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*

| III-2 proband exome sequencing | → | 311814 Variants |

| ANNOVAR Variants Reduction Pipeline and Sanger Sequencing Validation | → | 39 Variants |

| Genotyping of Affected and Unaffected Members of Family 616 | → | 3 Variants |

| Bioinformatic Functional Predictions and Tissue-Specific Protein Expression | → | 1 Variant: *SEC23B* c.1781T>G (p.Val594Gly) → |

Human (Q15437)   … PFLQ**V**FNNS …
Chimpanzee (H2QK08) … PFLQ**V**FNNS …
Rabbit (G1T8P7)   … PFLQ**V**FNNS …
Mouse (Q9D662)   … PFLQ**V**FNNS …
Dog (E2RH29)   … PFLQ**V**FNNS …
Cat (M3WJJ8)   … PFLQ**V**FNNS …
Cow (Q3SZN2)   … PFLQ**V**FNNS …
Zebrafish (Q6AZ98)   … PFLQ**V**FNNS …
Clawed frog (Q6DJE0) … PFLQ**V**FNNS …
Worm (Q9U2Z1)   … QFLQ**V**FNNS …
Fruit fly (Q9VNF8)   … QFLQ**V**FNNS …

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
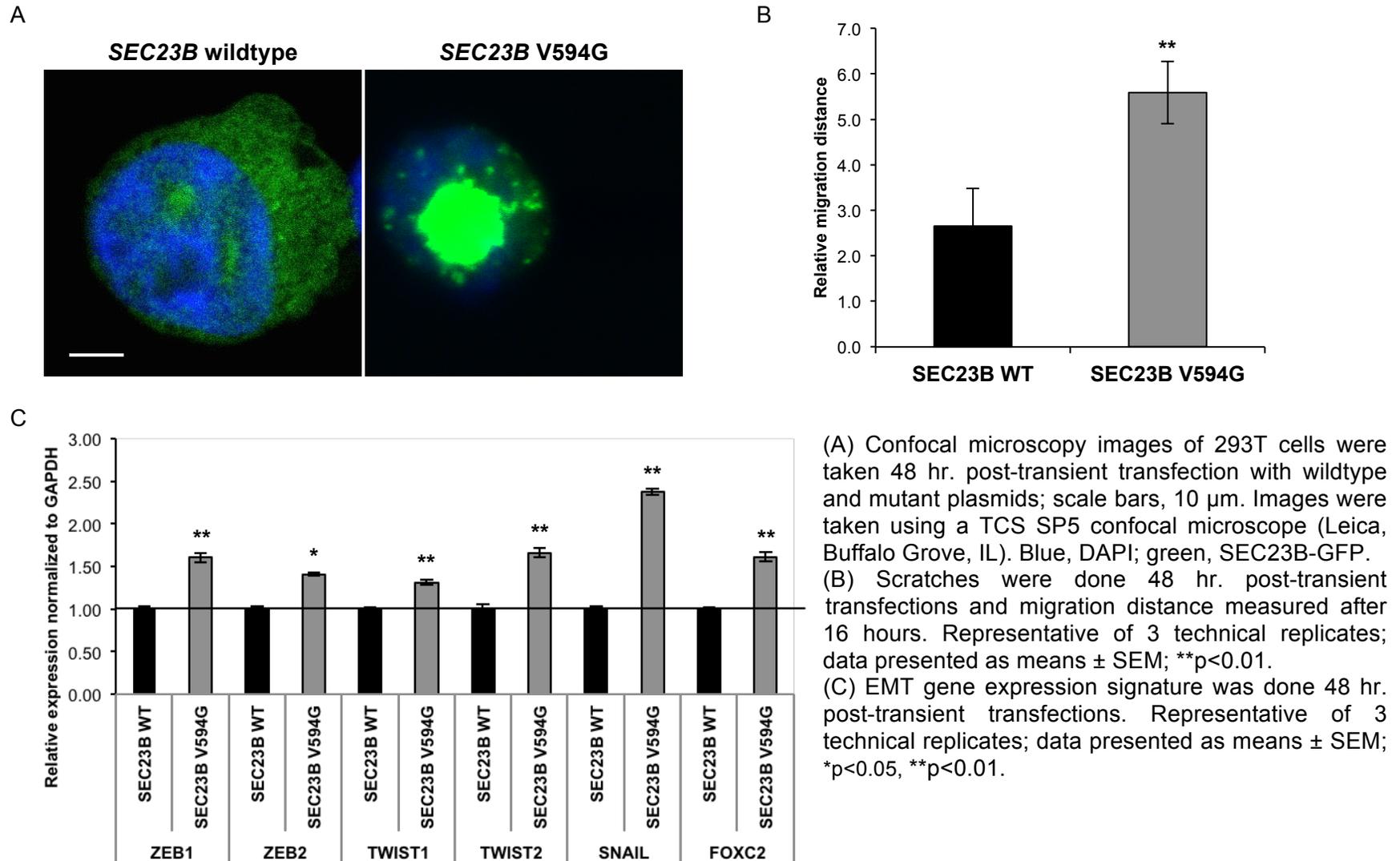
A

SEC23B wildtype | *SEC23B* V594G

B

C

(A) Confocal microscopy images of 293T cells were taken 48 hr. post-transient transfection with wildtype and mutant plasmids; scale bars, 10 μm. Images were taken using a TCS SP5 confocal microscope (Leica, Buffalo Grove, IL). Blue, DAPI; green, SEC23B-GFP.

(B) Scratches were done 48 hr. post-transient transfections and migration distance measured after 16 hours. Representative of 3 technical replicates; data presented as means ± SEM; **$p<0.01$.

(C) EMT gene expression signature was done 48 hr. post-transient transfections. Representative of 3 technical replicates; data presented as means ± SEM; *$p<0.05$, **$p<0.01$.

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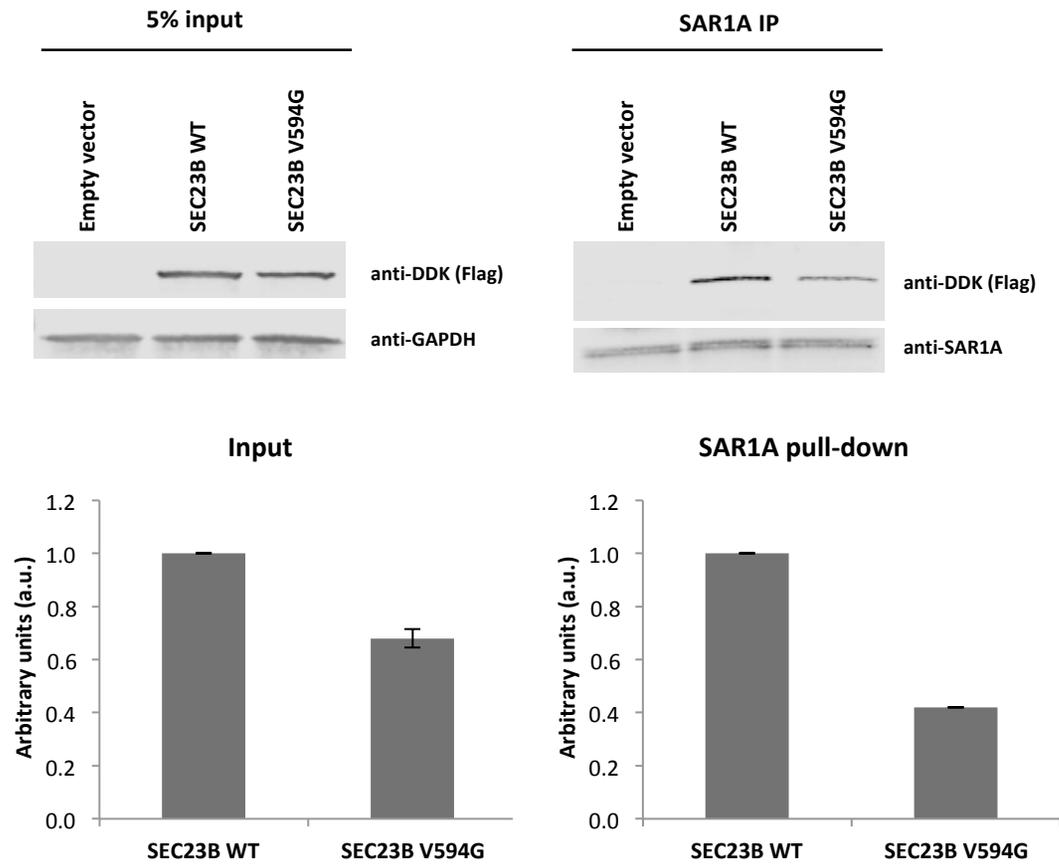

Immunoprecipitation of SEC23B and SAR1A in transiently transfected 293T cells shows decreased interaction of SAR1A with the mutant SEC23B V594G protein. Bands were quantified using Image J software (NIH, Bethesda, MD; http://imagej.nih.gov/ij/) and normalized to GAPDH. Representative of n=2 biological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

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

Nthy-*SEC23B* WT	**Nthy-*SEC23B* V594G**	**Nthy-*SEC23B* E109K**
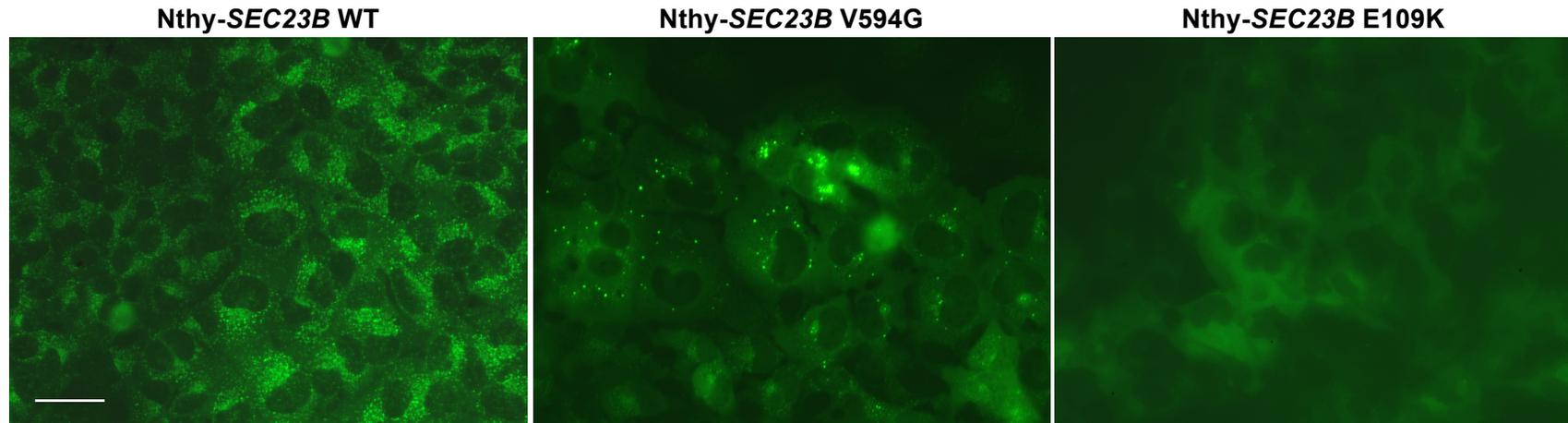
Nthy-*SEC23B* WT cells show an expression pattern typical of ER and Golgi resident proteins. Consistent with our findings in 293T transiently transfected cells, a population of Nthy-*SEC23B* V594G cells (representing the mutation existing in family 616) show aberrant accumulation of SEC23B protein. The Nthy-*SEC23B* E109K cells (representing a very common CDAII founder mutation) show faint cytoplasmic expression in the absence of the typical expression pattern, consistent with loss of SEC23B function. Scale bars, 50 µm. 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

B

C
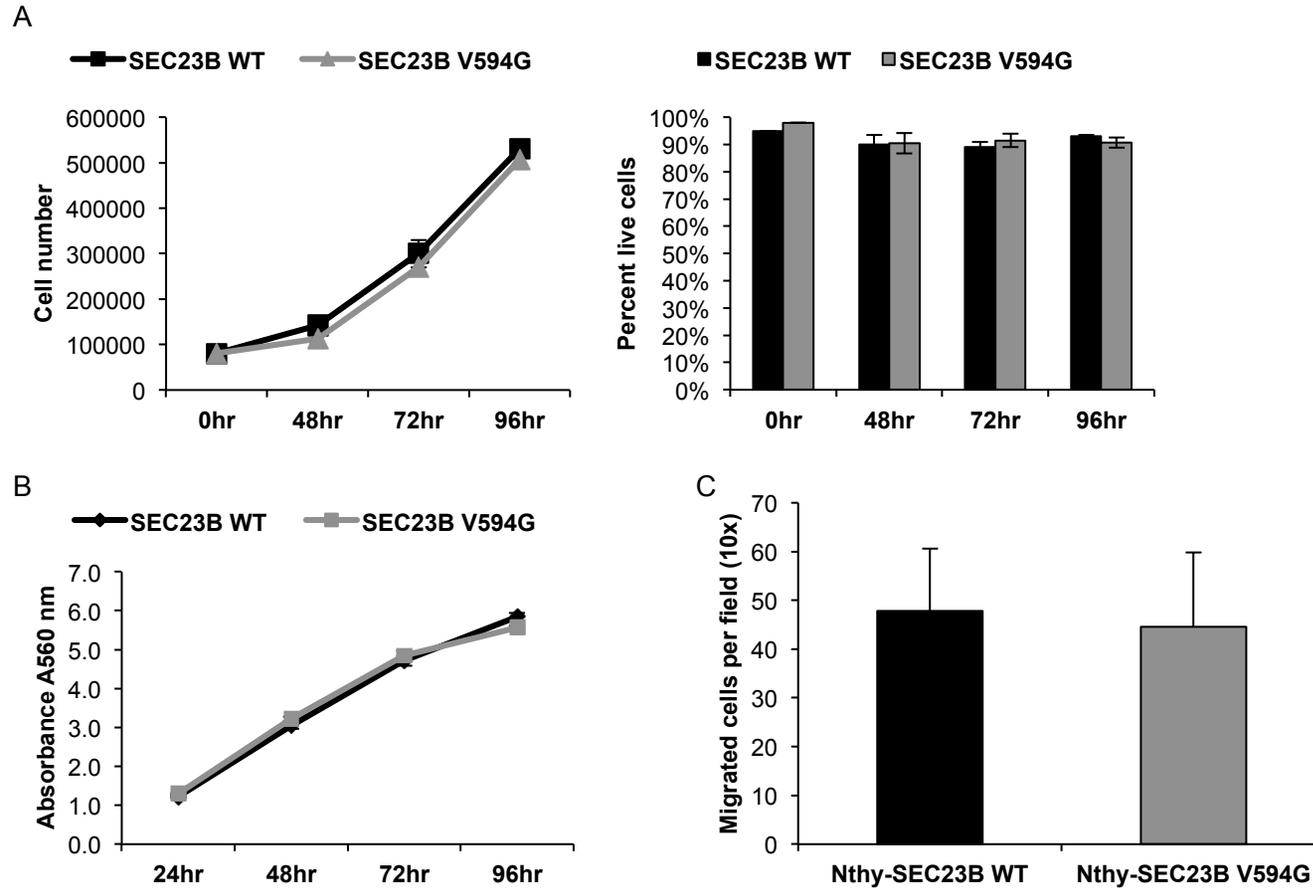
(A) Wildtype and mutant cells were counted for up to 96 hours post seeding. Trypan blue stain was used to count dead cells and assess viability. Data is representative of 3 technical replicates at each time point; data presented as means ± SEM.
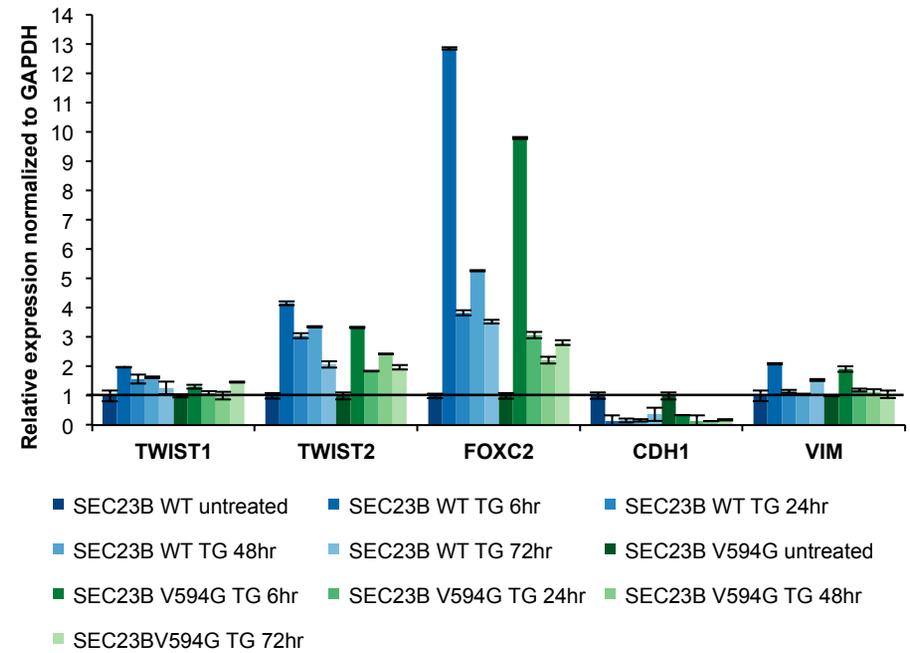(B) MTT assay done in triplicates for each genotype; data presented as means ± SEM.
(C) Transwell migration data is representative of n=2 biological replicates done in triplicates for each genotype; data presented as means ± SEM.

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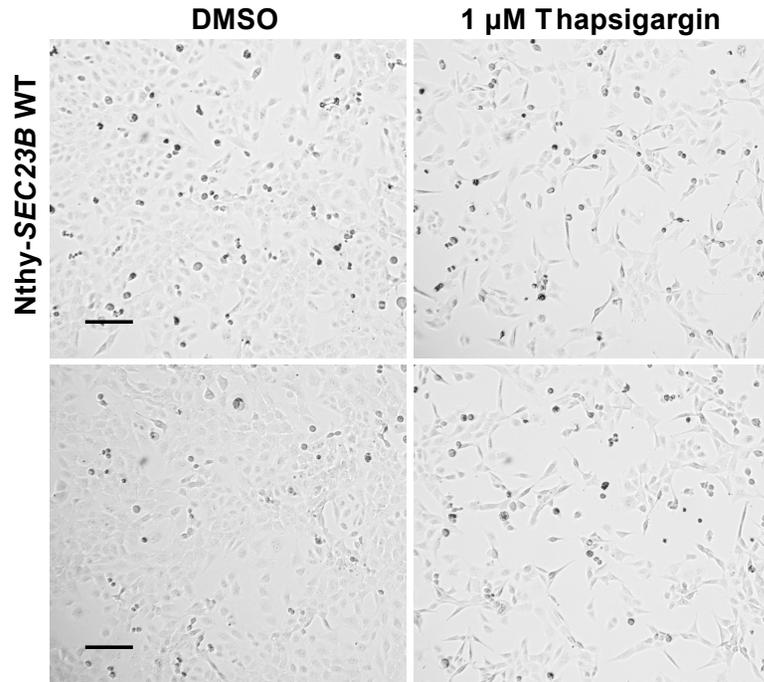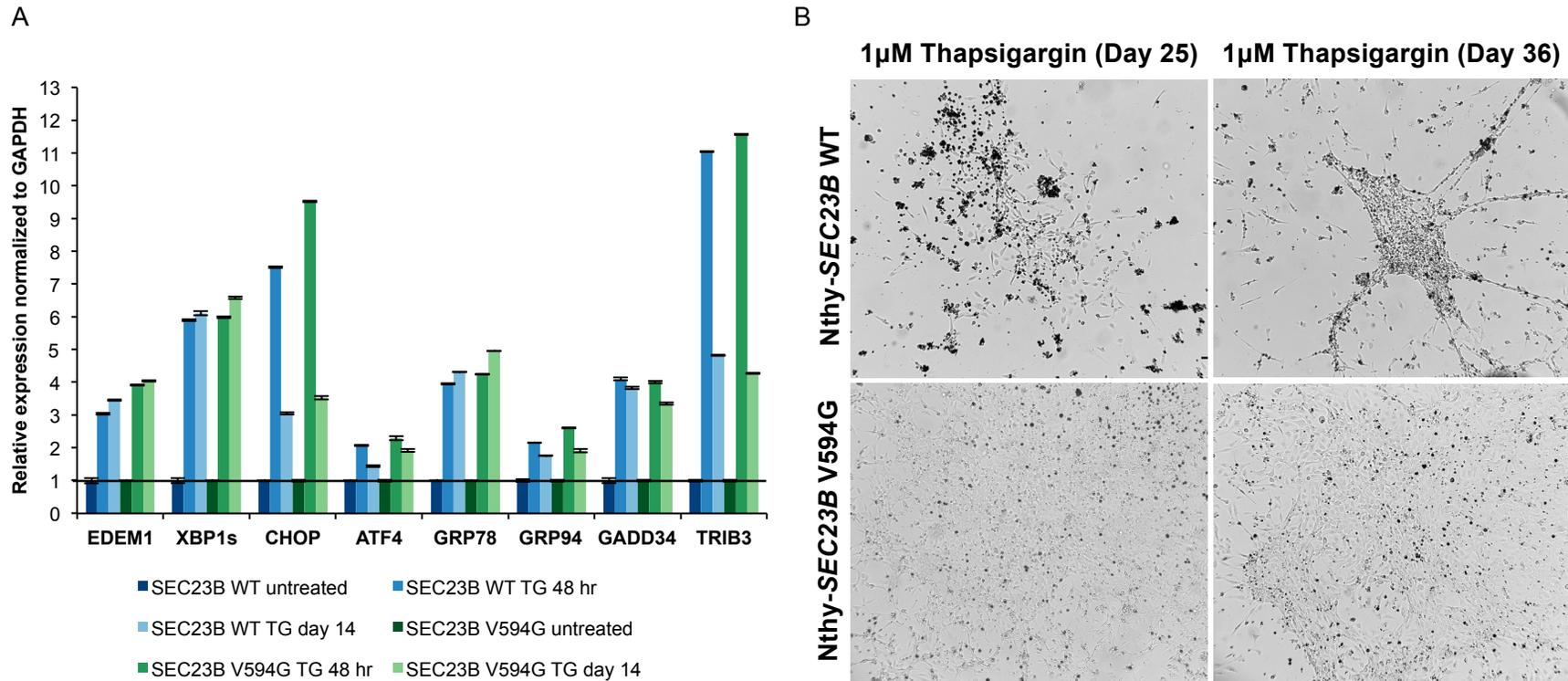

(A) Nthy-SEC23B WT and Nthy-SEC23B V594G cells were seeded at equal cell densities, grown overnight, and treated with 1 μM Thapsigargin. A cell morphological change consistent with EMT (spindle-shaped mesenchymal appearance) is evident by 24 hr. post treatment. Scale bars, 200 μm.

(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 ± SEM.
(B) Nthy-SEC23B WT and Nthy-SEC23B V594G cells were seeded at equal cell densities, grown overnight, and then treated with 1µM Thapsigargin. Epithelial cell colonies sustained for up to 36 days post Thapsigargin treatment. Scale bars, 200 µm. 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)

Pathognomonic	Major	Minor
Adult Lhermitte-Duclos disease	Breast cancer	Other thyroid lesions (eg, adenoma, multinodular goiter)
Mucocutaneous lesions	Thyroid cancer (nonmedullary)	Mental retardation (i.e., IQ ≤ 75)
Trichilemmomas, facial	Macrocephaly (i.e., ≥ 97th percentile)	GI hamartomas
Acral keratoses	Endometrial cancer	Fibrocystic breast disease
Papillomatous papules		Lipomas
Mucosal lesions		Fibromas
		Genitourinary tumors (especially renal cell carcinoma)
		Genitourinary malformations
		Uterine fibroids

Operational diagnosis in an individual
Any of following:
  Mucocutaneous lesions alone, if ≥ six facial papules (three of which must be trichilemmomas)
  Cutaneous facial papules and oral mucosal papillomatosis
  Oral mucosal papillomatosis and acral keratoses
  ≥ Six palmoplantar keratoses
≥ Two major criteria (one of which must be macrocephaly or LDD)
One major and ≥ three minor criteria
≥ Four minor criteria

Operational diagnosis in a family where one individual is diagnostic for CS
Any one pathognomonic criterion
Any one major criteria ± 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
ANK2	NM_001148: c.9859A>T, p.Ile3287Phe	+	+		+	+	+	+	
ARHGAP40	NM_001164431: c.872T>C, p.Ile291Thr	+			+	+			
BTBD7	NM_001002860: c.2291C>T, p.Pro764Leu; c.2272_2273insGGAGTTTGAGACCAGATTGGGCAACATAGGGAAATCC, p.Leu758_Pro759delinsWSLRPDWAT*; c.2275_2288del, p.759_763del	-	NS	NS	NS	NS	NS	NS	NS
C16orf72	NM_014117: c.253T>C, p.Ser85Pro	+	+	+	+	+			
C4orf29	NM_001039717: c.1199G>A, p.Gly400Glu	+	+	+			+		
CBS	NM_001178009: c.832_833insCTGGGGTGGATCATCCAGGTGGGGCTTTTGCTGGGCTTGAGCCCTGAAGCCGCGCCCTCTGCAGATCA, p.Ile278_Gly279delinsTGVDHPGGAF	-	NS	NS	NS	NS	NS	NS	NS
DCTN1	NM_001190836: c.1379G>A, p.Arg460His	+	+		+		+	+	
DNAH14	NM_001373: c.5509G>A, p.Gly1837Arg	+							
ENPP4	NM_014936: c.643G>A, p.Gly215Ser	+							
ERMP1	NM_024896: c.2584G>A, p.Val862Met	+	+		+		+	+	
FAM161A	NM_001201543: c.336delA, p.Lys112Asnfs*2	+	+		+		+	+	
FRMD4A	NM_018027: c.2039G>A, p.Arg680Gln	+							
GBF1	NM_001199378: c.2050delT, p.Phe684fs	-	NS	NS	NS	NS	NS	NS	NS
GCNT1	NM_001097634: c.215_216insG, p.Val73Glyfs*4	+	+	+					
GLI2	NM_005270: c.4729G>A, p.Glu1577Lys	+							
GLI3	NM_000168: c.563C>A, p.Ser188Tyr	+	+	+	+			+	
LONRF3	NM_001031855: c.662C>A, p.Pro221Gln	+							
MIB1	NM_020774: c.769G>C, p.Asp257His	+	+	+	+	+		+	
MYO7A	NM_000260: c.3659C>T, p.Pro1220Leu	+			+				
NHSL1	NM_001144060: c.2209C>T, p.Arg737Trp	+			+				
NOX5	NM_001184780: c.1621C>T, p.Arg541Cys	+							

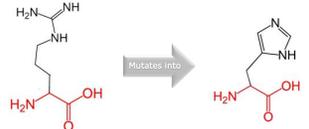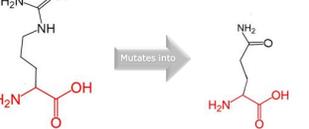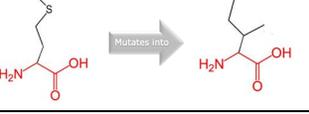| Gene	Variant	C1	C2	C3	C4	C5	C6	C7	C8
PACRGL	NM_145048: c.338T>G, p.Phe113Cys	+			+				
PARD6B	NM_032521: c.503C>A, p.Pro168His	+	+		+	+	+	+	
PATL1	NM_152716: c.1085G>A, p.Arg362His	+							
PLA2G4E	NM_001206670: c.1499G>A, p.Arg500His	+	+	+				+	
POLE	NM_006231: c.5071C>T, p.Arg1691Cys	+	+		+		+	+	
PRTFDC1	NM_020200: c.612_613insAGTCACCA, p.Tyr205fs	-	NS	NS	NS	NS	NS	NS	NS
PSKH2	NM_033126: c.572A>G, p.Tyr191Cys	+			+	+			
<u>*PTPN2*</u>	NM_002828: c.1204G>A, p.Ala402Thr	+	+	+	+	+			
RBM20	NM_001134363: c.3015T>G, p.Asp1005Glu	+							
RDH5	NM_002905: c.524A>T, p.Tyr175Phe	+			+				
RNF215	NM_001017981: c.884G>A, p.Arg295His	+	+	+			+	+	
ROBO2	NM_002942: c.3233C>A, p.Pro1078Gln	+	+					+	
RPS6KB2	NM_003952: c.941G>A, p.Gly314Asp	+			+				
SCN10A	NM_006514: c.4086G>C, p.Gln1362His	+	+				+		
<u>*SEC23B*</u>	NM_001172745: c.1781T>G, p.Val594Gly	+	+	+	+	+			
SLC18A1	NM_003053: c.1375T>C, p.Trp459Arg	+			+	+			
SPTBN1	NM_003128: c.3564+1G>T	-	NS	NS	NS	NS	NS	NS	NS
TIE1	NM_005424: c.710G>A, p.Cys237Tyr	+			+				
TUBB1	NM_030773: c.763G>A, p.Val255Met	+			+				
UTF1	NM_003577: c.962C>T, p.Ala321Val	+							
VPS13C	NM_001018088: c.8972_8975delTGTT, p.Lys2991Argfs*43	+	+						
WDR24	NM_032259: c.2363A>T, p.Glu788Val	+							
ZNF74	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]
Age	55 [18-80]
Sex	
Female	78
Male	18
CC score	12 [1-26]
Thyroid cancer	89
Follicular	25
Papillary	26
Follicular variant papillary	31
Hurthle Cell	3
Not otherwise specified (NOS)	2
Poorly differentiated	1
Anaplastic	1
Benign thyroid	
Goiter	52
Hashimoto's thyroiditis	21
Breast cancer	
Primary invasive	27
Ductal	18
Lobular	3
Mixed	2
NOS	4
Carcinoma *in situ*	25
Benign breast	43
Atypical ductal hyperplasia	4
Atypical lobular hyperplasia	1
Fibrocystic breast disease	36
Breast papilloma	4
Breast fibroadenoma	5
Female genitourinary cancer	
Uterine cancer	9
Endometrioid	4
NOS	5
Other CS/CSL features	
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
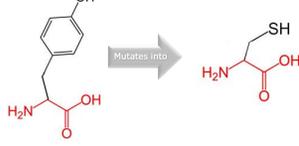
c.1484G>A, p.Arg495His		Residue part of Sec23/sec24 Beta-Sandwich Wildtype residue forms a salt bridge with glutamic acid at position 118, aspartic acid at position 395, aspartic acid at position 399; difference in charge will disturb the ionic interaction
c.1598T>G, p.Val533Gly		Residue part of Sec23/sec24, Helical Domain Disturbed core structure of the domain and loss of hydrophobic interactions in the protein core
c.1636C>T, p.Arg546Trp		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		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		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		Residue part of Gelsolin-Like Domain Possible loss of external interactions due to size differences
c.2101C>T, p.Arg701Cys		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)	C16orf72	c.416G>A, p.Arg139His c.365G>A, p.Arg122Gln	A=9/G=12985 A=20/G=12974	1 2	0.6%
	PTPN2	c.31G>C, p.Glu11Gln	0	1	0.2%
	SEC23B	c.1661G>A, p.Arg554Gln[b] c.1598T>G, p.Val533Gly[b] c.167A>G, p.Tyr56Cys c.301A>G, p.Ile101Val c.689+1G>C c.2101C>T, p.Arg701Cys c.1636C>T, p.Arg546Trp c.2031G>A, p.Met677Ile c.40C>T, p.Arg14Trp c.1484G>A, p.Arg495His c.490G>T, p.Val164Leu	0 0 0 0 0 0 T=1/C=13005 A=2/G=13004 T=3/C=13003 A=95/G=12911 T=97/G=12909	1 3 1 1 1 1 1 1 1 1 8	4.0%
BRCA (n=222)	C16orf72	c.365G>A, p.Arg122Gln	A=20/G=12974	3	1.4%
	PTPN2	c.1159A>G, p.Thr387Ala	0	1	0.5%
	SEC23B	c.884C>A, p.Pro295His c.985G>T, p.Ala329Ser c.74C>A, p.Pro25His c.490G>T, p.Val164Leu	0 T=4/G=13002 A=5/C=13001 T=97/G=12909	1 1 1 2	2.3%
UCEC (n=156)	C16orf72	-	-	0	0.0%
	PTPN2	c.739G>A, p.Val247Met	T=43/C=12959	1	0.7%
	SEC23B	c.389T>C, p.Ile130Thr c.649C>T, p.Arg217* c.490G>T, p.Val164Leu	0 T=1/C=13005 T=97/G=12909	1 1 2	2.6%

[a]Allele 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).
[b]Both 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	*SEC23B* 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

[a]MAF 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 ≤ MAF ≤ 0.008. MAF ≤ 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 ≤ MAF ≤ 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
SEC23B 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

CDAII 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)

<u>Abbreviations</u>: 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.