differences in *SEC23B* transcript and protein levels between variant carriers and wild-type individuals (Figure 3), suggesting that a probable change in function rather than reduced *SEC23B* levels (loss of function) is the likely mechanism in CS.

We then sought to functionally characterize the unique deleterious *SEC23B* variant (c.1781T>G [p.Val594Gly]) found in family 616. We transiently transfected HEK293T cells with wild-type and p.Val594Gly *SEC23B*. Mutant cells revealed aberrant aggregation of *SEC23B* and significantly increased cell migration and upregulation of epithelial-to-mesenchymal transition (EMT) genes, suggesting a cancer-relevant phenotypic effect (Figure S2). Notably, the p.Val594Gly variant occurs between the alpha-N and alpha-M regions within the *SEC23B* helical critical for binding to SARI,⁴⁹ a GTPase activated by *SEC23* and another component of the inner coat of COPII

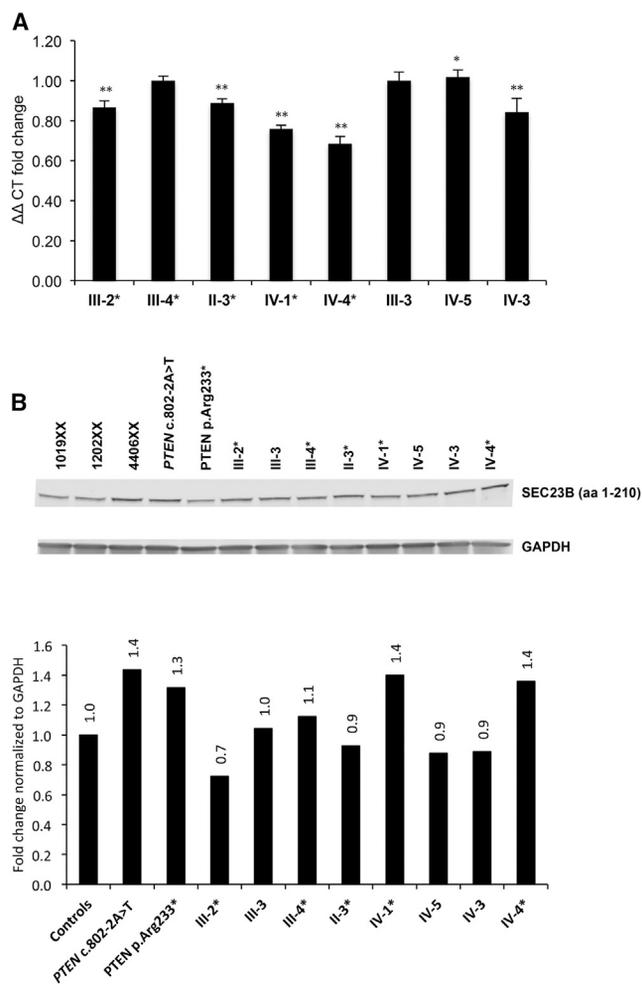


Figure 3. Lymphoblastoid Cells Derived from Family 616 Show No Differences in *SEC23B* Transcript and Protein Levels between p.Val594Gly Carriers and Wild-Type Individuals

(A) qRT-PCR expression showed no notable trend in *SEC23B* transcript expression between wild-type individuals and variant carriers. Values are normalized to unaffected individual III-3 and represent three technical replicates for which data represent mean values \pm SEM. ** $p < 0.01$, * $p < 0.05$ (two-sided Student's *t* test).

(B) *SEC23B* levels and quantification by western blot. *PTEN*-mutation-positive lymphoblastoid cells (LCLs) from other CS individuals were also used as internal exploratory controls. 1019XX, 1202XX, and 4406XX represent healthy control LCLs unrelated to the pedigree.

SEC23B-mutation carriers from the pedigree are indicated by asterisks in each panel. "aa" refers to the amino acids targeted by the *SEC23B* antibody.

vesicles. We speculated that the observed alteration might affect binding of *SEC23B* to *SAR1A* and confirmed this through immunoprecipitation of tagged *SEC23B* and *SAR1A* in 293T-*SEC23B*-WT and 293T-*SEC23B*-V594G cells (Figure S3). Pertinent to the cancer type (i.e., thyroid) enriched in family 616 and TCGA, we then generated stable normal thyroid follicle epithelial⁵⁰ cell lines expressing wild-type or p.Val594Gly *SEC23B* (Nthy-*SEC23B*-WT or Nthy-*SEC23B*-V594G, respectively). We validated that compared to Nthy-*SEC23B*-WT and Nthy-*SEC23B*-E109K cells (the latter of which represents one of two CDAI1

founder mutations accounting for ~50% of reported cases^{51,52}), the Nthy-*SEC23B*-V594G cells associated with the CS-affected family showed aberrant *SEC23B* aggregation (Figure 4A and Figure S4). The Nthy-*SEC23B*-E109K cells lacked a normal ER expression pattern, supporting the loss of altered protein in anemia. Relatedly, western blot analysis revealed similar levels of *SEC23B* and paralog *SEC23A* in wild-type and mutant Nthy-*SEC23B*-V594G cells, adding to the evidence that the mutation does not result in loss of function of *SEC23B* and does not lead to concomitant compensation by *SEC23A* (Figure 4B).⁵³

To our surprise, Nthy-*SEC23B*-WT and Nthy-*SEC23B*-V594G cells showed no differences in growth, viability, or migration ability (Figure S5). We hypothesized that another aspect of the EMT phenotype might be affected in normal thyroid cells, and so we turned our attention to investigating other EMT-related manifestations (or phenotypes). Indeed, Transwell invasion assays revealed a significantly higher (9-fold) increase in Matrigel invasion ($p = 2.31 \times 10^{-7}$) in Nthy-*SEC23B*-V594G cells (Figure 5A). We then demonstrated through gelatin zymography that these mutant cells showed increased activity of MMP-2 and MMP-9 (matrix-digesting metalloproteinases) in the supernatant (Figure 5B), thus explaining the invasive potential of these cells.

We continued to examine other cellular phenotypes linked to carcinogenesis. Given that loss of *SEC23B* in mice causes ER stress and degeneration of secretory organs,³⁶ we hypothesized that heterozygous *SEC23B* variants might prime mutant cells to respond differently to ER-stress stimuli. We first treated cells with thapsigargin, an ER Ca^{2+} ATPase inhibitor conventionally used to induce ER stress experimentally. Previous studies have reported that ER stress can induce an EMT phenotype in thyroid and lung alveolar epithelial cells.^{54–56} Consistent with these studies,^{54–56} starting at 6 hr after treatment with thapsigargin, both wild-type and mutant cells showed a cellular and molecular phenotype consistent with EMT (Figure S6). Intriguingly, cellular phenotypic differences between wild-type and Nthy-*SEC23B*-V594G cells became evident only after chronic exposure to thapsigargin (at ~14 days) in that mutant cells formed significantly larger colonies ($p = 1.84 \times 10^{-5}$) that survived for up to 36 days after treatment (Figure 5C and Figure S7). In support of these findings, treatment with tunicamycin, another standard ER-stress inducer (inhibitor of protein N-linked glycosylation),⁵⁷ also resulted in the formation of distinct colonies and conferred an apparent growth advantage at lower doses in mutant cells (Figure 5D). Overall, our experimental findings demonstrate examples of how a *SEC23B* germline variant could account for increased susceptibility to cancer, particularly thyroid cancer.

Discussion

Our observations reveal that a subset of CS and CSL individuals without germline mutations in known genes

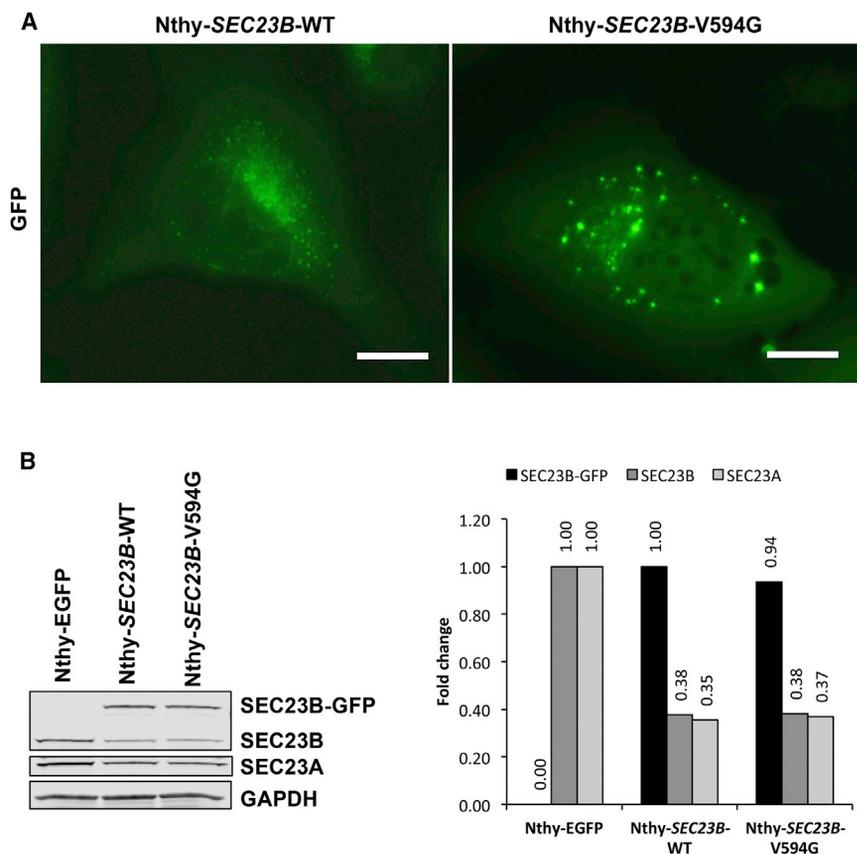


Figure 4. Cellular and Biochemical Characterization of Family 616 Variant p.Val594Gly in a Normal Thyroid Cell Line
 (A) Normal thyroid Nthy-ori 3-1 cells were stably transduced with wild-type and mutant c.1781T>G (p.Val594Gly) *SEC23B* fused with GFP. Wild-type cells (Nthy-*SEC23B*-WT) showed *SEC23B* localization typical of ER proteins, whereas a subset of mutant cells (Nthy-*SEC23B*-V594G) showed aberrant aggregation of *SEC23B*. Scale bars represent 20 μ m.

(B) Western blot and densitometric analysis of overexpressed *SEC23B*, endogenous *SEC23B* and paralog *SEC23A*, and *GAPDH* (loading control) from stable cell lines. Wild-type and mutant Nthy-*SEC23B*-V594G cells showed similar levels of *SEC23B* and *SEC23A*. The Nthy-EGFP cell line was used as a negative control to indicate baseline levels of *SEC23B* and *SEC23A* in the cell line (no *SEC23B* overexpression in these cells). Exogenous *SEC23B*-GFP normalized to the wild-type, and endogenous *SEC23B* and *SEC23A* normalized to levels in the Nthy-EGFP cell line.

might be accounted for by germline variants in *SEC23B*. Of the three *SEC23B* variants we identified in four unrelated CS and/or CSL individuals, the pathogenic c.1781T>G (p.Val594Gly) variant related to family 616 has never been described before; our functional analyses corroborate the pathogenicity of this missense mutation. Structurally, p.Val594Gly occurs between the alpha-N and alpha-M regions within the *SEC23B* helical domain critical for binding to SAR1, a GTPase activated by *SEC23* and another component of the inner coat of COPII vesicles. Interestingly, the p.Val164Leu variant present in two other unrelated CS and/or CSL individuals occurs within a different protein domain, the *SEC23* trunk domain. This region forms the dimer interface between *SEC23* and *SEC24* and also contacts SAR1,⁴¹ suggesting a possible common downstream effect. This variant has been reported in ESP6500 with a MAF of 0.0075 (97/12,909). Therefore, although we cannot ascertain causation, we bear in mind that cancer is a common phenotypic manifestation even in the general population. For example, according to SEER, the risk of thyroid cancer in the general population is 1%, that of female breast cancer is 12%, and that of endometrial cancer is 2.6%.¹³ Hence, it is possible that this *SEC23B* p.Val164Leu variant might be predisposing these individuals to cancer even in the general population.

It is also worth mentioning that *SEC23B* p.Val164Leu is reminiscent of two *SDHD* variants (c.34G>A [p.Gly12Ser] and c.149A>G [p.His50Arg]) with well-established functional roles in CS and CSL. Although c.34G>A and

c.149A>G are reported in ESP6500 with a MAF of 0.0080 (103/12,893) and 0.0063 (82/12,914), respectively, we have shown that these two vari-

ants result in increased reactive oxygen species (via PHD) and upregulation of p-AKT and p-ERK1/2 (mimicking loss of function of PTEN).⁹ Furthermore, they also destabilize p53 by altering the NAD/FAD ratio, which leads to decreased NQO1-p53 interaction with consequent non-MDM2 degradation of p53.¹⁰ More recently, we also elucidated that these same *SDHD* variants lead to nuclear mislocalization and oxidation of PTEN in CS-related thyroid cancer.⁵⁸ Therefore, although *SEC23B* might be a pertinent cancer-predisposing gene, as demonstrated by our functional evidence of the c.1781T>G (p.Val594Gly) variant, the relevance of the c.490G>T (p.Val164Leu) variant would require similar experimental interrogation for more conclusive evidence, although we believe that the latter could be a low- to moderate-penetrance allele. Lastly, because ESP6500 represents a set of population controls, it is plausible that some individuals might have unrecognized CS and CSL, given the subtlety of some of the clinical features and the existence of phenotypes often observed in the general population, such as cancer and non-specific skin features.

We also found heterozygous germline *SEC23B* variants in apparently sporadic breast, thyroid, and endometrial cancers in TCGA. Notably, these variants were overrepresented in individuals with thyroid cancer and enriched with unique *SEC23B* variants predicted to be deleterious and absent from public databases, suggesting a functional role in sporadic thyroid carcinogenesis. There is no overlap of mutation spectra between *CDAI1*⁵² and CS. However,

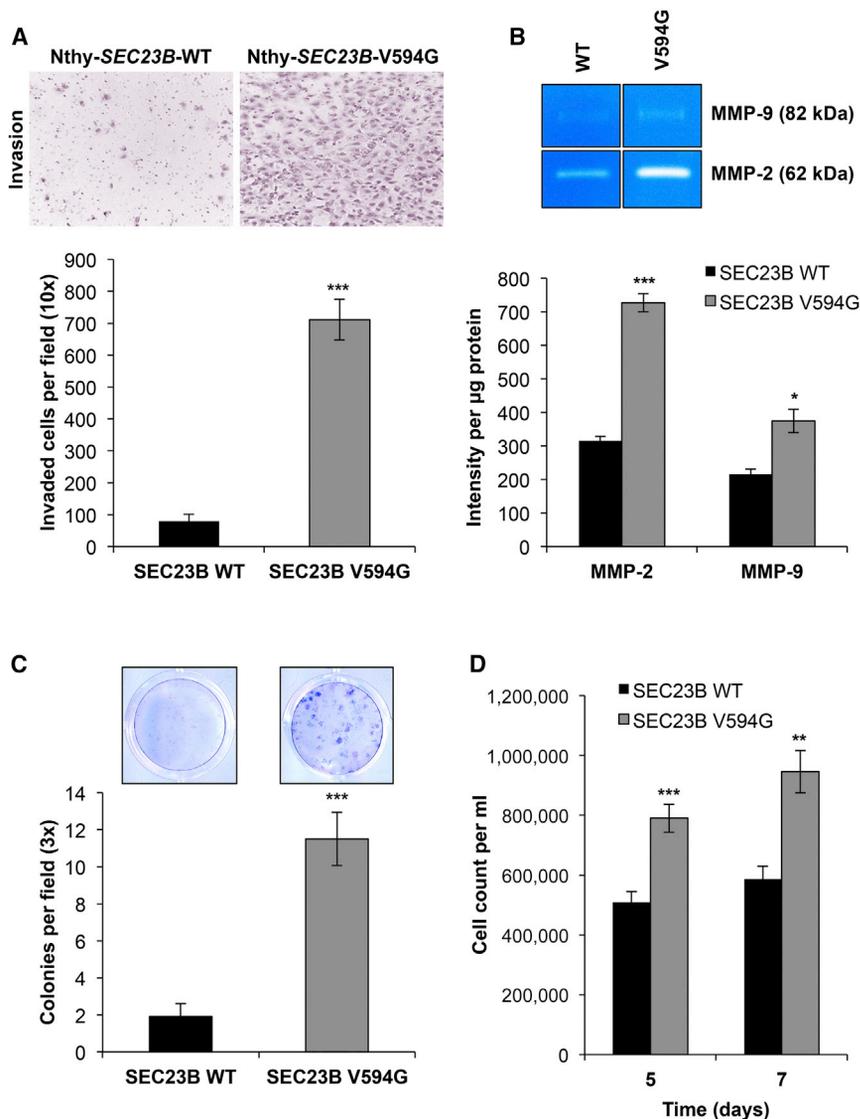


Figure 5. The *SEC23B* c.1781T>G (p.Val594Gly) Variant Confers Neoplastic Phenotypes upon ER Stress

(A) Representative images of invaded Nthy-*SEC23B*-WT (left) and Nthy-*SEC23B*-V594G (right) cells in a Transwell assay after overnight serum starvation. Invaded cells were counted after 24 hr in quadruplicate fields of view in triplicate membranes. Data represent mean values \pm SEM. *** p < 0.001 (two-sided Student's t test).

(B) Gelatin zymography was performed with conditioned media from wild-type and mutant cells and showed more MMP-2 and MMP-9 activity in supernatant originating from Nthy-*SEC23B*-V594G (right lanes) than in supernatant from Nthy-*SEC23B*-WT (left lanes) cells. Data represent triplicate lanes with band intensity normalized to protein mass. Data represent mean values \pm SEM. *** p < 0.001, * p < 0.05 (two-sided Student's t test).

(C) Wild-type and mutant cells were treated with 1 μ M thapsigargin (TG) for the induction of ER stress. Large colonies of cells expressing altered *SEC23B* formed, after which both Nthy-*SEC23B*-WT and Nthy-*SEC23B*-V594G cells were fixed and stained with Giemsa, and colonies were counted. Stained colonies represent cells 14 days after treatment, and a colony is standardly defined as a cluster of \geq 50 cells. Data shown represent three independent experiments done in triplicate for each genotype. Data represent mean values \pm SEM. *** p < 0.001 (two-sided Student's t test).

(D) Treatment of wild-type and mutant cells with 0.1 μ g/ml tunicamycin (TM), another ER-stress-inducing agent, resulted in more growth of mutant cells than of wild-type cells. Data shown represent two independent experiments done in triplicate for each genotype at each time point. Data represent pooled mean values \pm SEM. *** p < 0.001, ** p < 0.01 (two-sided Student's t test).

five TCGA individuals with cancer harbored germline heterozygous *SEC23B* variants (c.40C>T [p.Arg14Trp], c.649C>T [p.Arg217*], c.2101C>T [p.Arg701Cys], c.74C>A [p.Pro25His], and c.689+1G>C) known to be associated with CD411 in the homozygous or compound-heterozygous setting. Because we cannot ascertain anemia-related clinical phenotypes unrelated to cancer in these individuals, such variants warrant further investigation. Interestingly, irrespective of *SEC23B*-mutation status, an agnostic search in the Catalogue of Somatic Mutations in Cancer (COSMIC) showed that 625/828 (~75%) of apparently sporadic cancers of different types include overexpression of *SEC23B* transcripts (Z score values > 2) and that 84/828 (~10%) show high-level amplifications (gain of copy number), suggesting that at least in a subset of cancers, *SEC23B* could indeed be playing a previously unappreciated role in carcinogenesis at large.

Given the role of *SEC23B*, we speculate that because of the high demand for protein synthesis and trafficking

and, hence, increased sensitivity to perturbations in ER function, germline *SEC23B* variants associate with cancers in so-called professional secretory tissues (in this study, thyroid, breast, and endometrium). Our functional data also suggest that p.Val594Gly is a change-of-function substitution that favors tumorigenesis even, or especially, under stressful micro-environmental conditions. Cells respond to ER stress by activating the unfolded protein response (UPR), a signaling network that allows the recovery of ER function by upregulating adaptive factors, such as chaperone proteins and transcription factors, that regulate protein synthesis, growth arrest, and ER biogenesis among various pro-survival responses.⁵⁹ Cell fate is largely determined by the intensity and duration of the stress stimuli, and failure to adapt to ER stress ultimately leads to the activation of apoptotic pathways. Increasing evidence, however, suggests that cancer cells hijack or are addicted to ER-stress-induced responses as a survival mechanism and thus often confer resistance to anti-cancer therapies.^{60–65}

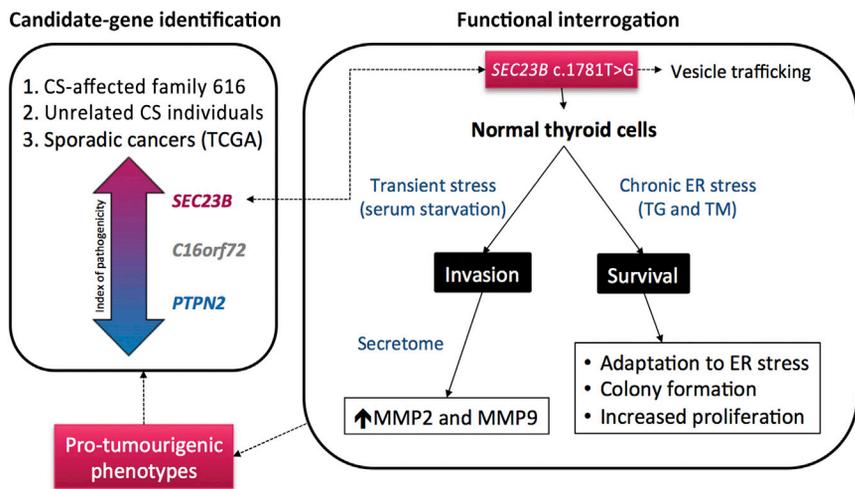


Figure 6. *SEC23B* Germline Heterozygous Variants Cause Predisposition to Hereditary and Apparently Sporadic Cancers Abbreviations are as follows: CS, Cowden syndrome; TG, thapsigargin; TM, tunicamycin.

indeed increasing evidence that monoallelic and biallelic mutations in a single gene can lead to completely different phenotypes.^{5,68,69} Relatedly, there are currently no reports of cancer predisposition in the obligate heterozygous parents of individuals with CDAII. One possibility is that the carried loss-of-function CDAII-associated

Mouse models completely deficient of *SEC23B* die shortly after birth.³⁶ Interestingly, these mice do not have anemia but rather suffer from UPR-induced degeneration of professional secretory tissues that seem to highly depend on *SEC23B* for efficient cargo transport and secretion. More recently, it was reported that mice with hematopoietic compartment localized deficiency of *SEC23B* also show no red blood cell phenotypes.⁶⁶ The difference in phenotypes observed between humans and mice was attributed to the existence of hypomorphic mutations in humans and a potential species-specific overlap of function with the *SEC23A* paralog. Interestingly, Korpál and colleagues agnostically identified that direct targeting of *SEC23A* by miR200s promotes metastatic lung colonization in mice by inhibiting the secretion of metastasis-suppressive proteins and promoting the mesenchymal-to-epithelial transition (MET).⁶⁷ The appearance of large epithelial colonies in our mutant Nthy-*SEC23B*-V594G cells after ER stress is suggestive of a phenotype favoring cell colonization. Whether this is conducive to primary tumor formation or metastatic cell seeding has yet to be determined. Moreover, increased MMP-2 and MMP-9 activity from the cell supernatant hints at the possibility that protein-trafficking alterations modulate the cell-derived secretome similarly to the newly discovered *SEC23A* role in trafficking metastasis-regulating proteins.⁶⁷ Furthermore, the different migration patterns between Nthy cells and HEK293T cells transfected with our mutant *SEC23B* might lend a clue. Although we could attribute differences to varying cellular contexts, HEK293T cells are immortalized with SV40 large T antigen, whereas Nthy cells are immortalized with SV40, where the origin of replication is absent. One way or the other, we speculate that the suspected heterozygous change-of-function *SEC23B* mutations in a cancer context could override the existence of wild-type *SEC23A* and *SEC23B*, at least in thyroid, breast, and endometrial cancers, potentially through ER stress and UPR “addiction.”

However, the precise mechanism leading to neoplastic transformation warrants further investigation. There is

mutations result in a less severe hypomorphic phenotype even in the heterozygous state, whereas we speculate that cancer-promoting heterozygous variants do not affect protein stability and result in change-of-function effects. The other possibility is that parents of CDAII individuals, especially if the latter are pediatric, might not be followed up long enough for the existence of cancer to be ascertained. This is confounded by the fact that CDAII is a rare disorder with a global series of only 205 affected individuals (172 unrelated),⁵² and so it is plausible that the existence of common cancers in these families might go unnoticed. This phenomenon is paralleled by the discovery of germline heterozygous mutations in *FH* (MIM: 136850), encoding fumarate hydratase, in hereditary leiomyomatosis and renal cell carcinoma (HLRCC [MIM: 150800]).⁶⁹ Homozygous or compound-heterozygous *FH* mutations cause fumarate deficiency (FMRD [MIM: 606812]), a lethal neuropathy, cardiopathy, and/or hepatopathy in infancy or early childhood.^{68,70} The heterozygous parents of the individuals with these rare diseases are not documented to have had cancer. Similarly, heterozygous *SDHA*, *SDHB*, *SDHC*, and *SDHD* mutations were described to cause endocrine neoplasias^{71–73} only 13 years ago, whereas germline homozygous or compound-heterozygous mutations in *SDHA*, *SDHB*, *SDHC*, and *SDHD* (any one of four subunits of succinate dehydrogenase or mitochondrial complex II) have long been known to cause Leigh syndrome, a lethal neuropathy that manifests shortly after birth.^{74,75} Yet, none of the parents of individuals with Leigh syndrome have been reported in the literature to have had component cancers. Quite similar to *Sec23b* mouse models, the existing murine models of *Sdhb*, *Sdhc*, and *Fh* do not mimic the human neoplasia conditions and indeed have little to no phenotype (reviewed by Lepoutre-Lussey et al.⁷⁶). It is tempting to speculate on a “common” mechanism that might explain the disparate phenotypes, namely cancer and non-neoplastic yet extremely serious phenotypes, between heterozygous and homozygous or compound-heterozygous germline mutations in seemingly disparate genes (*SEC23B*, *SDHA*, *SDHB*, *SDHC*, *SDHD*, and *FH*) and

the lack of the neoplastic phenotype in any of these heterozygous mouse models.

We can only speculate that tissue-specific gene-dosage thresholds and mutation-specific effects on encoded protein function might account for such differences. Notably, our observations in the context of heterozygous *SEC23B* mutations offer an important glimpse into how heritable gene mutations interact with the environment (e.g., an exposure leading to ER stress) to result in a particular disease phenotype (here, cancer)—and hence, genetic predisposition to cancer. Indeed, vesicular trafficking is a major contributor to carcinogenesis in that it modulates the trafficking of key proteins involved in cell polarity, motility, invasion, and cell-cell and cell-matrix interactions.⁷⁷ Our finding that heterozygous *SEC23B* germline variants exist in CS and CSL individuals and in individuals with apparently sporadic CS-component cancer is supported by our in vitro data to indicate a role for *SEC23B* in cancer predisposition when germline “change-of-function” variants constitutionally exist in the heterozygous context (Figure 6).

Accession Numbers

Exome sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRA: SRP059300.

Supplemental Data

Supplemental Data include seven figures and eight tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ajhg.2015.10.001>.

Acknowledgments

We are grateful to the members of family 616 and all Cowden syndrome (CS)- and CS-like-affected individuals who contributed to this study. We thank the Genomic Medicine Biorepository of the Cleveland Clinic Genomic Medicine Institute and our database and clinical research teams. We also thank Dr. Thomas LaFramboise and Darcie Seachrist (Case Western Reserve University) and Todd Romigh, Charissa Petersen, and Jessica Altemus (Cleveland Clinic Genomic Medicine Institute) for technical advice and helpful discussions. This study was funded in part by the National Cancer Institute (P01CA124570 and R01CA118989), American Cancer Society (RPG-02-151-01-CCE and Clinical Research Professorship), Breast Cancer Research Foundation, William Randolph Hearst Foundations, and Doris Duke Distinguished Clinical Scientist Award (all to C.E.) and the NHLBI (R01HL094505 to B.Z.). L.Y. is an International Fulbright Science and Technology Doctoral Fellow at the Cleveland Clinic Genomic Medicine Institute and recipient of the Dr. Michael H. Fakhri pre-doctoral endowed scholarship. Y.N. is a CoGEC Scholar funded in part by National Cancer Institute grant R25TCA094186. J.N. was and M.S. is an Ambrose Monell Foundation Cancer Genomic Medicine Fellow at the Cleveland Clinic Genomic Medicine Institute. C.E. is the Sondra J. and Stephen R. Hardis Chair of Cancer Genomic Medicine at the Cleveland Clinic and an American Cancer Society Clinical Research Professor.

Received: July 18, 2015

Accepted: September 30, 2015

Published: October 29, 2015

Web Resources

The URLs for data presented herein are as follows:

ANNOVAR, <http://www.openbioinformatics.org/annovar/>
CADD, <http://cadd.gs.washington.edu/>
Condel, <http://bg.upf.edu/fannsdb/>
COSMIC, <http://cancer.sanger.ac.uk/cosmic/>
GATK, <http://www.broadinstitute.org/gatk/>
Genomic Medicine Institute calculator for *PTEN* CC scores, <http://www.lerner.ccf.org/gmi/ccscore/>
HOPE, <http://www.cmbi.ru.nl/hope/>
Human Protein Atlas, <http://www.proteinatlas.org/>
IGV, <https://www.broadinstitute.org/igv/home/>
MutationTaster, <http://www.mutationtaster.org/>
MutPred Server, <http://mutpred.mutdb.org/>
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, <http://evs.gs.washington.edu/EVS/>
OMIM, <http://www.omim.org/>
OpenEpi, <http://www.openepi.com>
PolyPhen-2, <http://genetics.bwh.harvard.edu/pph2/>
SIFT, <http://sift.jcvi.org/>
TCGA Data Portal, <https://tcga-data.nci.nih.gov/tcga/>
UCSC Genome Browser, <https://genome.ucsc.edu/>

References

1. Marsh, D.J., Coulon, V., Lunetta, K.L., Rocca-Serra, P., Dahia, P.L., Zheng, Z., Liaw, D., Caron, S., Duboué, B., Lin, A.Y., et al. (1998). Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline *PTEN* mutation. *Hum. Mol. Genet.* 7, 507–515.
2. Orloff, M.S., and Eng, C. (2008). Genetic and phenotypic heterogeneity in the *PTEN* hamartoma tumour syndrome. *Oncogene* 27, 5387–5397.
3. Nelen, M.R., Padberg, G.W., Peeters, E.A., Lin, A.Y., van den Helm, B., Frants, R.R., Coulon, V., Goldstein, A.M., van Reen, M.M., Easton, D.F., et al. (1996). Localization of the gene for Cowden disease to chromosome 10q22–23. *Nat. Genet.* 13, 114–116.
4. Liaw, D., Marsh, D.J., Li, J., Dahia, P.L., Wang, S.I., Zheng, Z., Bose, S., Call, K.M., Tsou, H.C., Peacocke, M., et al. (1997). Germline mutations of the *PTEN* gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* 16, 64–67.
5. Zhou, X.P., Waite, K.A., Pilarski, R., Hampel, H., Fernandez, M.J., Bos, C., Dasouki, M., Feldman, G.L., Greenberg, L.A., Ivanovich, J., et al. (2003). Germline *PTEN* promoter mutations and deletions in Cowden/Bannayan-Riley-Ruvalcaba syndrome result in aberrant *PTEN* protein and dysregulation of the phosphoinositol-3-kinase/Akt pathway. *Am. J. Hum. Genet.* 73, 404–411.
6. Eng, C. (2000). Will the real Cowden syndrome please stand up: revised diagnostic criteria. *J. Med. Genet.* 37, 828–830.
7. Pilarski, R., and Eng, C. (2004). Will the real Cowden syndrome please stand up (again)? Expanding mutational and

- clinical spectra of the PTEN hamartoma tumour syndrome. *J. Med. Genet.* *41*, 323–326.
8. Tan, M.H., Mester, J., Peterson, C., Yang, Y., Chen, J.L., Rybicki, L.A., Milas, K., Pederson, H., Remzi, B., Orloff, M.S., and Eng, C. (2011). A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. *Am. J. Hum. Genet.* *88*, 42–56.
 9. Ni, Y., Zbuk, K.M., Sadler, T., Patocs, A., Lobo, G., Edelman, E., Platzer, P., Orloff, M.S., Waite, K.A., and Eng, C. (2008). Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. *Am. J. Hum. Genet.* *83*, 261–268.
 10. Ni, Y., He, X., Chen, J., Moline, J., Mester, J., Orloff, M.S., Ringel, M.D., and Eng, C. (2012). Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. *Hum. Mol. Genet.* *21*, 300–310.
 11. Bennett, K.L., Mester, J., and Eng, C. (2010). Germline epigenetic regulation of KILLIN in Cowden and Cowden-like syndrome. *JAMA* *304*, 2724–2731.
 12. Orloff, M.S., He, X., Peterson, C., Chen, F., Chen, J.L., Mester, J.L., and Eng, C. (2013). Germline PIK3CA and AKT1 mutations in Cowden and Cowden-like syndromes. *Am. J. Hum. Genet.* *92*, 76–80.
 13. Tan, M.H., Mester, J.L., Ngeow, J., Rybicki, L.A., Orloff, M.S., and Eng, C. (2012). Lifetime cancer risks in individuals with germline PTEN mutations. *Clin. Cancer Res.* *18*, 400–407.
 14. Ngeow, J., Mester, J., Rybicki, L.A., Ni, Y., Milas, M., and Eng, C. (2011). Incidence and clinical characteristics of thyroid cancer in prospective series of individuals with Cowden and Cowden-like syndrome characterized by germline PTEN, SDH, or KLLN alterations. *J. Clin. Endocrinol. Metab.* *96*, E2063–E2071.
 15. Goldgar, D.E., Easton, D.F., Cannon-Albright, L.A., and Skolnick, M.H. (1994). Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J. Natl. Cancer Inst.* *86*, 1600–1608.
 16. Hemminki, K., and Vaittinen, P. (1999). Familial cancers in a nationwide family cancer database: age distribution and prevalence. *Eur. J. Cancer* *35*, 1109–1117.
 17. Eng, C. (2000). Familial papillary thyroid cancer—many syndromes, too many genes? *J. Clin. Endocrinol. Metab.* *85*, 1755–1757.
 18. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* *25*, 1754–1760.
 19. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* *20*, 1297–1303.
 20. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* *43*, 491–498.
 21. 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.
 22. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* *38*, e164.
 23. González-Pérez, A., and López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, *Condel*. *Am. J. Hum. Genet.* *88*, 440–449.
 24. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* *31*, 3812–3814.
 25. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* *7*, 248–249.
 26. Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* *39*, e118.
 27. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.
 28. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* *14*, 178–192.
 29. Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O’Daniel, J.M., Ormond, K.E., et al.; American College of Medical Genetics and Genomics (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* *15*, 565–574.
 30. Schwarz, J.M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* *7*, 575–576.
 31. Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D.N., Mooney, S.D., and Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* *25*, 2744–2750.
 32. Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* *46*, 310–315.
 33. Venselaar, H., Te Beek, T.A., Kuipers, R.K., Hekkelman, M.L., and Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* *11*, 548.
 34. Mancias, J.D., and Goldberg, J. (2007). The transport signal on Sec22 for packaging into COPII-coated vesicles is a conformational epitope. *Mol. Cell* *26*, 403–414.
 35. Capes-Davis, A., Theodosopoulos, G., Atkin, I., Drexler, H.G., Kohara, A., MacLeod, R.A., Masters, J.R., Nakamura, Y., Reid, Y.A., Reddel, R.R., and Freshney, R.I. (2010). Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int. J. Cancer* *127*, 1–8.
 36. Tao, J., Zhu, M., Wang, H., Afelik, S., Vasievich, M.P., Chen, X.W., Zhu, G., Jensen, J., Ginsburg, D., and Zhang, B. (2012). SEC23B is required for the maintenance of murine professional secretory tissues. *Proc. Natl. Acad. Sci. USA* *109*, E2001–E2009.
 37. Liang, C.C., Park, A.Y., and Guan, J.L. (2007). In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.* *2*, 329–333.

38. Franken, N.A., Rodermond, H.M., Stap, J., Haveman, J., and van Bree, C. (2006). Clonogenic assay of cells in vitro. *Nat. Protoc.* *1*, 2315–2319.
39. Uhlén, M., Björling, E., Agaton, C., Szgyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C., et al. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* *4*, 1920–1932.
40. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* *347*, 1260419.
41. Lederkremer, G.Z., Cheng, Y., Petre, B.M., Vogan, E., Springer, S., Schekman, R., Walz, T., and Kirchhausen, T. (2001). Structure of the Sec23p/24p and Sec13p/31p complexes of COPII. *Proc. Natl. Acad. Sci. USA* *98*, 10704–10709.
42. Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., and Yao, X. (2009). DOG 1.0: illustrator of protein domain structures. *Cell Res.* *19*, 271–273.
43. Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M.F., Ravazzola, M., Amherdt, M., and Schekman, R. (1994). COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* *77*, 895–907.
44. Fromme, J.C., Orci, L., and Schekman, R. (2008). Coordination of COPII vesicle trafficking by Sec23. *Trends Cell Biol.* *18*, 330–336.
45. Schwarz, K., Iolascon, A., Verissimo, F., Trede, N.S., Horsley, W., Chen, W., Paw, B.H., Hopfner, K.P., Holzmann, K., Russo, R., et al. (2009). Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. *Nat. Genet.* *41*, 936–940.
46. Bianchi, P., Fermo, E., Vercellati, C., Boschetti, C., Barcellini, W., Iurlo, A., Marcello, A.P., Righetti, P.G., and Zanella, A. (2009). Congenital dyserythropoietic anemia type II (CDAII) is caused by mutations in the SEC23B gene. *Hum. Mutat.* *30*, 1292–1298.
47. Punzo, F., Bertoli-Avella, A.M., Scianguetta, S., Della Ragione, F., Casale, M., Ronzoni, L., Cappellini, M.D., Forni, G., Oostra, B.A., and Perrotta, S. (2011). Congenital dyserythropoietic anemia type II: molecular analysis and expression of the SEC23B gene. *Orphanet J. Rare Dis.* *6*, 89.
48. Russo, R., Langella, C., Esposito, M.R., Gambale, A., Vitiello, F., Vallefucio, F., Ek, T., Yang, E., and Iolascon, A. (2013). Hypomorphic mutations of SEC23B gene account for mild phenotypes of congenital dyserythropoietic anemia type II. *Blood Cells Mol. Dis.* *51*, 17–21.
49. Bi, X., Corpina, R.A., and Goldberg, J. (2002). Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. *Nature* *419*, 271–277.
50. Lemoine, N.R., Mayall, E.S., Jones, T., Sheer, D., McDermid, S., Kendall-Taylor, P., and Wynford-Thomas, D. (1989). Characterisation of human thyroid epithelial cells immortalised in vitro by simian virus 40 DNA transfection. *Br. J. Cancer* *60*, 897–903.
51. Russo, R., Gambale, A., Esposito, M.R., Serra, M.L., Troiano, A., De Maggio, I., Capasso, M., Luzzatto, L., Delaunay, J., Tamary, H., and Iolascon, A. (2011). Two founder mutations in the SEC23B gene account for the relatively high frequency of CDA II in the Italian population. *Am. J. Hematol.* *86*, 727–732.
52. Russo, R., Gambale, A., Langella, C., Andolfo, I., Unal, S., and Iolascon, A. (2014). Retrospective cohort study of 205 cases with congenital dyserythropoietic anemia type II: definition of clinical and molecular spectrum and identification of new diagnostic scores. *Am. J. Hematol.* *89*, E169–E175.
53. Satchwell, T.J., Pellegrin, S., Bianchi, P., Hawley, B.R., Gampel, A., Mordue, K.E., Budnik, A., Fermo, E., Barcellini, W., Stephens, D.J., et al. (2013). Characteristic phenotypes associated with congenital dyserythropoietic anemia (type II) manifest at different stages of erythropoiesis. *Haematologica* *98*, 1788–1796.
54. Ulianich, L., Garbi, C., Treglia, A.S., Punzi, D., Miele, C., Raciti, G.A., Beguinot, F., Consiglio, E., and Di Jeso, B. (2008). ER stress is associated with dedifferentiation and an epithelial-to-mesenchymal transition-like phenotype in PC C13 thyroid cells. *J. Cell Sci.* *121*, 477–486.
55. Zhong, Q., Zhou, B., Ann, D.K., Minoo, P., Liu, Y., Banfalvi, A., Krishnaveni, M.S., Dubourd, M., Demaiio, L., Willis, B.C., et al. (2011). Role of endoplasmic reticulum stress in epithelial-mesenchymal transition of alveolar epithelial cells: effects of misfolded surfactant protein. *Am. J. Respir. Cell Mol. Biol.* *45*, 498–509.
56. Tanjore, H., Cheng, D.S., Degryse, A.L., Zoz, D.F., Abdolrasulnia, R., Lawson, W.E., and Blackwell, T.S. (2011). Alveolar epithelial cells undergo epithelial-to-mesenchymal transition in response to endoplasmic reticulum stress. *J. Biol. Chem.* *286*, 30972–30980.
57. Lee, A.S. (2001). The glucose-regulated proteins: stress induction and clinical applications. *Trends Biochem. Sci.* *26*, 504–510.
58. Yu, W., He, X., Ni, Y., Ngeow, J., and Eng, C. (2015). Cowden syndrome-associated germline SDHD variants alter PTEN nuclear translocation through SRC-induced PTEN oxidation. *Hum. Mol. Genet.* *24*, 142–153.
59. Hetz, C. (2012). The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat. Rev. Mol. Cell Biol.* *13*, 89–102.
60. Scriven, P., Coulson, S., Haines, R., Balasubramanian, S., Cross, S., and Wyld, L. (2009). Activation and clinical significance of the unfolded protein response in breast cancer. *Br. J. Cancer* *101*, 1692–1698.
61. Wang, G., Yang, Z.Q., and Zhang, K. (2010). Endoplasmic reticulum stress response in cancer: molecular mechanism and therapeutic potential. *Am. J. Transl. Res.* *2*, 65–74.
62. Lee, E., Nichols, P., Spicer, D., Groshen, S., Yu, M.C., and Lee, A.S. (2006). GRP78 as a novel predictor of responsiveness to chemotherapy in breast cancer. *Cancer Res.* *66*, 7849–7853.
63. Zhang, J., Jiang, Y., Jia, Z., Li, Q., Gong, W., Wang, L., Wei, D., Yao, J., Fang, S., and Xie, K. (2006). Association of elevated GRP78 expression with increased lymph node metastasis and poor prognosis in patients with gastric cancer. *Clin. Exp. Metastasis* *23*, 401–410.
64. Daneshmand, S., Quek, M.L., Lin, E., Lee, C., Cote, R.J., Hawes, D., Cai, J., Groshen, S., Lieskovsky, G., Skinner, D.G., et al. (2007). Glucose-regulated protein GRP78 is up-regulated in prostate cancer and correlates with recurrence and survival. *Hum. Pathol.* *38*, 1547–1552.
65. Wu, X.Y., Fan, R.T., Yan, X.H., Cui, J., Xu, J.L., Gu, H., and Gao, Y.J. (2015). Endoplasmic reticulum stress protects human thyroid carcinoma cell lines against ionizing radiation-induced apoptosis. *Mol. Med. Rep.* *11*, 2341–2347.
66. Khoriaty, R., Vasievich, M.P., Jones, M., Everett, L., Chase, J., Tao, J., Siemieniak, D., Zhang, B., Maillard, I., and Ginsburg, D. (2014). Absence of a red blood cell phenotype in mice

- with hematopoietic deficiency of SEC23B. *Mol. Cell. Biol.* *34*, 3721–3734.
67. Korpala, M., Ell, B.J., Buffa, F.M., Ibrahim, T., Blanco, M.A., Celià-Terrassa, T., Mercatali, L., Khan, Z., Goodarzi, H., Hua, Y., et al. (2011). Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat. Med.* *17*, 1101–1108.
68. Gellera, C., Uziel, G., Rimoldi, M., Zeviani, M., Laverda, A., Carrara, F., and DiDonato, S. (1990). Fumarase deficiency is an autosomal recessive encephalopathy affecting both the mitochondrial and the cytosolic enzymes. *Neurology* *40*, 495–499.
69. Launonen, V., Vierimaa, O., Kiuru, M., Isola, J., Roth, S., Pukkala, E., Sistonen, P., Herva, R., and Aaltonen, L.A. (2001). Inherited susceptibility to uterine leiomyomas and renal cell cancer. *Proc. Natl. Acad. Sci. USA* *98*, 3387–3392.
70. Bourgeron, T., Chretien, D., Poggi-Bach, J., Doonan, S., Rabier, D., Letouzé, P., Munnich, A., Rötig, A., Landrieu, P., and Rustin, P. (1994). Mutation of the fumarase gene in two siblings with progressive encephalopathy and fumarase deficiency. *J. Clin. Invest.* *93*, 2514–2518.
71. Neumann, H.P., Bausch, B., McWhinney, S.R., Bender, B.U., Gimm, O., Franke, G., Schipper, J., Klisch, J., Althoefer, C., Zerres, K., et al.; Freiburg-Warsaw-Columbus Pheochromocytoma Study Group (2002). Germ-line mutations in nonsyndromic pheochromocytoma. *N. Engl. J. Med.* *346*, 1459–1466.
72. Vanharanta, S., Buchta, M., McWhinney, S.R., Virta, S.K., Peçzkowska, M., Morrison, C.D., Lehtonen, R., Januszewicz, A., Järvinen, H., Juhola, M., et al. (2004). Early-onset renal cell carcinoma as a novel extraparaganglial component of SDHB-associated heritable paraganglioma. *Am. J. Hum. Genet.* *74*, 153–159.
73. Burnichon, N., Brière, J.J., Libé, R., Vescovo, L., Rivière, J., Tissier, F., Jouanno, E., Jeunemaitre, X., Bénit, P., Tzagoloff, A., et al. (2010). SDHA is a tumor suppressor gene causing paraganglioma. *Hum. Mol. Genet.* *19*, 3011–3020.
74. Bourgeron, T., Rustin, P., Chretien, D., Birch-Machin, M., Bourgeois, M., Viegas-Péquignot, E., Munnich, A., and Rötig, A. (1995). Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nat. Genet.* *11*, 144–149.
75. Parfait, B., Chretien, D., Rötig, A., Marsac, C., Munnich, A., and Rustin, P. (2000). Compound heterozygous mutations in the flavoprotein gene of the respiratory chain complex II in a patient with Leigh syndrome. *Hum. Genet.* *106*, 236–243.
76. Lepoutre-Lussey, C., Thibault, C., Buffet, A., Morin, A., Badoual, C., Bénit, P., Rustin, P., Ottolenghi, C., Janin, M., Castro-Vega, L.J., et al. (2015). From Nf1 to Sdhb knockout: Successes and failures in the quest for animal models of pheochromocytoma. *Mol. Cell. Endocrinol.* Published online June 27, 2015. <http://dx.doi.org/10.1016/j.mce.2015.06.027>.
77. Goldenring, J.R. (2013). A central role for vesicle trafficking in epithelial neoplasia: intracellular highways to carcinogenesis. *Nat. Rev. Cancer* *13*, 813–820.

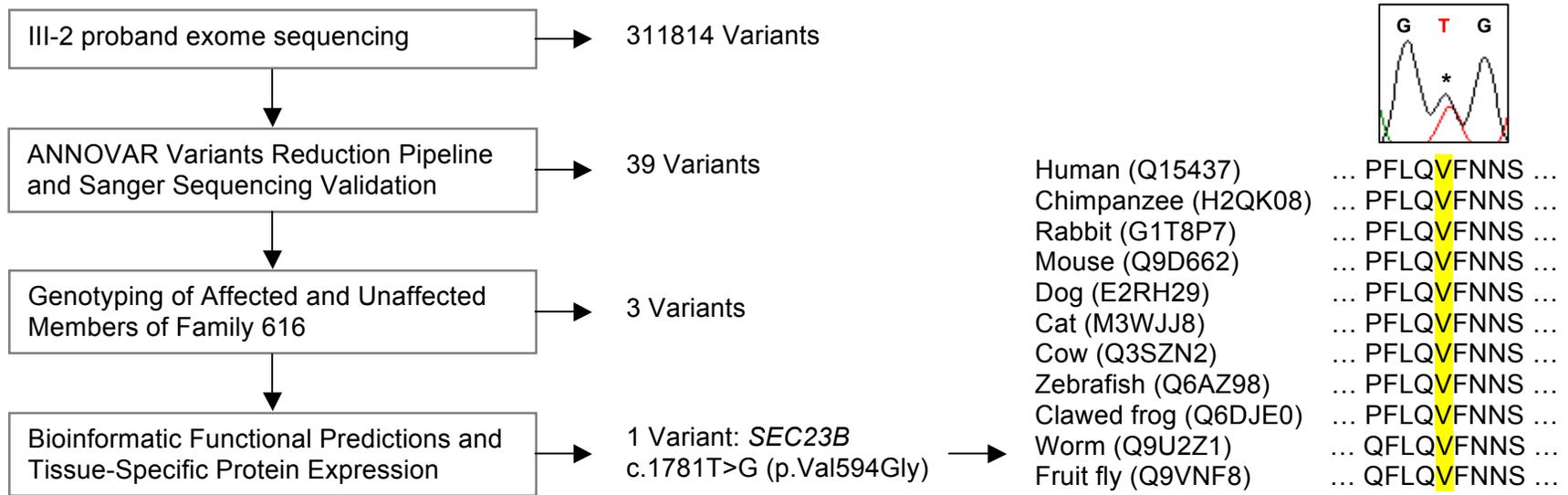
The American Journal of Human Genetics

Supplemental Data

**Germline Heterozygous Variants in *SEC23B*
Are Associated with Cowden Syndrome
and Enriched in Apparently Sporadic Thyroid Cancer**

Lamis Yehia, Farshad Niazi, Ying Ni, Joanne Ngeow, Madhav Sankunny, Zhigang Liu,
Wei Wei, Jessica L. Mester, Ruth A. Keri, Bin Zhang, and Charis Eng

Figure S1. Variant Filtering Strategy in Family 616 and Prioritization of *SEC23B*



Filtering criteria for gene prioritization are on the left and the number of variants retained shown on the right of the workflow. The missense heterozygous variant prioritized in family 616 shows high evolutionary conservation of the affected amino acid residue across ten different species besides human. A Uniprot accession code is given in parentheses for each species and the affected amino acid is highlighted in yellow. Above the sequence is a representative chromatogram from the proband (III-2).

Figure S2. Cellular Phenotype in 293T-*SEC23B* V594G Cells Shows Aberrant *SEC23B* Protein Accumulation, Increased Migration and Upregulation of Expression of Epithelial-to-Mesenchymal Transition Genes

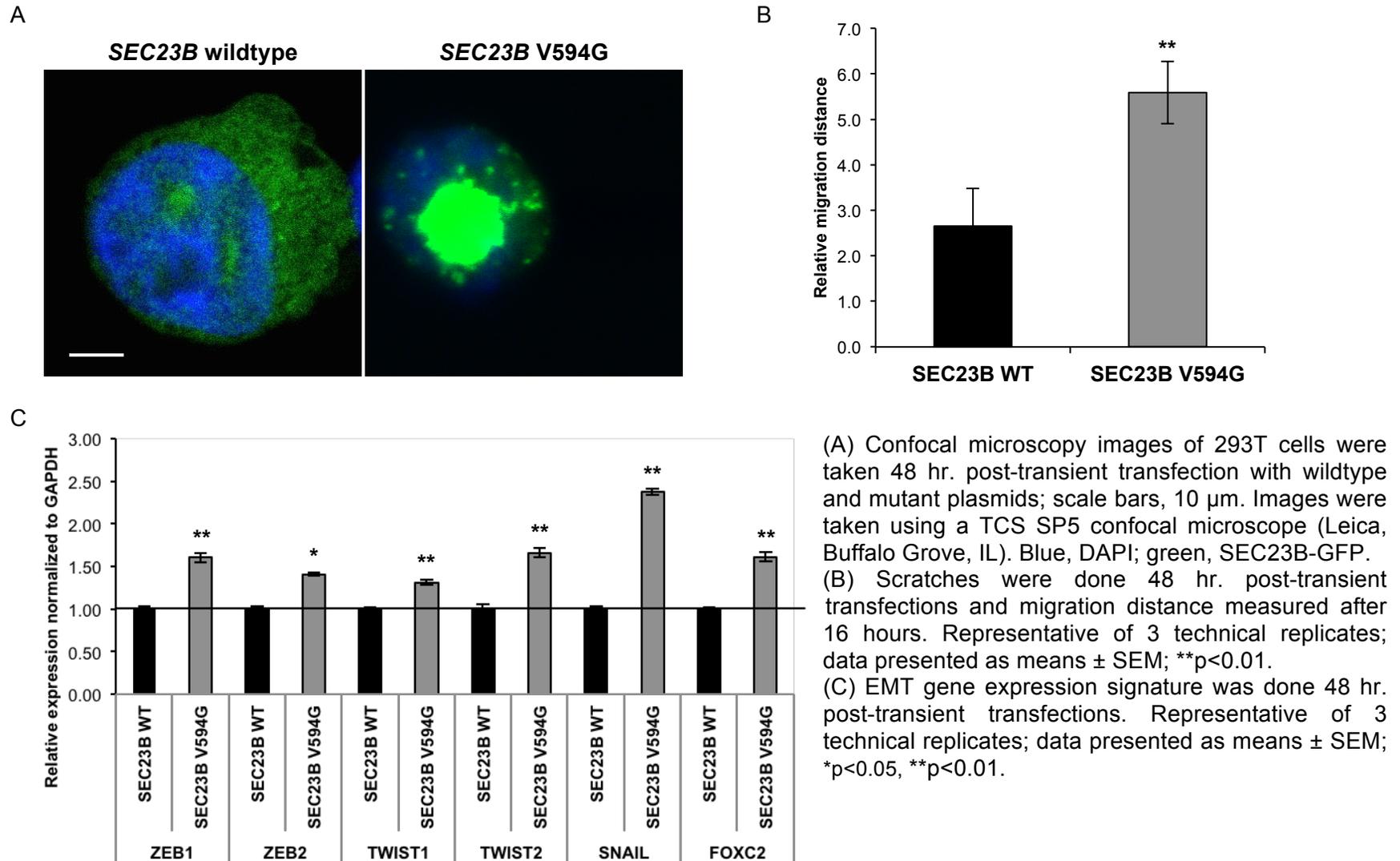
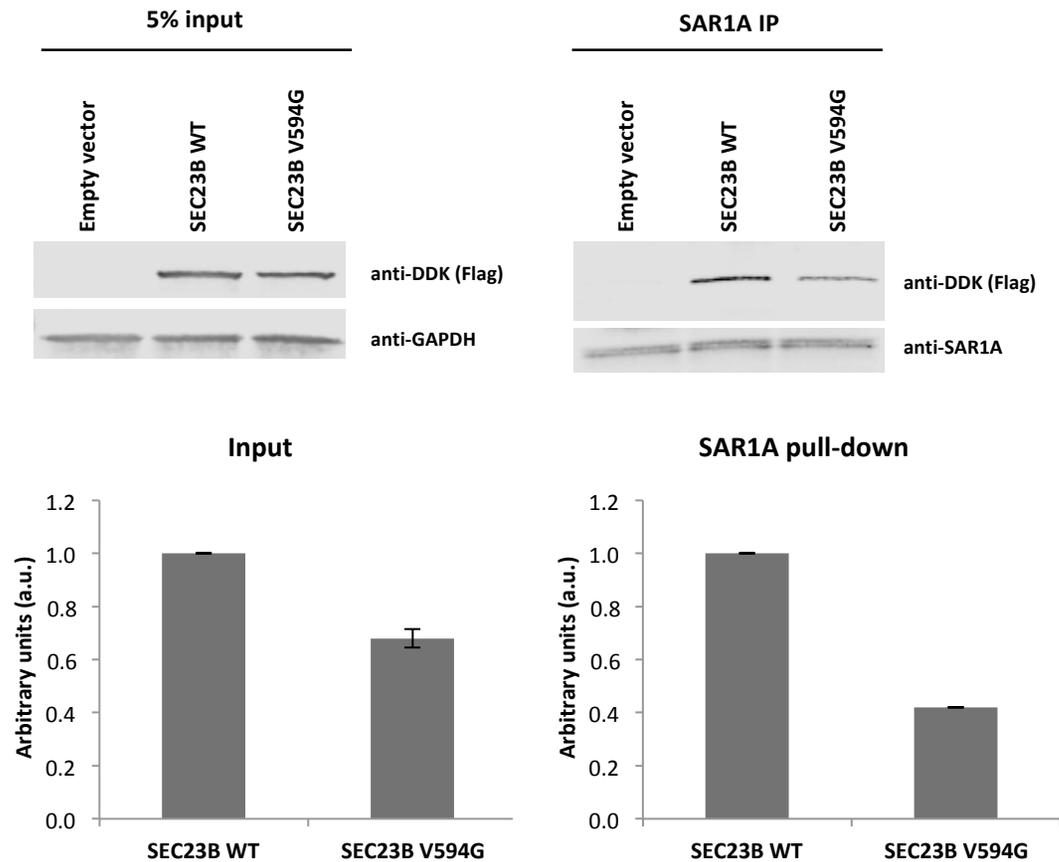
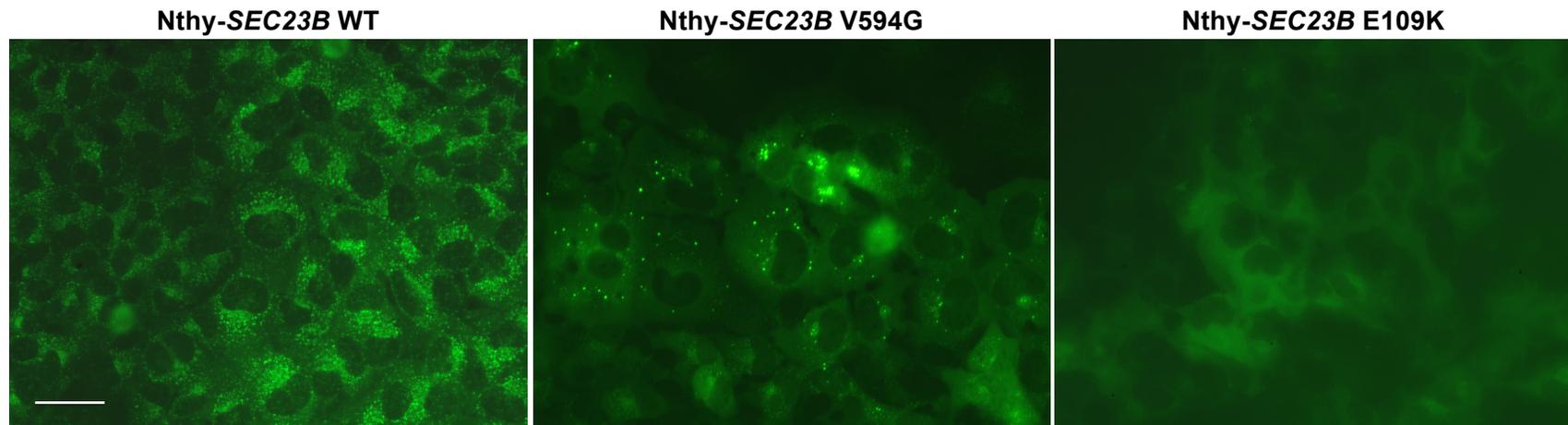


Figure S3: Immunoprecipitation of SEC23B and SAR1A in 293T-SEC23B WT and 293T-SEC23B V594G cells



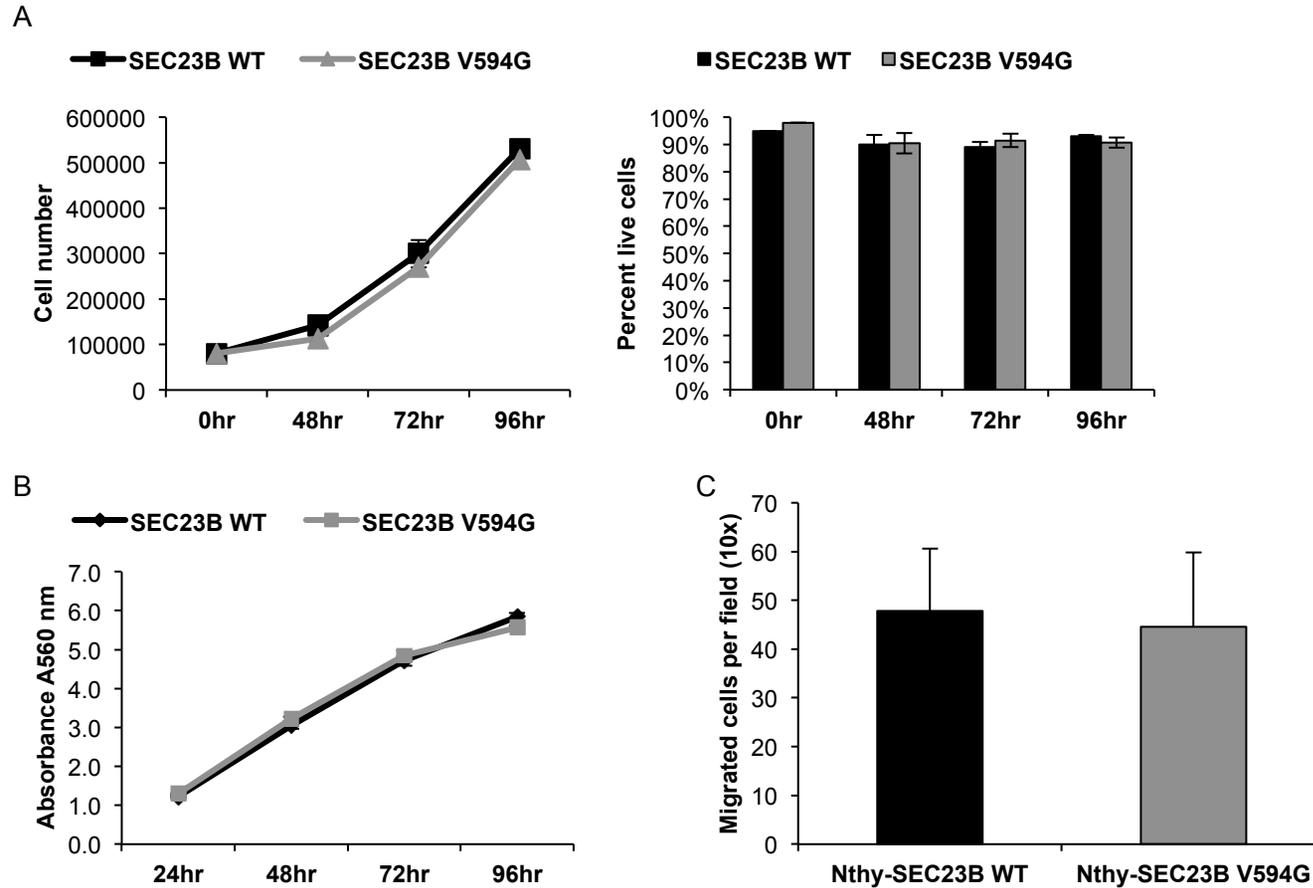
Immunoprecipitation of SEC23B and SAR1A in transiently transfected 293T cells shows decreased interaction of SAR1A with the mutant SEC23B V594G protein. Bands were quantified using Image J software (NIH, Bethesda, MD; <http://imagej.nih.gov/ij/>) and normalized to GAPDH. Representative of n=2 biological replicates with data presented as means \pm SEM.

Figure S4. SEC23B-GFP Expression Pattern in Nthy-*SEC23B* WT, Nthy-*SEC23B* V594G, and Nthy-*SEC23B* E109K Stable Cell Lines



Nthy-*SEC23B* WT cells show an expression pattern typical of ER and Golgi resident proteins. Consistent with our findings in 293T transiently transfected cells, a population of Nthy-*SEC23B* V594G cells (representing the mutation existing in family 616) show aberrant accumulation of SEC23B protein. The Nthy-*SEC23B* E109K cells (representing a very common CDAll founder mutation) show faint cytoplasmic expression in the absence of the typical expression pattern, consistent with loss of SEC23B function. Scale bars, 50 μ m. Images were taken using a Leica DMI3000B manual inverted microscope (Leica, Buffalo Grove, IL).

Figure S5. Nthy-SEC23B WT and Nthy-SEC23B V594G Cells Show Similar Growth, Viability, And Migration



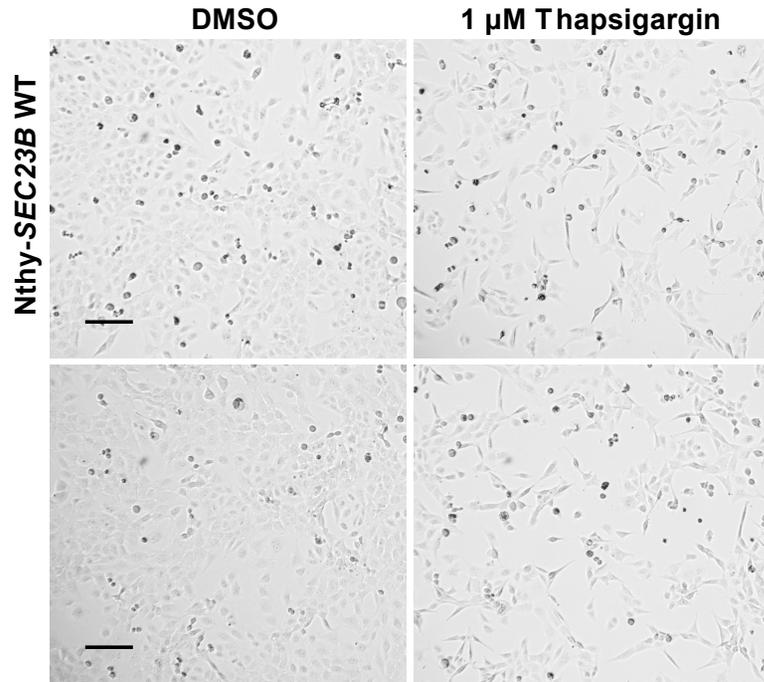
(A) Wildtype and mutant cells were counted for up to 96 hours post seeding. Trypan blue stain was used to count dead cells and assess viability. Data is representative of 3 technical replicates at each time point; data presented as means \pm SEM.

(B) MTT assay done in triplicates for each genotype; data presented as means \pm SEM.

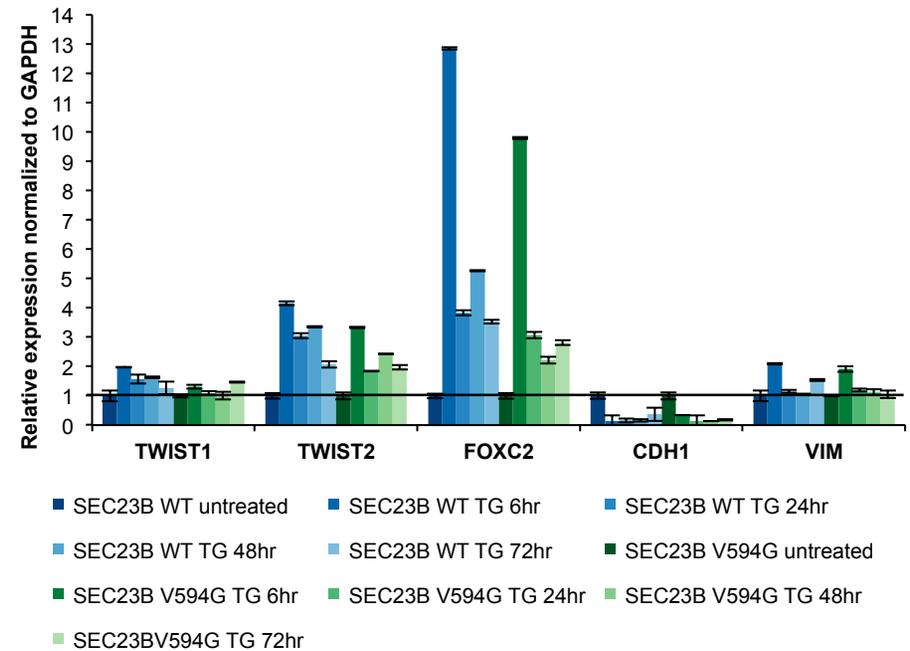
(C) Transwell migration data is representative of n=2 biological replicates done in triplicates for each genotype; data presented as means \pm SEM.

Figure S6. Thapsigargin Induces an Epithelial-to-Mesenchymal Transition (EMT) Phenotype in Wildtype and Mutant Cell Lines

A



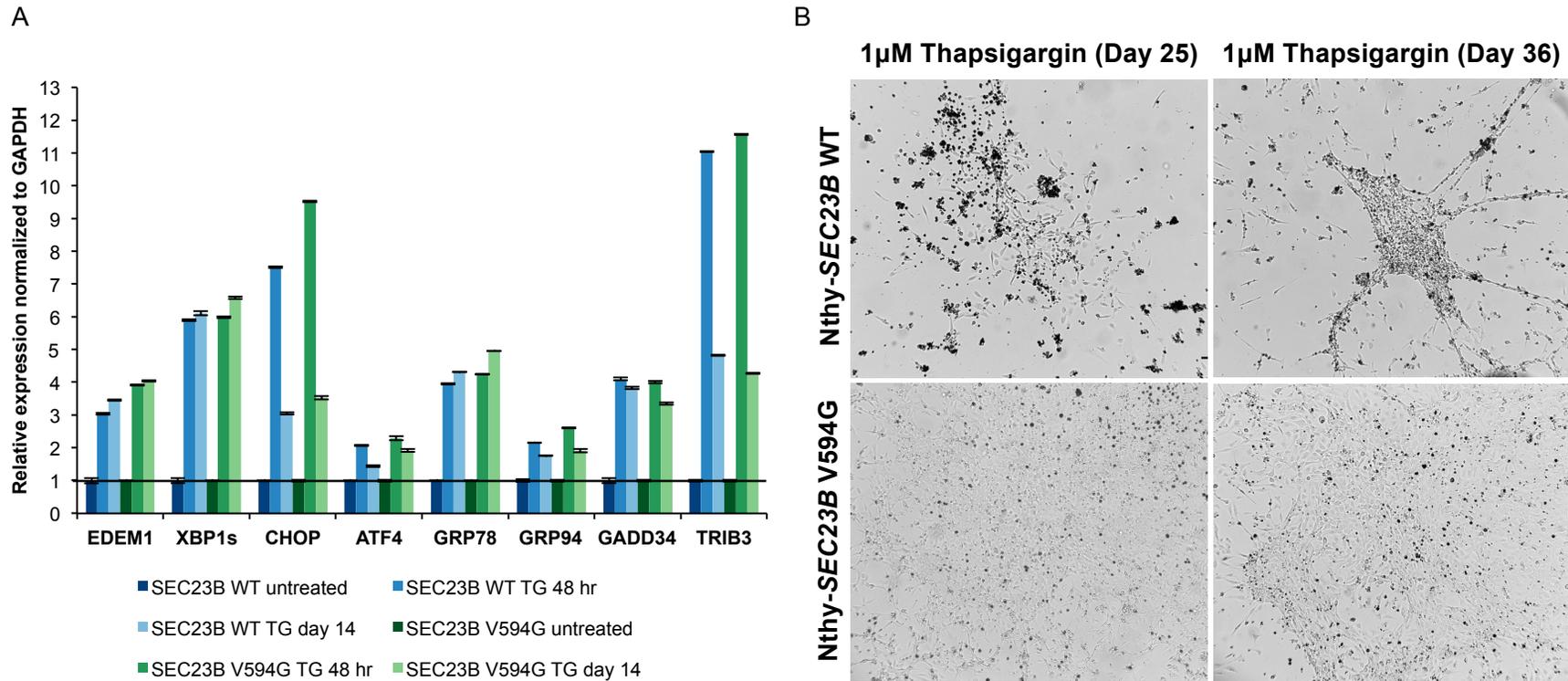
B



(A) Nthy-SEC23B WT and Nthy-SEC23B V594G cells were seeded at equal cell densities, grown overnight, and treated with 1 μM Thapsigargin. A cell morphological change consistent with EMT (spindle-shaped mesenchymal appearance) is evident by 24 hr. post treatment. Scale bars, 200 μm.

(B) EMT signature genes are upregulated at the transcriptional level as early as 6 hr. post treatment with 1 μM Thapsigargin. TG, Thapsigargin. qPCR data representative of 3 technical replicates; data presented as means ± SEM.

Figure S7. ER Stress-Activated Unfolded Protein Response (UPR) Pathway Gene Expression Remains Upregulated at 14 Days Post Thapsigargin Treatment and Nthy-SEC23B V594G Mutant Cell Colonies Survive Chronic ER Stress



(A) UPR activity remains elevated after 14 days of Thapsigargin treatment, suggesting that chronic ER stress had been induced and still exists when we observe phenotype differences between the surviving fractions of wildtype and mutant cells. TG, Thapsigargin. Representative of 3 technical replicates; data presented as means \pm SEM.

(B) Nthy-SEC23B WT and Nthy-SEC23B V594G cells were seeded at equal cell densities, grown overnight, and then treated with 1 μ M Thapsigargin. Epithelial cell colonies sustained for up to 36 days post Thapsigargin treatment. Scale bars, 200 μ m. Representative of 3 biological replicates, with images shown originating from 2 independent experiments.

Table S1. International Cowden Syndrome Consortium Operational Criteria for the Diagnosis of Cowden Syndrome (Ver. 2006)

<u>Pathognomonic</u> Adult Lhermitte-Duclos disease Mucocutaneous lesions Trichilemmomas, facial Acral keratoses Papillomatous papules Mucosal lesions	<u>Major</u> Breast cancer Thyroid cancer (nonmedullary) Macrocephaly (i.e., \geq 97th percentile) Endometrial cancer	<u>Minor</u> Other thyroid lesions (eg, adenoma, multinodular goiter) Mental retardation (i.e., $\text{IQ} \leq 75$) GI hamartomas Fibrocystic breast disease Lipomas Fibromas Genitourinary tumors (especially renal cell carcinoma) Genitourinary malformations Uterine fibroids
<u>Operational diagnosis in an individual</u> Any of following: Mucocutaneous lesions alone, if \geq six facial papules (three of which must be trichilemmomas) Cutaneous facial papules and oral mucosal papillomatosis Oral mucosal papillomatosis and acral keratoses \geq Six palmoplantar keratoses \geq Two major criteria (one of which must be macrocephaly or LDD) One major and \geq three minor criteria \geq Four minor criteria		
<u>Operational diagnosis in a family where one individual is diagnostic for CS</u> Any one pathognomonic criterion Any one major criteria \pm minor criteria Two minor criteria History of Bannayan-Riley-Ruvalcaba syndrome		

Table S2. Validation of Prioritised Variants through Sanger Sequencing in the Proband and Members of Pedigree 616

Genes	Gene variants	Proband	Affected	Affected	Affected	Affected	Unaffected	Unaffected	Unaffected
		III-2	III-4	IV-4	IV-1	II-3	IV-5	IV-3	III-3
<i>ANK2</i>	NM_001148: c.9859A>T, p.Ile3287Phe	+	+		+	+	+	+	
<i>ARHGAP40</i>	NM_001164431: c.872T>C, p.Ile291Thr	+			+	+			
<i>BTBD7</i>	NM_001002860: c.2291C>T, p.Pro764Leu; c.2272_2273insGGAGTTTGAGACCAGAT TGGGCAACATAGGGAAATCC, p.Leu758_Pro759delinsWSLRPDWAT*; c.2275_2288del, p.759_763del	-	NS	NS	NS	NS	NS	NS	NS
<i>C16orf72</i>	NM_014117: c.253T>C, p.Ser85Pro	+	+	+	+	+			
<i>C4orf29</i>	NM_001039717: c.1199G>A, p.Gly400Glu	+	+	+			+		
<i>CBS</i>	NM_001178009: c.832_833insCTGGGGTGGATCATCCAG GTGGGGCTTTTGTGGGCTTGAGCCCT GAAGCCGCGCCCTCTGCAGATCA, p.Ile278_Gly279delinsTGVDHPGGAF	-	NS	NS	NS	NS	NS	NS	NS
<i>DCTN1</i>	NM_001190836: c.1379G>A, p.Arg460His	+	+		+		+	+	
<i>DNAH14</i>	NM_001373: c.5509G>A, p.Gly1837Arg	+							
<i>ENPP4</i>	NM_014936: c.643G>A, p.Gly215Ser	+							
<i>ERMP1</i>	NM_024896: c.2584G>A, p.Val862Met	+	+		+		+	+	
<i>FAM161A</i>	NM_001201543: c.336delA, p.Lys112Asnfs*2	+	+		+		+	+	
<i>FRMD4A</i>	NM_018027: c.2039G>A, p.Arg680Gln	+							
<i>GBF1</i>	NM_001199378: c.2050delT, p.Phe684fs	-	NS	NS	NS	NS	NS	NS	NS
<i>GCNT1</i>	NM_001097634: c.215_216insG, p. Val73Glyfs*4	+	+	+					
<i>GLI2</i>	NM_005270: c.4729G>A, p.Glu1577Lys	+							
<i>GLI3</i>	NM_000168: c.563C>A, p.Ser188Tyr	+	+	+	+			+	
<i>LONRF3</i>	NM_001031855: c.662C>A, p.Pro221Gln	+							
<i>MIB1</i>	NM_020774: c.769G>C, p.Asp257His	+	+	+	+	+		+	
<i>MYO7A</i>	NM_000260: c.3659C>T, p.Pro1220Leu	+			+				
<i>NHSL1</i>	NM_001144060: c.2209C>T, p.Arg737Trp	+			+				
<i>NOX5</i>	NM_001184780: c.1621C>T, p.Arg541Cys	+							

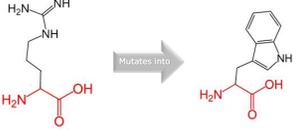
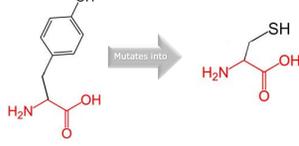
<i>PACRGL</i>	NM_145048: c.338T>G, p.Phe113Cys	+			+				
<i>PARD6B</i>	NM_032521: c.503C>A, p.Pro168His	+	+		+	+	+	+	
<i>PATL1</i>	NM_152716: c.1085G>A, p.Arg362His	+							
<i>PLA2G4E</i>	NM_001206670: c.1499G>A, p.Arg500His	+	+	+				+	
<i>POLE</i>	NM_006231: c.5071C>T, p.Arg1691Cys	+	+		+		+	+	
<i>PRTFDC1</i>	NM_020200: c.612_613insAGTCACCA, p.Tyr205fs	-	NS						
<i>PSKH2</i>	NM_033126: c.572A>G, p.Tyr191Cys	+			+	+			
<i>PTPN2</i>	NM_002828: c.1204G>A, p.Ala402Thr	+	+	+	+	+			
<i>RBM20</i>	NM_001134363: c.3015T>G, p.Asp1005Glu	+							
<i>RDH5</i>	NM_002905: c.524A>T, p.Tyr175Phe	+			+				
<i>RNF215</i>	NM_001017981: c.884G>A, p.Arg295His	+	+	+		+	+		
<i>ROBO2</i>	NM_002942: c.3233C>A, p.Pro1078Gln	+	+				+		
<i>RPS6KB2</i>	NM_003952: c.941G>A, p.Gly314Asp	+			+				
<i>SCN10A</i>	NM_006514: c.4086G>C, p.Gln1362His	+	+			+			
<u><i>SEC23B</i></u>	NM_001172745: c.1781T>G, p.Val594Gly	+	+	+	+	+			
<i>SLC18A1</i>	NM_003053: c.1375T>C, p.Trp459Arg	+			+	+			
<i>SPTBN1</i>	NM_003128: c.3564+1G>T	-	NS						
<i>TIE1</i>	NM_005424: c.710G>A, p.Cys237Tyr	+			+				
<i>TUBB1</i>	NM_030773: c.763G>A, p.Val255Met	+			+				
<i>UTF1</i>	NM_003577: c.962C>T, p.Ala321Val	+							
<i>VPS13C</i>	NM_001018088: c.8972_8975delTGTT, p.Lys2991Argfs*43	+	+						
<i>WDR24</i>	NM_032259: c.2363A>T, p.Glu788Val	+							
<i>ZNF74</i>	NM_001256525: c.632G>T, p.Gly211Val	+	+			+	+		

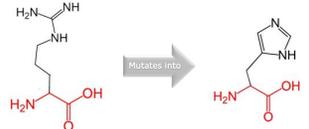
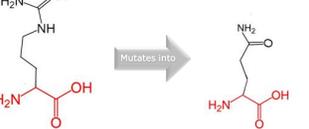
NS = not sequenced since these variants were false positive upon Sanger sequencing in the proband (III-2)
Underlined genes are with variants occurring in all affected members of Family 616

Table S3. Demographic and Clinical Characteristics of CS/CSL Patients (n=96) Selected for *C16orf72*, *PTPN2*, and *SEC23B* Germline Mutation Screening

Demographic and clinical characteristics	Mean or number [range]
<i>Age</i>	55 [18-80]
<i>Sex</i>	
Female	78
Male	18
<i>CC score</i>	12 [1-26]
<i>Thyroid cancer</i>	89
Follicular	25
Papillary	26
Follicular variant papillary	31
Hurthle Cell	3
Not otherwise specified (NOS)	2
Poorly differentiated	1
Anaplastic	1
<i>Benign thyroid</i>	
Goiter	52
Hashimoto's thyroiditis	21
<i>Breast cancer</i>	
Primary invasive	27
Ductal	18
Lobular	3
Mixed	2
NOS	4
Carcinoma <i>in situ</i>	25
<i>Benign breast</i>	43
Atypical ductal hyperplasia	4
Atypical lobular hyperplasia	1
Fibrocystic breast disease	36
Breast papilloma	4
Breast fibroadenoma	5
<i>Female genitourinary cancer</i>	
Uterine cancer	9
Endometrioid	4
NOS	5
<i>Other CS/CSL features</i>	
Macrocephaly	40
Trichilemmoma	11
Acral keratosis	8
Papillomatous papules	17
Lipoma	27
Fibroma	5
Hemangioma	14
Melanoma	3
GI polyps	27
GI cancer	2
Uterine fibroids	30
Renal cell carcinoma	9

Table S4. Structural Effects of Missense *SEC23B* Variants Observed in CS and Apparently Sporadic Thyroid, Breast, and Endometrial Cancers

Missense variants	Structural change	Protein domain and predicted impact on protein structure
c.40C>T, p.Arg14Trp		Residue at the surface of a domain of unknown function Disturbed ionic interaction with glutamic acid at position 13, aspartic acid at position 15
c.74C>A, p.Pro25His		Residue buried in the core of a domain Disturbed core structure of the domain and loss of hydrophobic interactions in the protein core
c.167A>G, p.Tyr56Cys		Residue part of Zinc Finger, Sec23/sec24-Type domain Disturbed interaction with residues in other domains and suspected impact on protein function
c.301A>G, p.Ile101Val		Residue part of Zinc Finger, Sec23/sec24-Type domain Disturbed interaction with residues in other domains and empty space in the protein core due to size differences
c.389T>C, p.Ile130Thr		Residue part of Sec23/sec24, Trunk Domain Empty space in the protein core due to size differences and loss of hydrophobic interactions in the protein core
c.490G>T, p.Val164Leu		Residue part of Sec23/sec24, Trunk Domain Disturbed core structure of the domain and slight destabilization of local secondary structural conformation
c.884C>A, p.Pro295His		Residue part of Sec23/sec24, Trunk Domain Disturbed core structure of the domain and loss of hydrophobic interactions in the protein core
c.985G>T, p.Ala329Ser		Residue part of Sec23/sec24, Trunk Domain Disturbed core structure of the domain and loss of hydrophobic interactions in the protein core

c.1484G>A, p.Arg495His	 <p>The diagram shows the chemical structure of Arginine (left) with its characteristic guanidinium group, which mutates into the structure of Histidine (right) with its imidazole ring. An arrow labeled 'Mutates into' points from the Arginine structure to the Histidine structure.</p>	Residue part of Sec23/sec24 Beta-Sandwich Wildtype residue forms a salt bridge with glutamic acid at position 118, aspartic acid at position 395, aspartic acid at position 399; difference in charge will disturb the ionic interaction
c.1598T>G, p.Val533Gly	 <p>The diagram shows the chemical structure of Valine (left) with its isopropyl side chain, which mutates into the structure of Glycine (right) with a single hydrogen atom as a side chain. An arrow labeled 'Mutates into' points from the Valine structure to the Glycine structure.</p>	Residue part of Sec23/sec24, Helical Domain Disturbed core structure of the domain and loss of hydrophobic interactions in the protein core
c.1636C>T, p.Arg546Trp	 <p>The diagram shows the chemical structure of Arginine (left) with its guanidinium group, which mutates into the structure of Tryptophan (right) with its indole ring system. An arrow labeled 'Mutates into' points from the Arginine structure to the Tryptophan structure.</p>	Residue part of Sec23/sec24, Helical Domain Wildtype residue forms a salt bridge with aspartic acid at position 543; difference in charge will disturb the ionic interaction
c.1661G>A, p.Arg554Gln	 <p>The diagram shows the chemical structure of Arginine (left) with its guanidinium group, which mutates into the structure of Glutamine (right) with its primary amide group. An arrow labeled 'Mutates into' points from the Arginine structure to the Glutamine structure.</p>	Residue part of Sec23/sec24, Helical Domain Wildtype residue forms a salt bridge with cysteine at position 767; difference in charge will disturb the ionic interaction
c.1781T>G, p.Val594Gly	 <p>The diagram shows the chemical structure of Valine (left) with its isopropyl side chain, which mutates into the structure of Glycine (right) with a single hydrogen atom as a side chain. An arrow labeled 'Mutates into' points from the Valine structure to the Glycine structure.</p>	Residue part of Sec23/sec24, Helical Domain Disturbed core structure of the domain and loss of hydrophobic interactions in the protein core
c.2031G>A, p.Met677Ile	 <p>The diagram shows the chemical structure of Methionine (left) with its methylsulfanyl side chain, which mutates into the structure of Isoleucine (right) with its sec-butyl side chain. An arrow labeled 'Mutates into' points from the Methionine structure to the Isoleucine structure.</p>	Residue part of Gelsolin-Like Domain Possible loss of external interactions due to size differences
c.2101C>T, p.Arg701Cys	 <p>The diagram shows the chemical structure of Arginine (left) with its guanidinium group, which mutates into the structure of Cysteine (right) with its thiol side chain. An arrow labeled 'Mutates into' points from the Arginine structure to the Cysteine structure.</p>	Residue part of Gelsolin-Like Domain Wildtype residue forms a hydrogen bond and salt bridge with aspartic acid at position 653; size and charge differences disrupt these interactions resulting in aberrant protein folding

Data extracted from the HOPE server (<http://www.cmbi.ru.nl/hope/>)

Protein Data Bank (PDB) input accession code: Protein transport protein Sec23B - Q15437 (SC23B_HUMAN).

Table S5. Germline *SEC23B*, *C16orf72* and *PTPN2* in TCGA Thyroid, Breast, and Endometrial Cancer Datasets

Cancer datasets	Gene	Variants	Allele frequency ^a	Number of cases	Frequency	
THCA (n=494)	<i>C16orf72</i>	c.416G>A, p.Arg139His	A=9/G=12985	1	0.6%	
		c.365G>A, p.Arg122Gln	A=20/G=12974	2		
	<i>PTPN2</i>	c.31G>C, p.Glu11Gln	0	1	0.2%	
		<i>SEC23B</i>	c.1661G>A, p.Arg554Gln ^b	0	1	4.0%
			c.1598T>G, p.Val533Gly ^b	0	3	
			c.167A>G, p.Tyr56Cys	0	1	
			c.301A>G, p.Ile101Val	0	1	
			c.689+1G>C	0	1	
			c.2101C>T, p.Arg701Cys	0	1	
			c.1636C>T, p.Arg546Trp	T=1/C=13005	1	
c.2031G>A, p.Met677Ile			A=2/G=13004	1		
c.40C>T, p.Arg14Trp			T=3/C=13003	1		
c.1484G>A, p.Arg495His	A=95/G=12911		1			
c.490G>T, p.Val164Leu	T=97/G=12909	8				
BRCA (n=222)	<i>C16orf72</i>	c.365G>A, p.Arg122Gln	A=20/G=12974	3	1.4%	
		<i>PTPN2</i>	c.1159A>G, p.Thr387Ala	0	1	0.5%
	<i>SEC23B</i>		c.884C>A, p.Pro295His	0	1	2.3%
		c.985G>T, p.Ala329Ser	T=4/G=13002	1		
UCEC (n=156)	<i>C16orf72</i>	-	-	0	0.0%	
		<i>PTPN2</i>	c.739G>A, p.Val247Met	T=43/C=12959	1	0.7%
			<i>SEC23B</i>	c.389T>C, p.Ile130Thr	0	1
c.649C>T, p.Arg217*	T=1/C=13005	1				
		c.490G>T, p.Val164Leu	T=97/G=12909	2		

^aAllele frequency data was extracted from the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) Exome Variant Server (<http://evs.gs.washington.edu/EVS/>) v.0.0.28 (accessed May 9, 2015).

^bBoth variants occur in the germline of the same thyroid cancer patient.

Abbreviations: THCA, thyroid cancer; BRCA, breast cancer; UCEC, uterine corpus endometrioid carcinoma.

Table S6. Germline *SEC23B* Variants Identified in the National Heart, Lung, and Blood Institute Exome Sequencing Project (NHLBI-ESP6500)

Chr	Position	ID	Variant type	Variant	Protein	Allele count	MAF	PolyPhen-2 (PP2)	PP2 score
20	18491496	rs369908885	Missense	c.17A>T	p.Glu6Val	2/13,004	0.000154	probably-damaging	0.975
20	18491519	rs121918222	Missense	c.40C>T	p.Arg14Trp	3/13,003	0.000231	probably-damaging	0.999
20	18491553	rs6045440	Missense	c.74C>A	p.Pro25His	5/13,001	0.000384	probably-damaging	1.000
20	18491560	rs139419360	Missense	c.81C>A	p.Ser27Arg	1/13,005	0.000077	probably-damaging	0.985
20	18491567	rs145142114	Missense	c.88G>A	p.Glu30Lys	1/13,005	0.000077	probably-damaging	0.997
20	18491582	rs147576961	Missense	c.103G>A	p.Val35Ile	1/13,005	0.000077	possibly-damaging	0.949
20	18491598	rs376192392	Missense	c.119G>A	p.Cys40Tyr	1/13,005	0.000077	probably-damaging	0.987
20	18491621	rs142919912	Missense	c.142C>T	p.Arg48Cys	2/13,004	0.000154	probably-damaging	0.998
20	18491697	rs141084862	Missense	c.218T>G	p.Leu73Arg	1/13,005	0.000077	possibly-damaging	0.953
20	18492882	rs150263014	Stop-gained	c.235C>T	p.Arg79*	1/12,973	0.000077	-	-
20	18496339	rs121918221	Missense	c.325G>A	p.Glu109Lys	3/13,003	0.000231	probably-damaging	1.000
20	18496372	rs372784283	Missense	c.358G>A	p.Val120Met	1/13,005	0.000077	possibly-damaging	0.650
20	18505120	rs145716758	Missense	c.410C>T	p.Thr137Ile	2/13,004	0.000154	possibly-damaging	0.795
20	18505200	rs36023150	Missense	c.490G>T	p.Val164Leu	97/12,909	0.007458	probably-damaging	0.994
20	18505624	rs121918226	Stop-gained	c.649C>T	p.Arg217*	1/13,005	0.000077	-	-
20	18506532	rs121918224	Stop-gained	c.790C>T	p.Arg264*	1/13,005	0.000077	-	-
20	18506533	rs148239360	Missense	c.791G>A	p.Arg264Gln	1/13,005	0.000077	probably-damaging	1.000
20	18507015	rs371646735	Splice	c.835-2A>G	-	2/13,004	0.000154	-	-
20	18507059	rs375028100	Missense	c.877G>A	p.Gly293Arg	1/13,005	0.000077	probably-damaging	0.994
20	18507167	rs143417821	Missense	c.985G>T	p.Ala329Ser	4/13,002	0.000308	possibly-damaging	0.878
20	18508181	rs140387877	Missense	c.1035C>G	p.His345Gln	1/13,005	0.000077	probably-damaging	0.975
20	18508197	rs150042408	Missense	c.1051G>C	p.Ala351Pro	1/13,005	0.000077	probably-damaging	0.994
20	18516354	rs141002390	Missense	c.1372A>G	p.Thr458Ala	2/13,004	0.000154	possibly-damaging	0.833
20	18516370	-	Frameshift indel	c.1389_1390delTG	p.Phe463Leufs*57	2/12,516	0.000160	-	-
20	18516382	rs150210442	Missense	c.1400A>G	p.Asn467Ser	1/13,005	0.000077	possibly-damaging	0.740
20	18522989	rs373944854	Missense	c.1454C>T	p.Thr485Met	1/13,005	0.000077	probably-damaging	1.000

20	18523025	rs368084647	Missense	c.1490G>A	p.Arg497His	1/13,005	0.000077	probably-damaging	1.000
20	18523740	rs368545054	Missense	c.1589G>A	p.Arg530Gln	1/13,005	0.000077	probably-damaging	1.000
20	18523755	rs377313915	Missense	c.1604G>T	p.Arg535Leu	1/13,005	0.000077	probably-damaging	0.974
20	18523787	rs147135162	Missense	c.1636C>T	p.Arg546Trp	1/13,005	0.000077	probably-damaging	1.000
20	18523799	rs199939108	Stop-gained	c.1648C>T	p.Arg550*	1/13,005	0.000077	-	-
20	18529262	rs150733820	Missense	c.1753C>T	p.His585Tyr	5/13,001	0.000384	possibly-damaging	0.935
20	18529307	rs373800759	Missense	c.1798G>A	p.Asp600Asn	1/13,005	0.000077	probably-damaging	1.000
20	18529382	rs368458175	Missense	c.1873C>A	p.Leu625Ile	1/13,005	0.000077	probably-damaging	0.996
20	18529410	rs372083109	Missense	c.1901C>T	p.Pro634Leu	1/13,005	0.000077	possibly-damaging	0.929
20	18535752	rs374644863	Missense near-splice	c.2149G>T	p.Ala717Ser	1/13,005	0.000077	possibly-damaging	0.859
20	18535813	rs139750050	Missense	c.2210G>T	p.Gly737Val	1/13,005	0.000077	probably-damaging	0.998
20	18541330	rs368409065	Missense	c.2250C>G	p.Ser750Arg	1/13,005	0.000077	probably-damaging	0.999

Data was extracted from the NHLBI-ESP Exome Variant Server (<http://evs.gs.washington.edu/EVS/>) v.0.0.30 (accessed May 30, 2015). We filtered variants to include those in coding regions of *SEC23B* and predicted to be damaging missense variants, splicing variants, and frameshift insertions and deletions, with a minor allele frequency (MAF) < 0.01 (1%) in both African-American and European-American populations. Gene transcripts used for variant calling: NM_001172745 for all variants reported except for NM_032986.3: c.1389_1390delTG.

Abbreviations: Chr, chromosome; MAF, minor allele frequency.

Table S7. Odds Ratios for Presence of *SEC23B* Variants in Different Patient Populations

MAF ^a	Category	<i>SEC23B</i> variant		Frequency	OR	95% CI	P-value
		+	-				
≤ 0.008	NHLBI-ESP	155	12,851	0.0121	1.000	0.7990 - 1.252	>0.9999
	TCGA-THCA	20	986	0.0203	1.682	1.026 - 2.645	0.0401
	TCGA-BRCA	5	439	0.0114	0.9443	0.3402 - 2.146	0.9533
	TCGA-UCEC	4	308	0.0130	1.077	0.3364 - 2.665	0.8303
	TCGA-all	29	1,715	0.0170	1.402	0.9255 - 2.068	0.1068
≤ 0.0004	NHLBI-ESP	58	12,948	0.0045	1.000	0.6931 - 1.443	>0.9999
	TCGA-THCA	11	977	0.0113	2.513	1.256 - 4.676	0.0116
	TCGA-BRCA	3	441	0.0070	1.519	0.3752 - 4.333	0.4689
	TCGA-UCEC	2	310	0.0065	1.440	0.2354 - 4.974	0.5778
	TCGA-all	16	1,728	0.0093	2.067	1.153 - 3.546	0.0166

^aMAF cut-off derivation: we found both unique (absent in public databases) and previously reported *SEC23B* variants in our cancer patient populations (CS and TCGA). Previously reported variants had $0.00008 \leq \text{MAF} \leq 0.008$. $\text{MAF} \leq 0.008$ was hence set as the upper limit to indicate all observed variants. The next most common variants had a MAF of 0.0004 (5/13,001), hence used as the upper cut-off threshold for more rare variants ($0.00008 \leq \text{MAF} \leq 0.0004$).

Abbreviations: MAF, minor allele frequency; OR, odds ratio; 95% CI, 95% confidence interval; THCA, thyroid cancer; BRCA, breast cancer; UCEC, uterine corpus endometrioid carcinoma; all, THCA, BRCA, and UCEC.

Table S8. Hematologic Phenotypes of CS Cancer Patients with Identified Germline *SEC23B* Variants

Patients	III-2	III-4	IV-1	IV-4	CCF05664	CCF04372
<i>SEC23B</i> variant	c.1781T>G, p.Val594Gly	c.1781T>G, p.Val594Gly	c.1781T>G, p.Val594Gly	c.1781T>G, p.Val594Gly	c.490G>T, p.Val164Leu	c.1512T>C, p.(=)
Cancer (age)	FvPTC (43)	FvPTC (51)	PTC (21)	Endometrial ca. (35)	PTC (45)	PTC (45)
Age at CBC	59	63	35	35	65	45
RBC count (10 ⁶ /μL)	4.09 (3.80-5.10)	3.41 (NA)	4.72 (3.77-5.28)	3.05 (3.80-5.20)	4.51 (3.95-5.40)	4.60 (3.80-5.10)
Hb (g/dL)	13.0 (11.7-15.5)	11.9 (NA)	14.4 (11.1-15.9)	8.6 (11.6-15.5)	14.6 (11.9-15.9)	13.7 (11.7-15.5)
MCV (fL)	94 (80-100)	101 (NA)	90 (79-97)	87 (80-100)	95 (78-95)	87.8 (81-100)
Total bilirubin (mg/gL)	-	-	0.4 (0.0-1.2)	-	1.1 (0.2-1.3)	-
Comments	Normal	Abnormal WBC differential	Normal	Concurrent with prolonged bleeding (metromenorrhagia)	Normal	Normal

CDAll mean observed levels (n=patients)*:

RBC (106/μl): 3.2 ± 0.1 (1.2-5.8; n=93)

Hb (g/dl): 9.6 ± 0.2 (3.6-16.4; n=161)

MCV (fL): 87.3 ± 1.0 (60.0-114.0; n=98)

Total bilirubin: 2.8 ± 0.2 (0.2-12.7; n=108)

*Iolascon A. et al. American Journal of Hematology 89 (10): E169-E175 (2014)

Abbreviations: FvPTC, follicular variant PTC; PTC, papillary thyroid cancer; ca., carcinoma; CBC, complete blood count; Hb, hemoglobin; MCV, mean cell volume; NA, not available; WBC, white blood cell.