COLD SPRING HARBOR
Molecular Case Studies

RESEARCH ARTICLE

Exome sequencing reveals germline gain-of-function *EGFR* mutation in an adult with Lhermitte–Duclos disease

Samantha Colby,[1,2,3,9] Lamis Yehia,[1,2,4,9] Farshad Niazi,[1,2] JinLian Chen,[1,2] Ying Ni,[1,2] Jessica L. Mester,[1,2] and Charis Eng[1,2,5,6,7,8]

[1]Genomic Medicine Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA; [2]Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA; [3]Case Western Reserve University School of Medicine, Cleveland, Ohio 44195, USA; [4]Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA; [5]Taussig Cancer Institute, Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA; [6]Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA; [7]CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio 44106, USA; [8]Germline High Risk Focus Group, CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio 44106, USA

Abstract Lhermitte–Duclos disease (LDD) is a rare cerebellar disorder believed to be pathognomonic for Cowden syndrome. Presently, the only known etiology is germline *PTEN* mutation. We report a 41-yr-old white female diagnosed with LDD and wild-type for *PTEN*. Exome sequencing revealed a germline heterozygous *EGFR* mutation that breaks a disulfide bond in the receptor's extracellular domain, resulting in constitutive activation. Functional studies demonstrate activation of ERK/AKT signaling pathways, mimicking *PTEN* loss-of-function downstream effects. The identification of *EGFR* as a candidate LDD susceptibility gene contributes to advancement of molecular diagnosis and targeted therapy for this rare condition with limited treatment options.

[Supplemental material is available for this article.]

Corresponding author: engc@ccf.org

Ontology terms: neoplasm of the nervous system

Published by Cold Spring Harbor Laboratory Press

doi: 10.1101/mcs.a001230

INTRODUCTION

Presently, there are more than 400 hereditary cancer syndromes, accounting for 5%–10% of all cancers (Nagy et al. 2004; Garber and Offit 2005). There are 12 million individuals in the United States and 40 million in the world living with heritable cancer, generating a considerable burden of morbidity and mortality as well as substantial societal costs. The continued identification and characterization of cancer-predisposing genes has allowed for the establishment of more accurate risk assessment, predictive testing for family members, and gene-specific tailored clinical management, all of which can aid in early cancer detection or even prevention.

Lhermitte–Duclos disease (LDD), also known as dysplastic gangliocytoma of the cerebellum, is a hamartomatous lesion characterized by a slowly enlarging mass within the cerebellar cortex that results in global thickening of the cerebellar folia (Nowak et al. 2001). The presentation of LDD can mimic many common disorders including headache, epilepsy, and stroke. LDD is believed to be a pathognomonic feature of Cowden syndrome (CS), a multisystem autosomal dominant hereditary disorder characterized by multiple hamartomas and an

[9]These authors contributed equally to this work.

elevated risk for benign and malignant neoplasms (Marsh et al. 1998; Orloff and Eng 2008). Presently, the only known etiology for LDD is germline phosphatase and tensin homolog (*PTEN*) mutation (Zhou et al. 2003).

CS is estimated to affect 1 in 200,000 individuals (Nelen et al. 1999), although this is likely an underestimate considering the extensive degree of phenotypic variability observed in this disorder. Indeed, CS is a clinical mimic with individual features resembling normal variation or alternative disorders. Thus, it is difficult to recognize and hence, remains underdiagnosed. In such situations, having a molecular diagnostic would be desirable. The first identified CS susceptibility gene, *PTEN*, encodes a ubiquitously expressed dual specificity phosphatase that regulates critical cellular processes such as cell cycle, cell migration, and apoptosis (Nelen et al. 1996). Although the discovery of *PTEN* as a CS susceptibility gene laid the foundation for gene-enabled clinical management of *PTEN* mutation-positive patients (Tan et al. 2012), this subset of patients accounts for only ~25% of CS (Tan et al. 2011). Relatedly, even with subsequent identification of germline alterations in *SDHB-D* (Ni et al. 2008), *PIK3CA/AKT1* (Orloff et al. 2013), and *KLLN* (Bennett et al. 2010) in a subset of *PTEN* wild-type CS patients, ~50% remain wild-type for all known CS predisposition genes.

Although adult-onset LDD has been believed to be pathognomonic for *PTEN* mutation, over the years, adults with LDD have been found not to have *PTEN* or pathway (*PIK3CA/AKT1*) alterations. Therefore, despite the rarity of this disease entity (estimated at 1 in 1,000,000), LDD provides a valuable patient population for discovery of other CS-relevant predisposition genes, which like *PTEN*, have far-reaching somatic implications in common sporadic neoplasias.

In this study, we utilized next-generation sequencing to analyze exomes of eight unrelated CS probands with LDD and who are wild-type for all known CS susceptibility genes. Variant prioritization revealed a pathogenic germline heterozygous variant in the receptor tyrosine kinase epidermal growth factor receptor (*EGFR* NM_005228: c.977G>T, p.Cys326Phe) in one proband that disrupts a disulfide bridge in the extracellular domain of the receptor. We show the *EGFR* p.Cys326Phe variant is an activating mutation resulting in increased EGFR autophosphorylation, indicating enhanced receptor activation. We also observed increased downstream phosphorylation of ERK1/2 and AKT, consistent with pathway activation mimicking loss-of-function of PTEN (Wu et al. 1998; Dahia et al. 1999; Weng et al. 1999, 2001a,b). Our findings suggest that activating mutations in *EGFR* contribute to the pathogenesis of *PTEN* wild-type LDD. Finding the underlying molecular etiology of identical phenotypes facilitates specific medical management and correct choice of targeted therapy or in the future, targeted prevention.

RESULTS

Clinical Presentation and Family History

We discuss a 43-yr-old female patient who presented to the emergency department at age 41 with left arm weakness and slurred speech. Investigation of the underlying cause of the presentation resulted in the incidental finding of a 5.0-cm mass within the right cerebellar hemisphere on a computed tomography (CT) brain scan without contrast. The mass was predominantly low density and demonstrated some mass effect on the fourth ventricle. The mass was further characterized using magnetic resonance imaging (MRI) with and without contrast, which revealed a striated appearance and thickening of the cerebellar folia, both characteristic of Lhermitte–Duclos disease. The MRI also revealed a focus of restricted diffusion in the right parietal convexity consistent with an area of acute infarction in the distribution of the right middle cerebral artery, which was believed to be the etiology for the

patient's chief complaint. The patient was referred to our institution for clinical and genetic workup relevant to CS, because LDD is a pathognomic feature of the disease. She was found to have macrocephaly (occipital frontal circumference of 58.5 cm) but tested negative for germline alterations in genes known to be associated with CS (*PTEN, SDHB-D, PIK3CA/AKT1, KLLN*).

Family pedigree was notable for overrepresentation of colorectal cancer from the paternal side, as well as the existence of autism spectrum disorder (both CS-related features) in the 14-yr-old daughter of the proband (Supplemental Fig. S1). Unfortunately, no genomic DNA could be obtained from other family members for mutation testing, so we proceeded by whole-exome sequencing (WES) of germline genomic DNA from our patient and seven other unrelated CS probands, also diagnosed with LDD and wild-type for all known CS susceptibility genes (Fig. 1). The median age at diagnosis of LDD was 35 (range 9–44 yr). The majority of patients presented with other clinical manifestations of CS in addition to the LDD diagnosis (Supplemental Table S2).

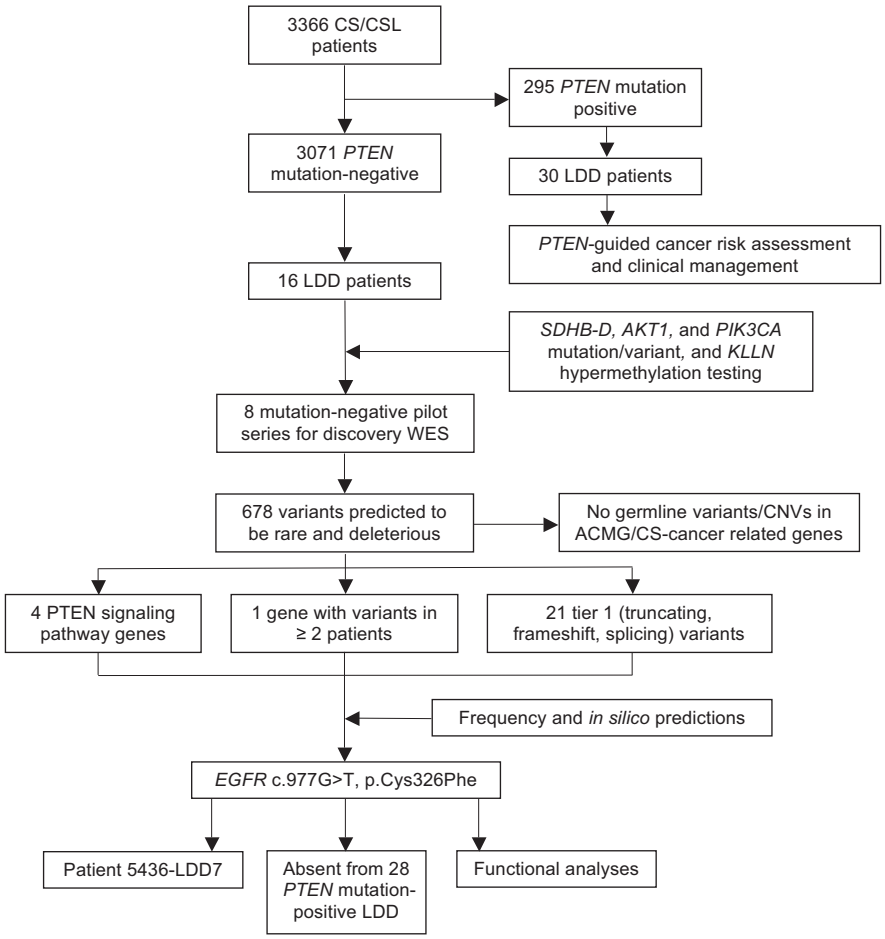

Figure 1. Experimental design for candidate gene prioritization and functional analysis. The pilot series of eight was selected from a larger pool of Lhermitte–Duclos disease (LDD) patients shown to be wild-type for *PTEN, SDHB-D, AKT1*, and *PIK3CA* variants/mutations and *KLLN* hypermethylation. CS/CSL, Cowden syndrome/CS-like; WES, whole-exome sequencing; CNVs, copy-number variants; ACMG, American College of Medical Genetics and Genomics.

Genomic Analyses

Filtration of variants using the ANNOVAR (Wang et al. 2010) variants reduction pipeline (http://www.openbioinformatics.org/annovar/) resulted in a prioritized gene variant count ranging 43–127 for each patient. As expected, we found no germline variants in the known CS susceptibility genes in the patient exomes, thus validating our initial mutation scanning results. We also analyzed our exomes for germline mutations in cancer-associated genes according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Green et al. 2013) and other genes known to be associated with hereditary cancer syndromes with clinical phenotypic overlap with Cowden syndrome. This analysis also included exon-level copy-number variant (CNV) analysis using the eXome-Hidden Markov Model (XHMM) algorithm (Fromer and Purcell 2014). Interestingly, we found no germline mutations or CNVs in these genes within the studied exomes. Accordingly, and because LDD is a highly specific feature of CS, we first looked for variants in genes that are mutated in two or more patients. We found that only two of the eight patients shared germline missense heterozygous variants in TITIN (*TTN*), the largest gene in the human genome that has not been implicated in a protumorigenic overgrowth phenotype. Because *PTEN* is the only known susceptibility gene in LDD, we then hypothesized that LDD patients may be harboring germline variants in genes belonging to the PTEN signaling pathway. Indeed, we observed four missense heterozygous variants in genes that are upstream of or downstream from PTEN (Table 1). Among the four variants, only the epidermal growth factor receptor (*EGFR*) variant (NM_005228: c.977G>T, p.Cys326Phe) has never been reported in known public databases (dbSNP138, 1000 Genomes, NHLBI-ESP6500, and ExAC). We then examined in silico the predicted functional impact and consistently noted a higher index of pathogenicity for the *EGFR* variant (Tables 2 and 3). Further structural modeling predicted that the *EGFR* p.Cys326Phe variant has a severe impact on protein structure because of breakage of a disulfide bond in the extracellular receptor domain and replacement of a 100% conserved cysteine with a phenylalanine residue (Fig. 2A,B). The *EGFR* p.Cys326Phe variant occurred in proband 5436-LDD7 (Fig. 1) and we found no other *EGFR* mutations in the seven unrelated LDD probands. Moreover, this variant was absent from germline genomic DNA of 28 LDD patients with known germline pathogenic *PTEN* mutations. Finally, exome sequencing and in silico exon-level CNV analysis of 60 unrelated *PTEN* wild-type patients with CS or Bannayan–Riley–Ruvalcaba syndrome (a hamartomatous overgrowth syndrome allelic to CS) without LDD did not reveal other *EGFR* mutations or copy-number alterations.

Functional Analyses

To investigate the in vitro impact of the *EGFR* p.Cys326Phe variant, we transiently transfected HEK293T cells with empty vector, wild-type, and p.Cys326Phe *EGFR* plasmids.

Table 1. Characteristics of four variants identified within the PTEN signaling pathway

Gene	Chromosome	HGVS DNA Reference	HGVS protein reference	Variant type	dbSNP ID	Genotype
EGFR	7:55223610	NM_005228	p.Cys326Phe	Substitution	—	Heterozygous
INSRR	1:156823922	NM_014215	p.Arg87Cys	Substitution	rs374390533	Heterozygous
ILK	11:6629361	NM_001014795	p.Arg59Trp	Substitution	—	Heterozygous
MET	7:116339877	NM_001127500	p.Ile247Val	Substitution	—	Heterozygous

PTEN, phosphatase and tensin homolog; HGVS, Human Genome Variation Society; dbSNP, Database for Short Genetic Variations.

Table 2. Prioritization of the PTEN signaling pathway candidate genes identified from eight unrelated LDD patients

	Gene			
Database	*EGFR* c.977G>T, p. Cys326Phe (Patient 5436-LDD7)	*INSRR* c.259C>T, p. Arg87Cys (Patient 4660-LDD6)	*ILK* c.175C>T, p.Arg59Trp (Patient 3249-LDD5)	*MET* c.739A>G, p. Ile247Val (Patient 344-LDD1)
1000 Genomes	0	2/5008 (0.0004) rs374390533	0	0
NHLBI-ESP6500	0	2/13006 (0.0002)	0	0
ExAC	0	64/120676 (0.0005)	2/121380 (1.648e-05)	5/120756 (4.141e-05)
MutationTaster	Disease-causing	Disease-causing	Disease-causing	Disease-causing
MutPred	Probability of deleterious mutation: 0.895	Probability of deleterious mutation: 0.798	Probability of deleterious mutation: 0.665	Probability of deleterious mutation: 0.657
Condel	Disease-causing	Neutral	Neutral	Neutral
Project HOPE	Loss of cysteine bridge → severe effect on 3D structure of protein Mutation of 100% conserved residue	Disturbance of salt bridge made by original wild-type residue	Disturbance of salt bridge made by original wild-type residue Wild-type residue not conserved at this position	Homologous proteins exist with mutant residue at this position → mutation possibly not damaging

PTEN, phosphatase and tensin homolog; LDD, Lhermitte–Duclos disease; NHLBI, National Heart, Lung, and Blood Institute; ExAC, Exome Aggregation Consortium; Condel, Consensus Deleteriousness.

Consistent with in silico predictions, immunoblotting of proteins downstream from EGFR demonstrated significantly increased phosphorylation of ERK1/2 and a slight increase of AKT phosphorylation at serine residue 473 in mutant cells relative to empty vector and *EGFR* wild-type cells (Fig. 3A). Furthermore, we also observed increased phosphorylation of EGFR at its autophosphorylation tyrosine sites Y992, Y1045, and Y1068, providing further evidence supporting enhanced mutant receptor activation (Fig. 3B). In support of our biochemical data, fluorescent visualization of green fluorescent protein (GFP)-tagged EGFR in transiently transfected HEK293T cells revealed pronounced clustering of EGFR in 293T-*EGFR*-C326F mutant cells relative to wild-type (Fig. 3C). Finally, we observed notable hypertrophy of 293T-*EGFR*-C326F mutant cells, a hallmark feature consistent with the pathobiology of human LDD phenotype (Supplemental Fig. S2).

DISCUSSION

EGFR is a transmembrane growth factor receptor that plays a well-established role in carcinogenesis (Voldborg et al. 1997). Raised awareness for *EGFR* mutations occurred with the discovery of a subset of non-small-cell lung cancers with specific activating somatic *EGFR* mutations that confer increased sensitivity to tyrosine kinase inhibitor therapy (Lynch et al. 2004; Pao et al. 2004). Somatic *EGFR* mutations and amplifications occur in multiple cancer

Table 3. Sequencing parameters for the exome-sequenced patient 5436-LDD7

Exome mean coverage	Bases covered ≥10 reads (%)	Targeted regions without reads (%)	Coverage at *EGFR* reported variant
195.9×	94.9	0.25	154× (50% mutated allele)

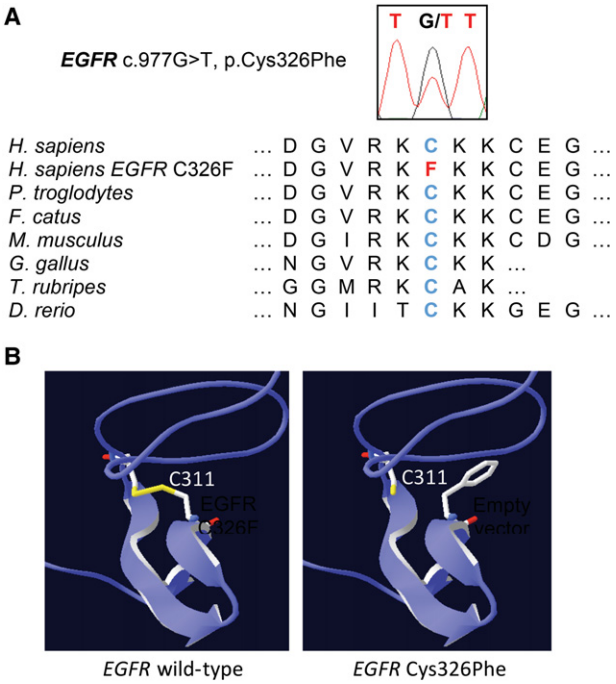

Figure 2. *EGFR* p.Cys326Phe mutation is conserved and disrupts an extracellular disulfide bridge. (*A*) A sequence chromatogram of the *EGFR* codon affected by the nucleotide mutation and conservation of the C326 amino acid across multiple species. (*B*) The protein structure of EGFR, with an enlarged view of the cysteine residue at position 326 forming the Cys311–Cys326 disulfide bond in wild-type EGFR and the phenylalanine residue and disruption of the same disulfide bond in mutant EGFR. Structural modeling was performed using Swiss-PdbViewer (http://spdbv.vital-it.ch) and EGFR PDB structure 3NJP.

types (Ciardiello and Tortora 2008), whereas germline *EGFR* mutations are rare (Centeno et al. 2011) but most commonly occur in familial lung cancer (Bell et al. 2005) and in a small subset of "never-smoking" lung cancer patients (Pao et al. 2004; Prim et al. 2014). Although the germline gain-of-function *EGFR* p.Cys326Phe mutation has not been previously reported, a somatic mutation at the same position (p.Cys326Arg) and with similar predicted gain-of-function effects is reported in The Cancer Genome Atlas (TCGA). Interestingly, this mutation was found in a renal cell carcinoma, a major CS-component cancer.

Knowing that *PTEN* is the only known LDD susceptibility gene and that LDD is highly pathognomonic of CS, our findings reveal that the *PTEN* wild-type subset of LDD patients indeed represents a genotypically heterogeneous population, especially when variable CS-overlapping clinical phenotypes are present (Supplemental Table S2). This may partly explain the absence of a shared genetic etiology among our pilot series of sequenced LDD patients. This is also confounded by the fact that LDD is rare and often reported as simplex cases—further complicated by the immense difficulty in obtaining cerebellar tissue for functional analyses. However, our data indicate that multiple genes within the PTEN signaling pathway may converge on and contribute to the same downstream signaling effects and, hence, disease etiology (Table 1). We show this biochemically for the *EGFR* p.Cys326Phe mutation, with increased activation of the ERK and AKT signaling pathways, mimicking loss-of-function *PTEN* mutations. Notably, although the proband does not have a family history of LDD, her pedigree shows an overrepresentation of colon cancer (Supplemental Fig. S1). Interestingly, *EGFR* germline mutations have never been reported in colon cancer and somatic mutations are rare (Barber et al. 2004). Nonetheless, EGFR-targeted antibodies

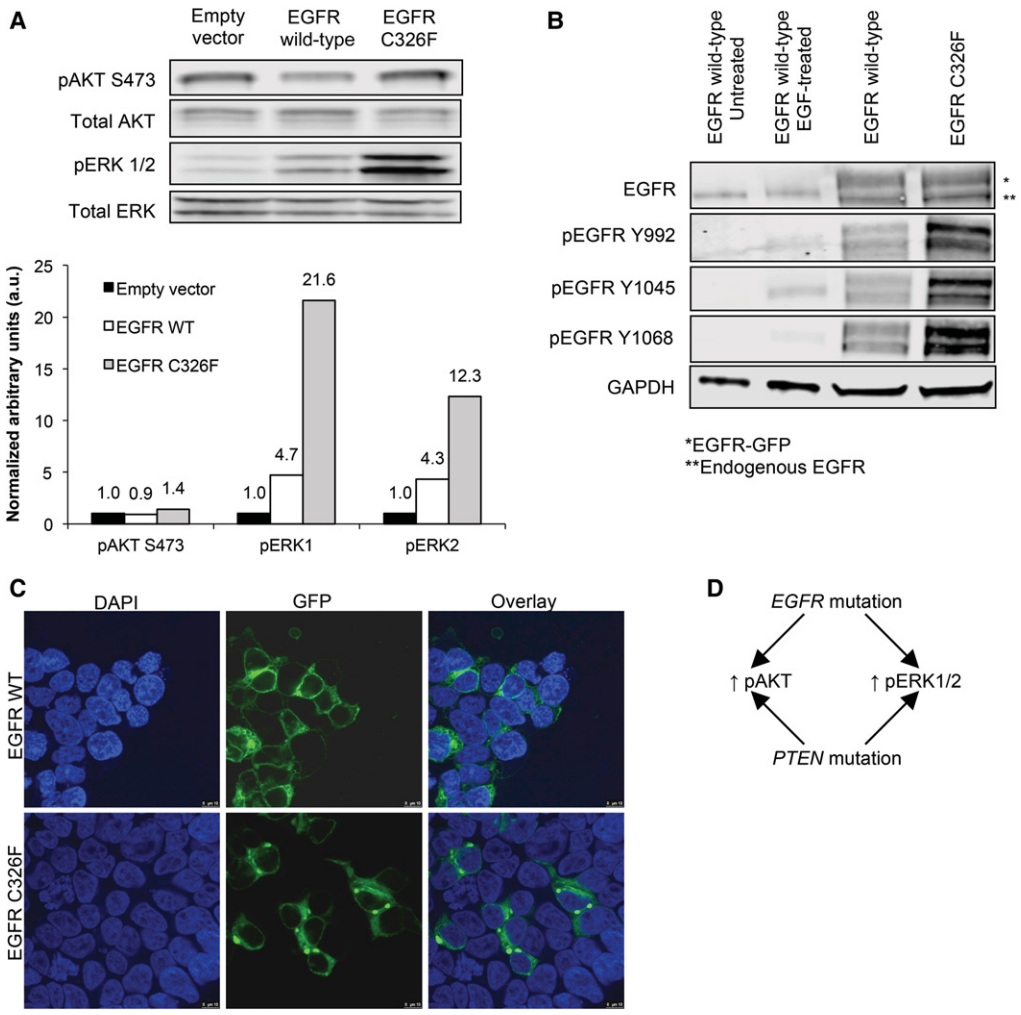

Figure 3. Functional analysis of *EGFR* p.Cys326Phe mutation. (*A*) A western blot assessing phospho-AKT, total AKT, phospho-ERK1/2, and total ERK1/2 expression in transiently transfected HEK293T cells. The quantitation bar plot appears below the western blot, with expression of phosphorylated proteins normalized to corresponding total protein. (*B*) A western blot assessing total EGFR, phospho-EGFR-Y992, phospho-EGFR-Y1045, and phospho-EGFR-Y1068 expression in transiently transfected HEK293T cells. The first two lanes represent commercially available cell lysates that serve as negative and positive controls of EGFR activation. The lysates are derived from A431 cells, known to have a large number of EGF binding sites, and representing an epidermal carcinoma of the vulva. The negative control (lane *1*) represents lysates extracted from overnight serum-starved A431 cells, hence signifying nonphosphorylated EGFR. The positive control (lane *2*) represents lysates extracted from A431 cells treated with 100 ng/mL EGF for 5 min, hence signifying active phosphorylated EGFR. Protein lysates from mutant cells (lane *4*) reveal notable EGFR phosphorylation compared with wild-type cells (lane *3*) even in the absence of EGF stimulation, reflecting constitutive receptor activation. (*C*) Confocal microscopy images of transiently transfected wild-type and p.Cys326Phe-mutant HEK293T cells, with the mutant cells demonstrating increased receptor clustering relative to wild-type cells (green). (*D*) The proposed model of the relationship between *PTEN* loss-of-function and *EGFR* gain-of-function mutations in Cowden syndrome (CS)-Lhermitte–Duclos disease (LDD). WT, wild-type.

(cetuximab) have been FDA-approved for colon cancer treatment, chiefly because of the important role EGFR signaling plays in colon carcinogenesis.

Our observations here may be pertinent in the context that LDD is surgically difficult to completely resect. Thus, nonsurgical treatment is germane. Typically, a poly ADP ribose

polymerase (PARP) inhibitor could be considered for germline *PTEN* mutation positive CS/CSL individuals and individuals with *PTEN*-related LDD. However, if an individual with LDD does not have the typical *PTEN* germline mutation but instead has a germline *EGFR* mutation, then some type of tyrosine kinase inhibitor (TKI) or antibody could be considered. TKIs have been utilized in a subset of lung cancers with somatic *EGFR* mutations with the majority of the latter occurring in the TK domain (Lynch et al. 2004; Pao et al. 2004). Most of these TK domain mutations alter the structure of EGFR such that there is an increased binding affinity of the mutant receptor to the TKI instead of ATP (Carey et al. 2006). Our LDD-*EGFR* mutation occurs in the extracellular cysteine-rich domain, where constitutive receptor dimerization and activation is predicted to occur. This should not yield the differential affinity of TKI to our mutant EGFR. Thus, in this situation, an antibody might be more appropriate. However, if indeed, as we show there is increased ERK signaling downstream from our *EGFR* mutant, then there is the theoretical possibility of using a mitogen-activated protein kinase (MAPK) inhibitor. Similarly, we also show a modest increase in phospho-AKT downstream from our constitutively activated mutant EGFR and perhaps an AKT inhibitor could also be considered.

The presence of a germline *EGFR* mutation in LDD, a cerebellar lesion, is particularly relevant because of the critical role EGFR plays in brain development (Novak et al. 2001). Interestingly, somatic-activating *EGFR* mutations and loss-of-function *PTEN* mutations have both been highly associated with glioblastoma multiforme (Smith et al. 2001). As an established oncogene, it is not surprising that a gain-of-function mutation will result in an overgrowth phenotype such as LDD. However, this is paralleled by the discovery of somatic-inactivating mutations of *LKB1* (also known as *STK11*, encoding serine/threonine kinase 11) in lung adenocarcinomas and squamous cell carcinomas (Ji et al. 2007). Intriguingly, germline mutations in *LKB1* result in Peutz–Jeghers syndrome (considered in differential diagnosis for CS) characterized by intestinal hamartomas and increased lifetime cancer risk. Importantly, unlike somatic mutations in such cancer-relevant genes, heritable germline mutations result in elevated cancer lifetime risk across multiple generations. What remains elusive is whether our patient is at increased risk of other extracerebellar tumors known to be associated with *EGFR* aberrations, including head and neck, colon (given family history), or renal cell carcinoma as being associated with a similar mutation in TCGA. This indeed underscores the importance of pursuing genetic testing in individuals with relevant family history. Detection of pathogenic cancer-predisposing germline mutations not only prompts the initiation of gene-informed high-risk surveillance and/or preventive clinical measures but also includes the theoretical possibility of an *EGFR*-targeted therapy.

METHODS

Research Participants and Clinical Data

A total of 3399 individuals were accrued prospectively. These patients were recruited from both community and academic medical centers throughout North America, Europe (with >85% originating from these two continents), and Asia using a standard protocol (Tan et al. 2012). Eligible probands were CS-LDD patients who tested negative for germline alterations in *PTEN*, *SDHB*, *SDHC*, *SDHD*, *AKT1*, *PIK3CA*, and *KLLN* and met at least the relaxed International Cowden Consortium (ICC) operational diagnostic criteria (Supplemental Table S1). Scanning of genomic DNA for known gene candidate mutations was performed as we previously reported (Ni et al. 2008), using a combination of denaturing gradient gel electrophoresis (*PTEN*), high-resolution melting curve analysis (*PTEN* and *SDHx*), and direct Sanger sequencing (all CS-known genes). Sanger sequencing of the *PTEN* promoter region was also performed, as we previously described (Teresi et al. 2007). Deletion analysis was performed

for *PTEN* and *SDHx* using the multiplex ligation-dependent probe amplification (MLPA) assay, according to the manufacturer's protocol (Schouten et al. 2002). *KLLN* promoter methylation analysis was performed using the Sequenom MassArray assay, as previously reported (Mahdi et al. 2015). For all selected individuals, we reviewed the Cleveland Clinic (CC) score (Tan et al. 2011), a semiquantitative score that is based on weighting clinical features and that estimates the pretest probability of finding a germline *PTEN* mutation (http://www.lerner.ccf.org/gmi/ccscore). Given that all individuals are wild-type for *PTEN*, we used the CC score as a surrogate of phenotypic burden. 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 Institutional Review Board approval (protocol #8458) and written informed consents 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 probands 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 (Li and Durbin 2009) (BWA v.0.6.1; http://bio-bwa.sourceforge.net). Insertion or deletion (indel) realignment, base- and quality-score recalibrations, and removal of polymerase chain reaction (PCR) duplicates from the resultant binary alignment map (BAM) files were performed with the Genome Analysis Toolkit (McKenna et al. 2010) (GATK; http://www.broadinstitute.org/gatk/), Sequence Alignment/Map (Li et al. 2009) (SAMtools), and Picard (http://broadinstitute.github.io/picard/). Variant discovery and genotype calling of single-nucleotide variants (SNVs) and short (<50-bp) indels were performed with the GATK Haplotype Caller.

Variant Filtration and Annotation

To prioritize causal variants, we applied ANNOVAR (Wang et al. 2010) (http://www.openbioinformatics.org/annovar/). First, we discarded synonymous variants and intronic variants >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 (or 0.05%) in the 1000 Genomes Project (August 2015 release) or the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) Exome Variant Server (ESP6500 data set) and variants reported in the Database for Short Genetic Variations (dbSNP)138 non-flagged database (excluding clinically associated single-nucleotide polymorphisms [SNPs]).

To predict the potential impact of missense variants on protein function, we used the program Consensus Deleteriousness (Gonzalez-Perez and Lopez-Bigas 2011) (Condel v.1.5; http://bg.upf.edu/fannsdb/), which combines multiple prediction software algorithms (SIFT [Sorting Intolerant from Tolerant], PolyPhen-2 [Polymorphism Phenotyping v2], and MutationAssessor) into a single weighted score. We inspected all resultant variants through the Integrative Genomics Viewer (Thorvaldsdottir et al. 2013) (IGV; https://www.broadinstitute.org/igv/home/). 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.

Copy-Number Variant Analysis

We implemented the XHMM algorithm to call CNVs from our exome-sequencing data, using default parameters. XHMM makes use of principal component analysis to normalize sequencing read depth and a hidden Markov model, while providing corresponding quality metrics for downstream prioritization (Fromer et al. 2012; Poultney et al. 2013). For the purposes of this study, we extracted data corresponding to cancer-associated genes according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Green et al. 2013), and other genes known to be associated with hereditary cancer syndromes with clinical phenotypic overlap with Cowden syndrome. The mean per-target depth of coverage across all samples was 60. An XHMM quality score (SQ or Q_some) of 60 was used as a cutoff threshold for CNV filtration and prioritization.

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 (Supplemental Table S3). 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). Primer sequences are listed in Supplemental Table S3. 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 (Green et al. 2013).

Variant Interpretation and In Silico Pathogenicity Predictions

Variants retained after prioritization were considered potentially pathogenic. Further prioritization was performed through selection of shared genes with variants in multiple patients and variants belonging to genes in the *PTEN* signaling pathway. In silico assessment of the mutations' effects on protein integrity was performed through a combination of mutation-prediction algorithms, namely Condel (http://bg.upf.edu/fannsdb/), MutationTaster (http://www.mutationtaster.org/), and MutPred (http://mutpred.mutdb.org/).

Predicted Structural Effects of Variants

We predicted the structural effects of variants by using the HOPE server (Venselaar et al. 2010), which integrates a series of standard algorithms to generate predictions. Structural information was obtained from the analysis of PDB structures 4KRP (EGFR), 4HI8 (ILK), 1IGR (INSRR), and 2UZX (MET). Annotations were obtained from UniProt entries EGFR_HUMAN, ILK_HUMAN, INSRR_HUMAN, and MET_HUMAN.

Cell Lines and Culture Conditions

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 Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin and streptomycin. Cell lines were maintained at 37°C and 5% CO_2 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 (Capes-Davis et al. 2010). Cell lines

purchased from ATCC after 2010 have been authenticated by short-tandem-repeat PCR analysis.

Plasmids, Mutagenesis, and Cell-Line Transfection

We generated the *EGFR* p.Cys326Phe mutant plasmid via site-directed mutagenesis of GFP-tagged wild-type *EGFR* plasmid #32751 (Carter and Sorkin 1998) (Addgene). 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. For transient overexpression, cell lines were transfected with GFP-tagged $EGFR^{WT}$, $EGFR^{C326F}$, or empty vector plasmids using Lipofectamine 3000 reagent (Thermo Fisher) according to the manufacturer's instructions. Cells were grown for at least 24 h before harvesting or downstream analysis.

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 sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto nitrocellulose membranes. We performed immunoblotting using protein lysates extracted from HEK293T cells expressing empty vector, *EGFR* wild-type, and *EGFR* p.Cys326Phe plasmids. We blotted for anti-phospho-AKT S473 (Cell Signaling #4060), anti-total AKT (Cell Signaling #9272), anti-phospho-ERK1/2 (Cell Signaling #9101), anti-ERK1/2 (Cell Signaling #9102), and anti-GAPDH (Cell Signaling #2118) antibodies, at the recommended dilutions. EGFR control cell extracts (Cell Signaling #5634) were used as negative and positive controls of EGFR phosphorylation/activation. Blots were scanned digitally and quantified with the Odyssey Infrared Imaging System (Li-Cor Biosciences).

Immunofluorescence

Cells were seeded on coverslips and were fixed with 4% paraformaldehyde for 5 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 SP8 confocal microscope (Leica).

ADDITIONAL INFORMATION

Data Deposition and Access

Exome-sequencing data has been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra/) under accession number SRX1802645. The *EGFR* c.977G>T, p.Cys326Phe mutation has been deposited to ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) under accession number SCV000299219.

Ethics Statement

Cleveland Clinic Institutional Review Board approval (protocol #8458) and written informed consent from all research participants were obtained for this study.

Acknowledgments

We are grateful to the patients who contributed to this study. We thank Todd Romigh and Kaitlin Sesock for technical advice and helpful discussions. We also thank the Genomic Medicine Biorepository of the Cleveland Clinic Genomic Medicine Institute and our database and clinical research teams. 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. Fakih predoctoral scholarship. C.E. is the Sondra J. and Stephen R. Hardis Chair of Cancer Genomic Medicine at the Cleveland Clinic and an ACS Clinical Research Professor.

Author Contributions

L.Y. and C.E. contributed to study design. J.L.M. and C.E. contributed to patient recruitment, clinical evaluation, and phenotyping. L.Y. and F.N. contributed to raw sequence data analysis. S.C. and L.Y. contributed to variant prioritization, validation, and functional analyses. J.C. contributed to EGFR protein modeling. S.C., L.Y., and C.E. contributed to manuscript preparation. All authors contributed to the revision of manuscript final version.

Competing Interest Statement

The authors have declared no competing interest.

Referees

Michele Carbone
Raymond Kim
Anonymous

Received May 28, 2016; accepted in revised form September 27, 2016.

Funding

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.). This work utilized the Leica SP8 confocal microscope that was purchased with funding from National Institutes of Health SIG grant 1S10OD019972-01.

REFERENCES

Barber TD, Vogelstein B, Kinzler KW, Velculescu VE. 2004. Somatic mutations of *EGFR* in colorectal cancers and glioblastomas. *N Engl J Med* **351:** 2883.

Bell DW, Gore I, Okimoto RA, Godin-Heymann N, Sordella R, Mulloy R, Sharma SV, Brannigan BW, Mohapatra G, Settleman J, et al. 2005. Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nat Genet* **37:** 1315–1316.

Bennett KL, Mester J, Eng C. 2010. Germline epigenetic regulation of KILLIN in Cowden and Cowden-like syndrome. *JAMA* **304:** 2724–2731.

Capes-Davis A, Theodosopoulos G, Atkin I, Drexler HG, Kohara A, MacLeod RA, Masters JR, Nakamura Y, Reid YA, Reddel RR, et al. 2010. Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int J Cancer* **127:** 1–8.

Carey KD, Garton AJ, Romero MS, Kahler J, Thomson S, Ross S, Park F, Haley JD, Gibson N, Sliwkowski MX. 2006. Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. *Cancer Res* **66:** 8163–8171.

Carter RE, Sorkin A. 1998. Endocytosis of functional epidermal growth factor receptor-green fluorescent protein chimera. *J Biol Chem* **273:** 35000–35007.

Centeno I, Blay P, Santamaria I, Astudillo A, Pitiot AS, Osorio FG, Gonzalez-Arriaga P, Iglesias F, Menendez P, Tardon A, et al. 2011. Germ-line mutations in epidermal growth factor receptor (*EGFR*) are rare but may contribute to oncogenesis: a novel germ-line mutation in *EGFR* detected in a patient with lung adenocarcinoma. *BMC Cancer* **11:** 172.

Ciardiello F, Tortora G. 2008. EGFR antagonists in cancer treatment. *N Engl J Med* **358:** 1160–1174.

Dahia PL, Aguiar RC, Alberta J, Kum JB, Caron S, Sill H, Marsh DJ, Ritz J, Freedman A, Stiles C, et al. 1999. PTEN is inversely correlated with the cell survival factor Akt/PKB and is inactivated via multiple mechanisms in haematological malignancies. *Hum Mol Genet* **8:** 185–193.

Fromer M, Purcell SM. 2014. Using XHMM software to detect copy number variation in whole-exome sequencing data. *Curr Protoc Hum Genet* **81:** 7 23 21–27 23 21.

Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, Handsaker RE, McCarroll SA, O'Donovan MC, Owen MJ, et al. 2012. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet* **91:** 597–607.

Garber JE, Offit K. 2005. Hereditary cancer predisposition syndromes. *J Clin Oncol* **23:** 276–292.

Gonzalez-Perez A, Lopez-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.

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, et al. 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* **15:** 565–574.

Ji H, Ramsey MR, Hayes DN, Fan C, McNamara K, Kozlowski P, Torrice C, Wu MC, Shimamura T, Perera SA, et al. 2007. LKB1 modulates lung cancer differentiation and metastasis. *Nature* **448:** 807–810.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, et al. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350:** 2129–2139.

Mahdi H, Mester JL, Nizialek EA, Ngeow J, Michener C, Eng C. 2015. Germline *PTEN*, *SDHB-D*, and *KLLN* alterations in endometrial cancer patients with Cowden and Cowden-like syndromes: an international, multicenter, prospective study. *Cancer* **121:** 688–696.

Marsh DJ, Coulon V, Lunetta KL, Rocca-Serra P, Dahia PL, Zheng Z, Liaw D, Caron S, Duboue B, Lin AY, 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.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303.

Nagy R, Sweet K, Eng C. 2004. Highly penetrant hereditary cancer syndromes. *Oncogene* **23:** 6445–6470.

Nelen MR, Padberg GW, Peeters EA, Lin AY, van den Helm B, Frants RR, Coulon V, Goldstein AM, van Reen MM, Easton DF, et al. 1996. Localization of the gene for Cowden disease to Chromosome 10q22–23. *Nat Genet* **13:** 114–116.

Nelen MR, Kremer H, Konings IB, Schoute F, van Essen AJ, Koch R, Woods CG, Fryns JP, Hamel B, Hoefsloot LH, et al. 1999. Novel *PTEN* mutations in patients with Cowden disease: absence of clear genotype–phenotype correlations. *Eur J Hum Genet* **7:** 267–273.

Ni Y, Zbuk KM, Sadler T, Patocs A, Lobo G, Edelman E, Platzer P, Orloff MS, Waite KA, 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.

Novak U, Walker F, Kaye A. 2001. Expression of EGFR-family proteins in the brain: role in development, health and disease. *J Clin Neurosci* **8:** 106–111.

Nowak DA, Trost HA, Porr A, Stolzle A, Lumenta CB. 2001. Lhermitte–Duclos disease (dysplastic gangliocytoma of the cerebellum). *Clin Neurol Neurosurg* **103:** 105–110.

Orloff MS, Eng C. 2008. Genetic and phenotypic heterogeneity in the *PTEN* hamartoma tumour syndrome. *Oncogene* **27:** 5387–5397.

Orloff MS, He X, Peterson C, Chen F, Chen JL, Mester JL, Eng C. 2013. Germline *PIK3CA* and *AKT1* mutations in Cowden and Cowden-like syndromes. *Am J Hum Genet* **92:** 76–80.

Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, et al. 2004. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci* **101:** 13306–13311.

Poultney CS, Goldberg AP, Drapeau E, Kou Y, Harony-Nicolas H, Kajiwara Y, De Rubeis S, Durand S, Stevens C, Rehnstrom K, et al. 2013. Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *Am J Hum Genet* **93:** 607–619.

Prim N, Legrain M, Guerin E, Mennecier B, Weingertner N, Voegeli AC, Guenot D, Maugard CM, Quoix AE, Beau-Faller M. 2014. Germ-line exon 21 *EGFR* mutations, V843I and P848L, in nonsmall cell lung cancer patients. *Eur Respir Rev* **23:** 390–392.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* **30:** e57.

Smith JS, Tachibana I, Passe SM, Huntley BK, Borell TJ, Iturria N, O'Fallon JR, Schaefer PL, Scheithauer BW, James CD, et al. 2001. PTEN mutation, EGFR amplification, and outcome in patients with anaplastic astrocytoma and glioblastoma multiforme. *J Natl Cancer Inst* **93:** 1246–1256.

Tan MH, Mester J, Peterson C, Yang Y, Chen JL, Rybicki LA, Milas K, Pederson H, Remzi B, Orloff MS, et al. 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.

Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. 2012. Lifetime cancer risks in individuals with germline *PTEN* mutations. *Clin Cancer Res* **18:** 400–407.

Teresi RE, Zbuk KM, Pezzolesi MG, Waite KA, Eng C. 2007. Cowden syndrome-affected patients with *PTEN* promoter mutations demonstrate abnormal protein translation. *Am J Hum Genet* **81:** 756–767.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178–192.

Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* **11:** 548.

Voldborg BR, Damstrup L, Spang-Thomsen M, Poulsen HS. 1997. Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials. *Ann Oncol* **8:** 1197–1206.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38:** e164.

Weng LP, Smith WM, Dahia PL, Ziebold U, Gil E, Lees JA, Eng C. 1999. PTEN suppresses breast cancer cell growth by phosphatase activity-dependent G_1 arrest followed by cell death. *Cancer Res* **59:** 5808–5814.

Weng LP, Brown JL, Eng C. 2001a. PTEN coordinates G(1) arrest by down-regulating cyclin D1 via its protein phosphatase activity and up-regulating p27 via its lipid phosphatase activity in a breast cancer model. *Hum Mol Genet* **10:** 599–604.

Weng LP, Smith WM, Brown JL, Eng C. 2001b. PTEN inhibits insulin-stimulated MEK/MAPK activation and cell growth by blocking IRS-1 phosphorylation and IRS-1/Grb-2/Sos complex formation in a breast cancer model. *Hum Mol Genet* **10:** 605–616.

Wu X, Senechal K, Neshat MS, Whang YE, Sawyers CL. 1998. The PTEN/MMAC1 tumor suppressor phosphatase functions as a negative regulator of the phosphoinositide 3-kinase/Akt pathway. *Proc Natl Acad Sci* **95:** 15587–15591.

Zhou XP, Marsh DJ, Morrison CD, Chaudhury AR, Maxwell M, Reifenberger G, Eng C. 2003. Germline inactivation of *PTEN* and dysregulation of the phosphoinositol-3-kinase/Akt pathway cause human Lhermitte–Duclos disease in adults. *Am J Hum Genet* **73:** 1191–1198.