Contemporary Clinical Trials Communications 21 (2021) 100733

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

A randomized double-blind controlled trial of everolimus in individuals with *PTEN* mutations: Study design and statistical considerations

Antonio Y. Hardan [a], Booil Jo [a], Thomas W. Frazier [b,c,d], Patricia Klaas [e], Robyn M. Busch [d,e,f], Kira A. Dies [g,h], Rajna Filip-Dhima [g,h], Anne V. Snow [i], Charis Eng [d,f,j,k,l], Rabi Hanna [f,i], Bo Zhang [g,*], Mustafa Sahin [g,h,**]

[a] Department of Psychiatry and Behavioral Science, Stanford University, Stanford, CA, USA
[b] Autism Speaks, New York, NY, USA
[c] Department of Psychology, John Carroll University, Cleveland Heights, Ohio, USA
[d] Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA
[e] Department of Neurology, Neurological Institute, Cleveland Clinic, Cleveland, OH, USA
[f] Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA
[g] Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[h] F.M. Kirby Neurobiology Center, Translational Neuroscience Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[i] Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[j] Department of Pediatric Hematology Oncology and Bone Marrow Transplantation, Cleveland Clinic Children's Hospital, Cleveland, OH, USA
[k] Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA
[l] Germline High Risk Cancer Focus Group, Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA

ARTICLE INFO

Keywords:
Everolimus
Neurobehavior and neurocognition
PTEN hamartoma tumor syndrome
Autism spectrum disorder
Randomized controlled trial

ABSTRACT

This randomized, double-blind controlled trial of everolimus in individuals with germline phosphatase and tensin homolog mutations (*PTEN*) was designed to evaluate the safety of everolimus compared with placebo and to evaluate the efficacy of everolimus on neurocognition and behavior compared to placebo as measured by standardized neurocognitive and motor measures as well as behavioral questionnaires. The safety profile of everolimus is characterized by manageable adverse events that are generally reversible and non-cumulative. The primary safety endpoint of this study was drop-out rate due to side effects, comparing everolimus versus placebo. We also sought to determine the frequency of adverse events by type and severity. The main efficacy endpoint was a neurocognitive composite computed in two ways: 1) an average for working memory, processing speed, and fine motor subtests; and 2) the same average as above except weighted 2/3, and an additional average based on all other available neurocognitive testing measures assessing the additional domains of nonverbal ability, visuomotor skills, verbal learning, and receptive and expressive language, weighted 1/3. Secondary efficacy endpoints examined the effect of everolimus on overall global clinical improvement, autism symptoms, behavioral problems, and adaptive abilities as measured by validated, standardized instruments. We predicted that the rate of adverse events would be no more than 10% higher in the everolimus group compared to placebo, and overall severity of side effects would be minimal. We also expected that individuals receiving everolimus would show more improvement, relative to those taking placebo, on the composite neurocognitive index.

1. Introduction

PTEN germline mutations are associated with a spectrum of clinical disorders characterized by neurocognitive deficits, intellectual disability, autism symptomatology, skin lesions, macrocephaly, hamartomatous overgrowth of tissues, and an increased risk of specific cancers [1–4]. In humans, *PTEN*-related research has historically focused on physical manifestations of the disease and the pathophysiology and treatment of the hamartomatous lesions that arise in affected patients and their predisposition for malignancy. In contrast, there is

* Corresponding author. Boston Children's Hospital, 300 Longwood Avenue, Boston, MA, 02115, USA.
** Corresponding author. Boston Children's Hospital, 300 Longwood Avenue, Boston, MA, 02115, USA.
E-mail addresses: Bo.Zhang@childrens.harvard.edu (B. Zhang), Mustafa.Sahin@childrens.harvard.edu (M. Sahin).

https://doi.org/10.1016/j.conctc.2021.100733
Received 13 August 2020; Received in revised form 3 January 2021; Accepted 16 January 2021
Available online 6 February 2021

much less research to date that focuses on the behavioral and cognitive features and their treatment [2,5,6]. Multiple murine CNS-conditional (CNS: central nervous system) knock-out models have established the role of *PTEN* in learning and control of social behavior [7,8]. Using a germline model, female *PTEN* heterozygous animals exhibited decreased social behavior [9]. More recently, a germline model that results in inappropriate *PTEN* subcellular localization showed inappropriate social behavior and clumsiness, reminiscent of a subset of *PTEN* mutation-positive autism spectrum disorder (ASD) patients [10–13]. Additionally, *PTEN* loss in mature neurons have led to diminished social behavior, an effect replicated in a model of *PTEN* loss in neuronal precursors [14,15]. The Nse-cre x $PTEN^{loxP/loxP}$ model also shows decreased social interaction and increased anxiety [16]. Interestingly, inhibition of mTOR complex 1, a downstream target of AKT signaling, improved social behavior in this model [15].

Everolimus, a novel derivative of rapamycin, has been in clinical development since 1996 as an immunosuppressant in solid organ transplantation and as a cancer growth blocker. In TSC, a genetic disorder with synaptic disruption and cognitive and behavioral features similar to those of PHTS, everolimus is used to block the growth of subependymal giant cell astrocytoma, renal angiomyolipoma, and hamartomatous lesions. It can also be used as adjunctive therapy for TSC-associated partial onset seizures. At the cellular and molecular level, everolimus acts as a signal transduction inhibitor that binds to FKB12 to selectively inhibit mTOR, a key and a highly conservative serine-threonine kinase in the PI3K/AKT signaling cascade, which is known to be dysregulated in a wide spectrum of human cancers [17]. mTOR is present in all cells and is a central regulator of protein synthesis and ultimately cell growth, cell proliferation, angiogenesis, and cell survival. It is currently the only known target of everolimus [17].

In 2003, everolimus was approved in Europe (trade name: Certican®) via the Mutual Recognition Procedure (MRP) for the prevention of organ rejection in patients with renal and cardiac transplantation. Everolimus 2.5 mg, 5 mg, 7.5 mg, and 10 mg tablets were approved under the trade name Afinitor® for patients with advanced renal cell carcinoma (RCC) after failure of treatment with Sutent® (sunitinib) or Nexavar® (sorafenib) in the United States (US), European Union (EU), and several other countries and is undergoing registration in other regions worldwide. In 2010, Afinitor® received US approval for patients with subependymal giant cell astrocytoma (SEGA) associated with tuberous sclerosis complex (TSC). Everolimus is also available as Votubia® in the EU for patients with SEGA associated with TSC who require therapeutic intervention but are not candidates for curative surgical resection. Afinitor® was approved for "progressive pancreatic neuroendocrine tumor (PNET) in patients with unresectable, locally advanced, or metastatic disease" in 2011 in various countries, including the US and Europe. In 2012 Afinitor® received approval for the treatment of postmenopausal women with advanced hormone receptor-positive, HER2-negative breast cancer in combination with exemestane, after failure of treatment with letrozole or anastrozole. Furthermore, in 2012, Afinitor® received approval for the treatment of adult patients with TSC who have renal angiomyolipoma not requiring immediate surgery. Afinitor® Diseperz was approved in 2018 for adjunctive treatment of adult and pediatric patients aged 2 year and older with TSC-associated partial-onset seizures. Everolimus is also approved to treat hamartomatous lesions in tuberous sclerosis complex (TSC), a genetic disorder with synaptic disruption and cognitive and behavioral features similar to those of *PTEN*.

We designed this Phase II 6-month, randomized, double-blind placebo-controlled trial to establish the short-term safety profile of everolimus in individuals aged 5–45 years (inclusive) with germline *PTEN* mutations. We planned to evaluate the associated cognitive and behavioral changes in this study, with the goal of generating plausible hypotheses to be tested in a future Phase III confirmatory trial. This designed trial was registered in clinicaltrials.gov as "RAD001 and Neurocognition in PTEN Hamartoma Tumor Syndrome" (NCT0299180). The purpose of the current article is to elaborate on the study design and statistical analysis plan of this trial in the context of clinical trials for rare disorders.

2. Study design and procedure

2.1. Study design and rationale

We designed our Phase II, double-blind, randomized, parallel group, placebo-controlled, three-center study to evaluate treatment with everolimus versus placebo in 4 different phases: pre-treatment (screening), a 6-month blinded treatment phase, a 6-month open-label phase for patients initially randomized to placebo, and a follow-up phase. Each of these phases is described in detail below. We planned to screen approximately 60 patients with *PTEN* gene mutations aged 5–45 years (inclusive) to identify 40 meeting inclusion and exclusion criteria. By maintaining a lower age limit of 5 years, we ensured a more accurate assessment of cognitive and behavioral outcomes, streamlined the neurobehavioral assessment battery, and achieved minimal loss to outcome measure scope and sensitivity, similar to the strategy employed in a previous everolimus trial in TSC (NCT01289912), performed by PIs Mustafa Sahin, M.D., Ph.D. and Darcy Krueger, M.D., Ph.D [18]. This study was conducted at three sites: Stanford University, Cleveland Clinic, and Boston Children's Hospital. The Institutional Review Boards at these three sites approved this study. Patients or parents/guardians signed an informed consent prior to participating in any study activities.

2.2. Pre-treatment screening phase

After patients/parents provided their signed informed consent form and eligibility was confirmed, the investigator or his/her designee registered the patient for randomization. The randomization ratio was 1:1, with one patient being randomly assigned to everolimus for every patient randomly assigned to placebo. Treatment was assigned via the data management center but was not disclosed prior to the baseline visit, to allow adequate time for pharmacy preparation.

Screening evaluations included demographics, relevant medical history, current medical conditions, a physical examination (including a neurological examination), suicidal ideation and behavior assessment, vital signs, laboratory assessments, and other additional study entry evaluations. To ensure that participants met inclusion criteria and were able to complete at least one of the primary neurocognitive endpoints, the Stanford-Binet Intelligence Scales, Fifth Edition (SB-5) [19], Conners' Continuous Performance Test, Third Edition (CPT-3) [20], and the Purdue Pegboard Test [21] were administered at the screening visit.

Participants were also screened for signs and/or symptoms of cancer and for suicidal ideation and/or behaviors using the Columbia Suicide Severity Rating Scale (C-SSRS) [22]. If a patient presented at the screening visit with a mass that could be malignant or benign, the study physician referred them to the appropriate health care providers. If the mass was benign and the patient met all other inclusion/exclusion criteria, the patient was enrolled in the study. Participants who were determined to have significant suicidal ideation and/or behaviors were ineligible to participate in the trial and were referred to the appropriate healthcare provider per the principal investigators' discretion.

Patients had screening evaluations performed within 6 weeks of the baseline visit to ensure they met all inclusion and exclusion criteria, listed in Table 1, at the time of the baseline visit. Results of all screening evaluations were reviewed by the site's Principal Investigator or his/her designee prior to enrollment of the patient.

If the participant, legal guardian, or the study physician made the decision to have the participant exit the study early, a follow-up visit was scheduled 28 days (±14 days) after termination. At this visit, the study physician reviewed the patient's medical history, concomitant medications, and the Dosage Record Treatment Emergent Symptom Scale (DOTES), a general rating scale published by the Early Clinical

Table 1
Inclusion and exclusion criteria for double-blind treatment phase and open-label phase.

Double-blind treatment phase inclusion criteria

1. Male and female outpatients between 5 and 45 years of age (inclusive).
2. Pathogenic *PTEN* mutation confirmed by clinical genetic testing.
3. Participant must be able to complete one of the following three standardized assessments: CPT-3 mean reaction time, SB-5 working memory, or the Purdue Pegboard Test.
4. Performance below the age-adjusted population mean on at least one of the above standardized measure: attention (CPT-3 mean reaction time), working memory (SB-5), or fine motor skills (Purdue Pegboard Test; either dominant hand, non-dominant hand, or both hands).
5. Adequate bone marrow function as shown by: a. platelets $\geq$80,000/mm^3,
   b. absolute neutrophil count $\geq$1000/mm^3,
   c. hemoglobin $\geq$9 g/dL.
6. Adequate liver function as shown by: a. Total serum bilirubin $<$1.5 x ULN,
   b. AST and ALT levels $<$2.5 x ULN,
   c. INR $\leq$2.
7. Adequate renal function: serum creatinine $<$1.5 x ULN.
8. Signed informed consent obtained prior to any screening procedures.
9. Individuals on psychotropic and anti-epileptic medications should maintain a stable dose for at least 2 months prior to the screening visit.
10. Negative serum pregnancy test for females at screening and no plans to become pregnant or conceive a child while participating in the study. The effects of mTOR inhibitors on the developing fetus at the doses used in this study are unknown. For this reason, women of child-bearing potential and men must agree to use adequate contraception prior to study entry and for the duration of the study. Estrogen-containing oral contraceptives are not recommended in women enrolled in this study. Abstinence or two effective non-estrogen or barrier methods of contraception (such as condoms + spermicidal foam) must be used.
11. No anticipated changes in the frequency and intensity of existing interventions such as behavioral and developmental treatments, in home services, and speech therapy.
12. No planned changes in school placement.
13. For individuals under 18 or who are otherwise incapable, there must be an available caregiver who can reliably bring subject to clinic visits and provide trustworthy data.
14. Able to communicate fluently in English.

Double-blind treatment phase exclusion criteria

1. Patients currently receiving anticancer therapies or who have received anticancer therapies within 4 weeks of the start of everolimus (including chemotherapy, radiation therapy, antibody-based therapy, etc.).
2. Known intolerance or hypersensitivity to everolimus or other rapamycin analogs (e.g. sirolimus, temsirolimus).
3. Known impairment of gastrointestinal function or gastrointestinal disease that may significantly alter the absorption of oral everolimus.
4. Uncontrolled diabetes mellitus as defined by HbA1c $>$8% despite adequate therapy. Patients with a known history of impaired fasting glucose or diabetes mellitus (DM) may be included, however blood glucose and antidiabetic treatment must be monitored closely throughout the trial and adjusted as necessary.
5. Patient with uncontrolled hyperlipidemia: fasting serum cholesterol $>$300 mg/dL OR $>$ 7.75 mmol/L AND fasting triglycerides $>$2.5 x ULN.
6. Patients who have any severe and/or uncontrolled medical or psychiatric conditions.
7. Chronic treatment with corticosteroids or other immunosuppressive agents. Topical or inhaled corticosteroids are allowed.
8. Known history of or seropositivity for hepatitis B, hepatitis C, or HIV.
9. Patients who have received live attenuated vaccines within 1 week of start of everolimus and during the study. Patients should also avoid close contact with others who have received live attenuated vaccines. Examples of live attenuated vaccines include intranasal influenza, measles, mumps, rubella, oral polio, BCG, yellow fever, varicella, and TY21a typhoid vaccines.
10. Patients who have a history of another primary malignancy, with the exceptions of:
    a. non-melanoma skin cancer, and
    b. carcinoma in situ of the cervix, uteri, or breast from which the patient has been disease free for $\geq$3 years.
11. Planned changes to concomitant medications.
12. Prior or concomitant therapy with known or possible anti-mTOR activity, including rapamycin (sirolimus).
13. Concomitant therapy with strong inhibitor (e.g., cyclosporine and ketoconazole) or inducer of CYP3A.
14. Active infection at time of enrollment.
15. Patients with a history of non-compliance to medical regimens or who are considered potentially unreliable or will not be able to complete the entire study.
16. Patients who are currently part of or have participated in any clinical investigation with an investigational drug within 1 month prior to dosing.
17. Pregnant or nursing (lactating) women.
18. Women of child-bearing potential, defined as all women physiologically capable of becoming pregnant, must use highly effective methods of contraception during the study and 8 weeks after. Highly effective contraception methods include either a combination of any two of the following:
    a. use of oral, injected, or implanted hormonal non-estrogen containing methods of contraception,
    b. placement of an intrauterine device system,
    c. barrier methods of contraception such as condom or occlusive cap (diaphragm or cervical/vault caps) with spermicidal foam/gel/film/cream/vaginal suppository,
    d. total abstinence, or
    e. male/female sterilization. Women are considered post-menopausal and not of child-bearing potential if they have had 12 months of natural (spontaneous) amenorrhea with an appropriate clinical profile (e.g. age appropriate, history of vasomotor symptoms) or have had surgical bilateral oophorectomy (with or without hysterectomy) or tubal ligation at least six weeks prior to randomization. In the case of oophorectomy alone, only when the reproductive status of the woman has been confirmed by follow up hormone level assessment, is she considered not of child-bearing potential.
19. Male patients whose sexual partner(s) are women of child-bearing potential who are not willing to use adequate contraception during the study and for 8 weeks after the end of treatment.
20. Major surgery, radiation therapy, or stereotactic radiosurgery within previous 4 weeks at time of screening.
21. Neurosurgery within prior 6 months at time of screening.

Open-label phase inclusion criteria

1. Patients who completed the double-blind phase of the study and were assigned to the placebo treatment arm.
2. Verbal consent (and assent, as appropriate) obtained prior to any open-label phase study procedures,

CPT-3: Conners' Continuous Performance Test, Third Edition; SB-5: Stanford-Binet Intelligence Scales, Fifth Edition.

Drug Evaluation Unit of the National Institute of Mental Health [23]. If the patient exited the trial due to an adverse event (AE), this event was reviewed and, if still present, the study team followed up with the patient/legal guardian(s) for 56 days or until the event subsided. At the completion of this follow-up visit, the participant could be unblinded and the results could be disclosed to the participant.

2.3. Blinded treatment phase

All baseline procedures and evaluations were completed within 6 weeks of the screening visit (study timeline is shown in Table 2). If a patient's safety lab results were out of range, and the investigator had reason to believe there were situational factors affecting the values, the screening labs could be repeated. If a patient's baseline visit was more than 6 weeks after the screening visit, the safety labs, physical and neurological exam, and medical history were redone and evaluated to confirm that there were no changes. If the screening safety evaluation and/or lab results were out-of-range, the participant was not eligible to participate in the study.

Patients started randomized, blinded treatment within 7 days of the baseline visit if they still met all inclusionary criteria and had not developed any exclusionary criteria. Patients/families were instructed to take the medication at the same each morning with a light, low-fat breakfast for 6 months. Treatment was only stopped if an intolerable toxicity occurred, consent was withdrawn, or the investigator decided to discontinue the patient from study treatment.

Safety evaluations were routinely performed (visit 3/month 1, visit 5/month 3, and visit 8/month 6). Patients were in a fasting state at the time of blood sampling for all laboratory evaluations, including the lipid profile. Hematology and biochemistry assessments were required at the screening visit, the month 1 visit, the month 3 visit, and the month 6 visit. All blood samples obtained at each visit were sent to LabConnect, LLC and associated testing facilities for analysis. If safety labs were out of

Table 2
Double-blind treatment phase study timeline.

Measurement	Screen (Visit 1)	Baseline[b] (Visit 2)	Month 1 (Visit 3)	Month 2[e] (Visit 4)	Month 3 (Visit 5)	Month 4[e] (Visit 6)	Month 5[e] (Visit 7)	Month 6 (Visit 8)	Follow-up	PRN
Visit Windows		Within 42d of Screen	±14d	±14d	±14d	±14d	±14d	±14d	28d after last visit ±14d	
Medical History	X	X	X	X	X	X	X	X	X	
Clinical Interview	X	X	X	X	X	X	X	X		
Inclusion/Exclusion Criteria	X	X								
Cgi Scales: Severity & Improvement		X[c]	X	X	X	X	X	X		
Vital Signs	X	X	X		X			X		
Physical/Neurological Exam	X							X		X
Dermoscopy		X			X			X		
Microbiome/Mycobiome Sample Collection		X			X			X		
Tanner Staging		X						X		
Side Effects (CTCAE V5.0 + DOTES)	X	X	X	X	X	X	X	X	X	
Laboratory Tests[a]	X		X		X			X		X
Everolimus Level (Only Done After M1 If Needed)			X							X
PTEN-Associated Proteins	X[d]	X[d]			X			X		
Concomitant Treatment Log	X	X	X	X	X	X	X	X	X	
Autism Diagnostic Interview-Revised		X								
Primary Efficacy Outcome Neurocognitive Composite		X			X			X	X	
Columbia-Suicide Severity Rating Scale	X		X		X			X		
Cpt-3/K-Cpt-2)	X	X			X			X		
SB-5/Mullen Scales Of Early Learning (MSEL)	X							X		
SB-5 Working Memory Subscale Only		X			X					
Purdue Pegboard (Pp)	X	X			X			X		
Wechsler Processing Speed Index		X			X			X		
Wide Range Assessment of Memory and Learning-2 (WRAML-2)		X			X			X		
Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4)		X			X			X		
Expressive Vocabulary Test – Second Edition (EVT-2)		X			X			X		
Autism Diagnostic Observation Schedule (ADOS-2)		X						X		
Social Responsiveness Scale – Second Edition (SRS-2)		X			X			X		
Repetitive Behavior Scale – Revised (RBS-R)		X			X			X		
Developmental Coordination Disorder Questionnaire (DCDQ)		X			X			X	X	
Behavior Rating Inventory of Executive Function (BRIEF)		X			X			X		
Adult/Child Behavior Checklist		X			X			X	X	
Sensory Profile Questionnaire - Short Form (SPQ)		X			x			x		
Vineland Adaptive Behavior Scales (VABS-III) – caregiver report		X			X			X		
Eye Tracking (Optional)		X			X			X		
Resting State EEG/Auditory Evoked Potentials (Optional)		X			X			X		
Participant Unblinding								X		X

**: primary outcome mesaure only.

[a] Coagulation testing will only be performed at screening, month 3, and month 6. Pregnancy testing, serum is done at screening and month 6 and urine may be used for onsite visits (if deemed clinically necessary) for women of childbearing potential.

[b] Monthly visit windows are calculated based on date of baseline visit.

[c] At the baseline visit, only the CGI: Severity scale will be completed.

[d] *PTEN*-associated protein blood collections will be collected at either the screening or baseline visit (as well as at the 3-month and 6-month visits). CGI: Clinical Global Impressions Scale; CPT: Conners' Continuous Performance Test; IE: Independent Evaluator; CTCAE: Common Terminology Criteria for Adverse Events; DOTES: Dosage Record and Treatment Emergent Symptom Scale; SB-5: Stanford-Binet Intelligence Scales, Fifth Edition. PRN: *Pro* Re *Nata* (visits as needed).

[e] May be performed by phone or in person.

range while the participant was randomized, a clinical assessment by the study physician took place that could include a repeat of blood work. Withdrawal from the study was considered by the principal investigator if clinically indicated and after discussion with the study team.

A pharmacokinetic (PK) sample was collected at the month 1 visit (1 month ±14 days from the baseline visit), when possible at steady state, prior to dose administration during this visit, within 24 ± 4 h after the last dose. Participants who received a dose adjustment after this visit had another PK sample taken at the study site or locally, by a designated and trained remote health care provider, 2 weeks (±1 week) after the new dose had been taken. A PK sample could also be collected from a patient experiencing an adverse event (AE) or side effect. The decision to obtain this sample was made by the site physician.

Visits 4, 6, and 7 (months 2, 4, and 5) could be in-person or by phone. If performed by phone, a phone interview was performed to collect safety and other data, review AEs, and assess changes in mood or behavior. All applicable case report forms were completed.

Tests conducted during the blinded treatment phase included laboratory tests for safety, physical exams (including a neurological assessment), vital signs, and neuropsychological assessments at the baseline, 3-month and 6-month visits. If unforeseen circumstances (i.e., unexpected personal reasons) prevented the patient from complying with the established visit schedule, the site could re-schedule the visit within 14 days of the expected visit date. The reason(s) for any visit or treatment delays were documented in the case report forms for the appropriate visit. All on-site visits (screening, baseline, month 1, month 3, and month 6) were required to be completed by the participant, and any missed visits would warrant termination from the study. Participants and families were informed that a summary of the neuropsychological results could be made available at study completion.

At the end of the double-blind phase (or early termination), the treatment code was broken, and the participant was informed of their assigned treatment arm. Participants who received the active compound entered the follow-up phase of the study and were referred back to their treating physicians. Individuals who were in the placebo group were invited to enter a 6-month open-label extension trial.

2.4. Open-label phase, follow-up phase, everolimus administration and treatment course

Details on open-label phase, follow-up phase, everolimus administration and treatment course are described in the appendix.

3. Statistical considerations

3.1. Study objectives and measures

3.1.1. Primary objectives

Our primary objectives were to evaluate the safety and efficacy of everolimus compared with placebo in patients with *PTEN*. Primary safety outcomes focused on Grades 3 and 4 AEs, SAEs, and Grades 3 and 4 laboratory toxicities, as defined by the Common Terminology Criteria for Adverse Events (CTCAE), version 4. We also aimed to determine the frequency of AEs by type and severity. Primary efficacy outcomes were evaluated using a composite neurocognitive score generated based on performance on objective measures.

3.1.2. Primary safety outcomes

The primary safety endpoint was drop-out rate due to AEs and side effects, comparing everolimus versus placebo. Based on the published literature and our experience in prior TSC trials, we hypothesized that the drop-out rate due to AEs and side effects in those receiving everolimus would be similar to those in the placebo group with minimal effect size (<10% difference). We predicted that the rate of AEs would be no more than 10% higher in the everolimus group compared to placebo, but overall severity would be minimal.

An AE was defined as the appearance of (or worsening of any pre-existing) undesirable sign(s), symptom(s), or medical condition(s) after the patient's signed informed consent had been obtained. Abnormal laboratory values or test results occurring after informed consent constituted AEs only if they induced clinical signs or symptoms, were considered clinically significant, required therapy (e.g., hematologic abnormality that requires transfusion or hematological stem cell support), or required changes in study medication(s).

Conditions or symptoms that were present at the time of informed consent were recorded on the Medical History case report form and were entered into the database. AEs that occurred after informed consent were recorded on the running AE log, as well as on the AE case report form, and were entered into the database. AE monitoring was continued for at least 28 days following the last dose of study treatment. AEs (including lab abnormalities that constituted AEs) were described using a diagnosis whenever possible, rather than individual underlying signs and symptoms. When a clear diagnosis could not be identified, each sign or symptom was reported as a separate AE.

AEs were assessed according to the CTCAE, version 4. If CTCAE grading did not exist for an AE, the severity of mild, moderate, severe, and life-threatening, corresponding to Grades 1–4, were used. The occurrence of AEs was sought by non-directive questioning of the patient at each visit during the study. AEs were also detected when they were volunteered by the patient during or between visits or through physical examination, laboratory test, or other assessments. As far as possible, each AE was evaluated to determine: the severity grade (CTCAE Grade 1–4), its duration (start and end dates or, if continuing at the Safety Follow-up Visit, its relationship to the study treatment, that is, Reasonable possibility that AE is related: No, Yes), action taken with respect to study or investigational treatment (none, dose adjusted, temporarily interrupted, permanently discontinued, hospitalized, unknown, not applicable), whether medication or therapy was given (no concomitant medication/non-drug therapy, concomitant medication/non-drug therapy), outcome (not recovered/not resolved, recovered/resolved, recovering/resolving, recovered/resolved with sequela, fatal, unknown), and seriousness, where a SAE was defined.

An SAE was defined as an AE that constituted a congenital anomaly/birth defect, resulted in persistent or significant disability/incapacity, or required inpatient hospitalization or prolongation of existing hospitalization, unless hospitalization was for (i) routine treatment or monitoring of the studied indication, not associated with any deterioration in condition; (ii) elective or pre-planned treatment for a pre-existing condition unrelated to the indication under study and that has not worsened since signing the informed consent; (iii) treatment on an emergency outpatient basis for an event not fulfilling any of the definitions of an SAE given above and not resulting in hospital admission; (iv) social reasons and respite care in the absence of any deterioration in the patient's general condition, whether fatal or life-threatening, and whether medically significant (i.e. defined as an event that jeopardized the patient or that may require medical or surgical intervention to prevent one of the outcomes listed above).

AEs that were Grade 1–2 and expected were reported through the Rare Diseases Clinical Research Network (RDCRN) database within 20 working days and logged appropriately. Determination of causality and relatedness were made by the site PI. AEs that were Grade 1–2 and unexpected were reported through the RDCRN database within 24 h, to local IRBs, and to the sponsor only if the event suggested that there were greater risks to study subjects than previously suspected. AEs that were Grade 3–4 were considered serious if they fit into one of the categories listed above (as defined by the FDA). If the event was classified as an SAE, the appropriate reporting guidelines were followed. All SAEs were further classified as SAEs or serious adverse drug reactions based on the events relation to the study drug. For SAEs, the final determination of causality was made by the Medical Review Officer. If the event was unrelated to the drug (probably not related or definitely not related), it was defined as an SAE. If the event was related to the drug (definitely

related, probably related, or possibly related), it was defined as a serious adverse drug reaction.

3.1.3. Primary efficacy outcomes

The primary efficacy outcome, a composite neurocognitive score, was generated based on performance on objective measures and was computed in two ways. The first was an average of measures evaluating working memory (SB-5 working memory subscale), processing speed (CPT-3 mean reaction time), and fine motor skills (Purdue Pegboard Test average of both hands) [2,24]. The second included the above average weighted by 2/3 and an average of the remaining standardized, norm-referenced neurocognitive measures (e.g., non-verbal ability, visuomotor skills, verbal learning, receptive and expressive language) weighted by 1/3. The first method was used if nearly complete data was observed on the three primary constructs (working memory, processing speed, and fine motor skills) in randomized patients. The second was used if missing data on the working memory/processing speed/motor composite was present in greater than 20% of randomized patients.

Neurocognitive measures were completed at baseline, 3-month, and 6-month timepoints. Alternate forms were used were possible to reduce practice effects. In order for any patient to be included in the study, they were required to complete measures from at least one of the three primary efficacy constructs. If patients could not complete the full SB-5, the Mullen Scales of Early Learning could be substituted. In a similar fashion, if the CPT-3 could not be completed, the processing speed index of the Wechsler scales could be substituted as an alternative measure of processing speed.

Measures were chosen based on previous empirical findings in a cross-sectional cohort study of *PTEN*-ASD [6,25]. In this study, measures of working memory and processing speed were differentially impaired relative to other cognitive measures, including IQ. Furthermore, the cognitive deficits seen in these patients were related to reduced *PTEN* protein levels and brain white matter abnormalities. Specifically, in a cross-sectional mediational model, reduced *PTEN* protein levels led to greater brain white matter abnormalities which, in turn, led to greater cognitive difficulties. Although not evaluated using standardized testing, in our prior cohort of *PTEN* patients with ASD, every patient had a history of fine motor difficulties, occupational therapy, and/or observations during testing of significant fine motor weaknesses. Motor difficulties are also consistent with brain white matter abnormalities [6]. Thus, the neurocognitive index included measures that had both empirical and biological bases for inclusion as outcome measures.

In the second version of the neurocognitive composite, the primary justifications for including the remaining neurocognitive measures were: 1) if significant missing data was seen on measures of processing speed, working memory, and fine motor skills, including a minor component of additional neurocognitive measures would ensure that more reliably obtained data was contributing to the measurement of change, and 2) it is reasonable to expect a range of neurocognitive functions might improve since *PTEN* expresses in all of the cells of the body, the brains of *PTEN* patients have shown widespread changes that included white matter abnormalities but also regional gray matter changes, and mouse models of *PTEN* loss have suggested that dendrite and synaptic dysfunction is also a consequence. Inclusion of other neurocognitive measures as the minority part of the index ensured that improvements in other domains, such as expressive language, also had an opportunity to contribute to individual patient outcome measurements.

3.1.4. Secondary objectives and outcomes

Our secondary objective was to evaluate the efficacy of everolimus on neurocognition and behavior in individuals with *PTEN* compared to placebo. The secondary efficacy endpoints examined the effect of everolimus on overall global clinical improvement, autism symptoms, other behavioral problems, and adaptive abilities. These endpoints include validated, standardized instruments including the Clinical Global Impression – Improvement scale (CGI-I), Autism Diagnostic Observation Schedule – Second Edition (ADOS-2 calibrated severity score), Social Responsiveness Scale – Second Edition (SRS-2), Adult/Child Behavior Checklist (CBCL ACF/CBCL), Wide Range Assessment of Memory and Learning – Second Edition (WRAML-2), and Vineland Adaptive Behavior Scales – Third Edition (VABS-III). The Clinical Global Impression – Severity scale (CGI-S) and CGI-I were completed by trained neuropsychologists or physicians with a focus on cognitive functioning after discussion with the patient and legal guardian as well as the review of the cognitive testing completed during the in-person visits. Efficacy measures included in the secondary outcome analyses are listed in Table 4.

The secondary safety endpoint was the Dosage Record and Treatment Emergent Symptoms (DOTES) scale, which has been widely used clinically for both children and adults to assess many central nervous system side effects as well as some behavioral side effects. It was completed at screening, baseline, and at each subsequent visit. In addition, patients were interviewed using the C-SSRS to proactively assess patients for suicidality and mood disturbances. If the patient's and/or parents' responses were of concern, the investigator performing the interview could refer them to a mental health professional at the hospital. Specific AEs were also classified and graded according to CTCAE. Use of CTCAE allowed direct comparison with past and current clinical trials with everolimus, including the neurocognitive trial described in previous sections (NCT01289912). Vital signs (blood pressure, pulse, and temperature), height, and weight were obtained during screening, baseline, and at each subsequent visit.

Appendix 2 elaborates on the collection and analysis of electroencephalogram and auditory evoked potential measures, as well as the plan to collect other measures.

3.2. Sample size calculation and power estimation

This is the first study to evaluate the efficacy of everolimus in patients with a germline heterozygous *PTEN* mutation. As such, analyses were predominantly descriptive with the intent of generating plausible hypotheses to be tested in a Phase III confirmatory trial. The planned recruitment sample size was based on three considerations: 1) the expected recruitment potential of the three sites, 2) a desire to ensure sensitivity to any enrichment in side effects or problems with tolerability in everolimus-treated patients, and 3) the need for adequate statistical power in examining primary and secondary efficacy outcomes.

For power estimation purposes, the variability in *PTEN*-ASD patients treated with an mTOR inhibitor was expected to be similar to the untreated sample. Preliminary data indicated that the measures comprising the dominant portion of the neurocognitive composite index (processing speed working memory, and fine motor) are highly sensitive to the brain dysfunction seen in PHTS patients with ASD. The effect sizes were approximately 1 standard deviation (SD) or larger. In a comparison between mTOR treated versus untreated PHTS with ASD patients, an effect size of at least 0.65 SD would be of interest and represent clinically meaningful improvement. For example, an effect size of at least 0.65 SD on cognitive measures would translate to a 10-point standard score improvement in the everolimus-treated group relative to placebo. An improvement of 10 or more points is approximately 2–3 times the standard error of measurement for most cognitive measures and is typically considered a reliable change when disordered populations are re-tested [27,28].

A sample size of 40 (20 per arm) patients permitted detection of endpoint (6-month follow-up) differences of at least 0.80 SD at the 5% significance level (one-sided $\alpha = 0.05$) with 80% power ($1 - \beta$). Statistical power would be weaker ($1 - \beta = .65$) if smaller cross-sectional differences were observed (0.65 SD). However, statistical power is very strong, even for smaller differences (Cohen's $d \geq .50$ equivalent to $\geq .50$ SD difference) when the full repeated measures nature of the design is considered—baseline, 3-month follow-up, 6-month follow-up.

Table 3
Open-label treatment phase study timeline.

Measurement	Baseline[b] (Visit 9)	Month 1 (Visit 10)	Month 2[d] (Visit 11)	Month 3 (Visit 12)	Month 4[d] (Visit 13)	Month 5[d] (Visit 14)	Month 6 (Visit 15)	Follow-up	PRN
Visit Windows	Within 2 weeks after double-blind treatment	±14d	±14d	±14d	±14d	±14d	±14d	28d after last visit ±14d	
Medical History	#	X	X	X	X	X	X	X	
Clinical Interview	#	X	X	X	X	X	X		
Inclusion/Exclusion criteria	X								
CGI Scale: Severity & Improvement	X[c]	X	X	X	X	X	X		
Vital Signs		X		X			X		
Physical/Neurological Exam							X		X
Dermoscopy				X			X		
Microbiome/Mycobiome Sample Collection				X			X		
Tanner Staging							X		
Side Effects (CTCAE v5.0 + DOTES)	#	X	X	X	X	X	X	X	
Columbia-Suicide Severity Rating Scale		X		X			X		
Laboratory Tests[a]	#	X		X			X		X
Everolimus Level (Only Done After M1 If Needed)		X							X
PTEN-associated Proteins				X			X		
Concomitant Treatment Log	#	X	X	X	X	X	X	X	
Primary Efficacy Outcome Neurocognitive Composite	#			X			X	X	
CPT-3/K-CPT-2)				X			X		
SB-5/Mullen Scales of Early Learning (MSEL)							X		
SB-5 Working Memory Subscale Only				X					
Purdue Pegboard (PP)				X			X		
Wechsler Processing Speed Index				X			X		
Wide Range Assessment of Memory and Learning-2 (WRAML-2)				X			X		
Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4)				X			X		
Expressive Vocabulary Test – Second Edition (EVT-2)				X			X		
Autism Diagnostic Observation Schedule (ADOS-2)							X		
Social Responsiveness Scale – Second Edition (SRS-2)				X			X		
Repetitive Behavior Scale – Revised (RBS-R)				X			X		
Developmental Coordination Disorder Questionnaire (DCDQ)				X			X	X	
Behavior Rating Inventory of Executive Function (BRIEF)				X			X		
Adult/Child Behavior Checklist				X			X	X	
Sensory Profile Questionnaire - Short Form (SPQ)				x			x		
Vineland Adaptive Behavior Scales (VABS-III) – caregiver report				X			X		
Eye Tracking (Optional)				X			X		
Resting State EEG/Auditory Evoked Potentials (Optional)				X			X		

**: primary outcome mesaure only.

[a] Coagulation testing will only be performed at screening, month 3 and month 6. Pregnancy testing, serum is done at screening and month 6 and urine may be used for onsite visits (if deemed clinically necessary) for women of childbearing potential.

[b] Monthly visit windows are calculated based on date of Baseline visit.

[c] At the baseline visit, only the CGI: Severity scale will be completed. CGI: Clinical Global Impressions Scale; IE: Independent Evaluator; CPT: Conners' Continuous Performance Test; CTCAE: Common Terminology Criteria for Adverse Events; DOTES: Dosage Record and Treatment Emergent Symptom Scale; SB-5: Stanford-Binet Intelligence Scales, Fifth Edition; PRN: *Pro* Re *Nata* (visits as needed).

[d] May be performed by phone or in person; #: procedures may be carried over from the double-blind Month 6 visit, if within 2 weeks of the visit.

Assuming even modest correlations between repeated measurements of outcome variables ($r \geq .30$), power to detect clinically meaningful differences of .50 SD is very good ($1 - \beta = .83$; one-sided $\alpha = .05$). If the primary outcome measure (neurocognitive composite) meets nominal statistical significance in the expected direction (one-tailed $p < .05$) or a secondary outcome measure meets a more stringent false discovery rate-corrected significance level in the expected direction, these outcomes will be considered candidates for Phase III confirmatory trial evaluation.

Assuming an overall dropout rate of approximately 10%, 44 patients needed to be enrolled to get 40 patients with complete trial data. However, following intent-to-treat procedures, all patients who were randomized, dosed, and for whom baseline data was collected were

Table 4
Primary and secondary efficacy outcome measures.

	Domain	Measure
Primary efficacy outcome	Neurocognitive Composite	Average working memory (SB-5 working memory), processing speed (CPT-3 mean reaction time) and fine motor (Purdue Pegboard-average of both hands) subtests weighted at 2/3 of the composite score and the average of all other available neurocognitive testing measures (receptive and expressive language, non-verbal ability, verbal learning, sustained attention, impulsivity, and visuomotor skills) weighted at 1/3 of the score.
Secondary efficacy outcomes	Global Ability	SB-5: Full scale IQ, Verbal IQ, and Non-Verbal IQ or the Mullen Scales of Learning: Cognitive IQ
	Attention	CPT-3 (CPT-3/K-CPT-2): Discriminability (d') and Omissions
	Processing Speed	CPT-3 (CPT-3/K-CPT-2): Mean Reaction Time; in cases where an individual cannot complete the CPT, we will administer the appropriate Wechsler processing speed index subtest.
	Impulsivity	CPT: Commissions and Bias
	Long-Term Memory	WRAML-2 Verbal Learning: Scaled score
	Language	Peabody Picture Vocabulary Test −4 (PPVT-4): Standard Score Expressive Vocabulary Test – 2 (EVT-2): Standard Score
	Motor Functioning	Purdue Pegboard (Pegs): Dominant and non-dominant hand standard scores
	Motor Coordination	Developmental Coordination Disorder Questionnaire (DCDQ): Total score
	Autism Symptoms	Autism Diagnostic Observation Schedule (ADOS): Calibrated severity score Social Responsiveness Scale – 2 (SRS-2): Total T-score Repetitive Behavior Scale – Revised (RBS-R): Total raw score
	Other Behavioral Symptoms	Behavior Rating Inventory of Executive Function (BRIEF) - Global Executive Composite: Standard Score Child Behavior Checklist - Total Problems: Standard Score
	Sensory Processing	Short Sensory Profile (SSP): Total Score
	Adaptive Behavior	Vineland Adaptive Behavior Scales (VABS-III): Composite Standard Score
	Global Severity and Improvement	Clinical Global Impressions – Severity (CGI-S) Clinical Global Impressions – Improvement (CGI-I)

CPT-3: Conners' Continuous Performance Test, Third Edition; SB-5: Stanford-Binet Intelligence Scales, Fifth Edition; WRAML-2: Wide Range Assessment of Memory and Learning, Second Edition.

included in statistical analyses. The sample size of 40 is underpowered to detect small differences in dropout rates or side effects between everolimus versus placebo-treated groups. Statistical power would only be adequate if at least a 35% difference in the proportion of dropout/side effects was observed ($1 - \beta \geq .81$), assuming a one-tailed Type 1 error rate of $\alpha = .05$. This is equivalent to 1/20 patients experiencing an AE/side effect in the placebo arm and 8/20 in the treated arm. Similarly, statistical power of survival analyses would only be fair ($1 - \beta \geq .57$) if a hazard ratio of 7.5 was observed, equivalent to a 5% dropout rate in the placebo group versus a 30% dropout rate in the everolimus group. However, because the goal was to maintain sensitivity to any potential enrichment in dropout or side effects, any indicator showing a difference of 10% or greater was considered a meaningful difference and was reported.

3.3. Data analysis plan and considerations

We will examine the safety and tolerability of everolimus by comparing the 1) dropout rates due to side effects, 2) dropout rates for any reason, and the 3) rates of specific side effects in those receiving everolimus versus placebo. Based on the published literature and our experience in the TSC trial [18], we predict that the dropout rates due to side effects will be very similar across everolimus and placebo groups ($< 10\%$ difference). However, we predict that everolimus will cause higher rates of medication-related side effects ($\geq 10\%$ difference) when compared to placebo. We will use chi-square statistics or Fisher's exact test to compare dropout rates and rates of side effects. Statistical significance will be determined using a one-tailed test with Type 1 error rate of $\alpha = .05$, as only increases with everolimus treatment are expected. No false discovery rate correction will be applied because the concern is maximizing sensitivity to any increase in side effect prevalence with everolimus treatment. For dropout rates, we will also compute Kaplan-Meier survival analyses with discontinuation due to side effects or any reason as separate status endpoints. Number of days from baseline to medication discontinuation will be the time variable. Survival analyses are expected to be under-powered due to the generally low dropout rates expected, but these analyses can be useful for describing temporal trends. To ensure any observed group differences are not influenced by other variables or randomization imbalance, Cox and logistic regression analysis will also be conducted with and without conditioning on relevant baseline covariates (e.g., demographics, language and cognitive ability, symptom severity).

We will also examine whether the everolimus group will show more improvement, compared to the placebo group, on the main neurocognitive efficacy composite outcome as well as on the secondary efficacy outcomes. Treatment group (everolimus versus placebo) differences on the neurocognitive composite scores and the secondary efficacy outcomes will be analyzed using mixed-effects models to fully utilize repeated measurements collected at baseline, 3-month follow-up, and 6-month follow-up. Changes from baseline to 3-month follow-up and/or 6-month follow-up the neurocognitive composite scores and the secondary efficacy outcomes will be tested. A significant interaction between time and treatment group, with the everolimus group showing a more favorable outcome trajectory, will support the primary efficacy hypothesis. The false discovery rate of multiple testing on the secondary efficacy outcomes will be controlled by using the Benjamini-Hochberg procedure.

Our data analyses will be conducted following the intention-to-treat principle. Data points that are missing due to subject attrition will be handled assuming that data are missing at random (MAR) conditional on observed information, which is less restrictive than missing completely at random (MCAR). In this procedure, all available cases including the ones with missing information will be included in the analyses. By including every subject who completed at least one follow-up assessment, we are not only more likely to conserve power, but also less likely to produce biased effect estimates. In mixed effects analyses, the slope of the outcome will be modeled as the key dependent variable predicted by treatment group. We will conduct the analysis with and without conditioning on relevant baseline covariates. The results of these longitudinal analyses can be easily converted to a cross-sectional group effect at each assessment time point to describe the magnitude and significance of group differences. Of particular interest is the group difference at end of treatment (6-month assessment). The mixed effects analyses will be repeated using a model where the baseline scores on each outcome are treated as a baseline covariate to control for (instead of as a part of) the repeated measures. This model can be useful if there is a considerable heterogeneity between before and after treatment processes. We will also analyze the data using analysis of covariance (ANCOVA) treating the neurocognitive composite scores as the dependent variable and controlling for the baseline neurocognitive composite scores. Further analyses will be conducted using the same analysis models but with

transformed data (e.g., log transformation) to check the sensitivity of the results to deviation from outcome normality. Statistical significance of all main effects and interaction terms as well as cross-sectional group differences will be determined using a one-tailed Type 1 error rate of $\alpha = .05$. The same analysis strategy will be employed for secondary efficacy outcomes and the Benjamini-Hochberg procedure as a false discovery rate correction will be implemented to control inflation of Type 1 error.

If the exploratory biosample objectives are pursued, we will explore whether baseline peripheral blood levels of *PTEN*-associated pathway molecules (PI3K/AKT, mTOR, MAPK, PS6K/S6K protein levels) change with treatment and if changes in peripheral pathway molecules correlate with clinical improvement. To determine whether *PTEN*-associated pathway molecules change with treatment, we will use a similar mixed effect (growth curve) modeling approach to that described above for secondary objective with screening, 3-month, and 6-month pathway measurements as outcome measures. The time-by-treatment group interaction will test whether the everolimus treatment group shows improvement/normalization of *PTEN*-pathway molecule levels relative to placebo. This component will be dependent on securing additional funds.

4. Discussion

The benefits of this study include potential improvement of cognitive and behavioral symptoms in individuals with germline heterozygous *PTEN* mutations. The information gathered from this study may help PHTS patients who are impacted by behavioral and cognitive dysfunctions by better describing the neurocognitive deficits and the safety and efficacy of available treatments. EEG recordings and eye tracking data obtained at baseline and at the end of trial can provide additional, more direct measures of the potential impact of everolimus on brain function. Biological specimens that obtained from skin or leftover tissue obtained by following a clinically indicated procedure can be used in future genetic and molecular studies.

The design of this trial is innovative in that it used a neurocognitive index as the main efficacy measure. The index was developed based on preliminary data obtained from individuals with *PTEN* mutations and optimized the chances of detecting a change with everolimus treatment [2,6,24]. Additionally, the secondary outcome measures allow the casting of a wide net to capture any improvement with the medication. Finally, the inclusion of biological measures at baseline and at the end of trial shed light on the neurobiology resulting from *PTEN* mutations and the changes in this biology in response to everolimus. All of these measures—biologic, cognitive, and behavioral—will allow the examination of predictors of response and help in identifying the group of patients who are more likely to respond to this medication.

This research study is the first trial examining the safety and efficacy of everolimus in individuals with *PTEN* mutation. This trial is broad in reach, as it includes biologic, cognitive, and behavioral outcomes which have been understudied in the *PTEN* literature to date. It has been designed to optimize the ability to detect changes in the *PTEN* presentation that are sensitive to pharmacologic intervention. If successful, this trial will detect a promising candidate to address the pervasive neurobehavioral deficits seen in PHTS patients and will have identified the first established personalized medicine for a genetic syndrome associated with ASD. A potential limitation of our clinical trial on everolimus is its duration. A 6-month trial can only evaluate short-term safety and efficacy of the drug. However, the duration of this trial has been longer than several other Phase II trials in similar neurodevelopmental disorders [29,30]. We hesitated to keep the participants on the drug that is potential toxic for a long duration without some evidence of efficacy. This is a Phase II trial that services as a precursor of the confirmatory Phase III trial. A longer period can be designed for a Phase III study or a post-market Phase IV study.

Abbreviations

ADOS-2	Autism Diagnostic Observation Schedule – Second Edition
AE	adverse event
AEP	auditory evoked potentials
ASD	autism spectrum disorder
CBCL ACF/CBCL	Adult/Child Behavior Checklist
CGI-I	Clinical Global Impression – Improvement scale
CGI-S	Clinical Global Impression – Severity scale
CPT-3	Conners' Continuous Performance Test Third Edition
C-SSRS	Columbia-Suicide Severity Rating Scale
CTCAE	Common Terminology Criteria for Adverse Events
DOTES	dosage record treatment emergent symptom scale
EEG	electroencephalography
mTORC1	mTOR complex 1
SAE	severe adverse event
SB-5	Stanford-Binet Intelligence Scales, Fifth Edition

Funding

The Developmental Synaptopathies Consortium (U54NS092090) is part of the National Center for Advancing Translational Sciences (NCATS) Rare Diseases Clinical Research Network (RDCRN) and is supported by the RDCRN Data Management and Coordinating Center (DMCC) (U2CTR002818). RDCRN is an initiative of the Office of Rare Diseases Research (ORDR), NCATS, funded through a collaboration between NCATS and the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (NINDS), Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) and National Institute of Mental Health (NIMH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH).

Declaration of competing interest

Authors declare no conflict of interest.

Acknowledgments

We are sincerely indebted to the generosity of the families and patients in *PTEN* clinics across the United States who contributed their time and effort to this study. We would also like to thank the *PTEN* Hamartoma Syndrome Foundation, *PTEN* Research and Novartis Pharmaceuticals for their support in this trial.

SRS-2 Social Responsiveness Scale – Second Edition
TSC tuberous sclerosis complex
VABS-III Vineland Adaptive Behavior Scales, Third Edition
WRAML-2 Wide Range Assessment of Memory and Learning, Second Edition

Appendix 1. Open-label phase, follow-up phase, everolimus administration and treatment course

Open-label phase

Individuals in the placebo group were invited to enter the 6-month open-label extension trial that would follow the same schedule as the blinded treatment phase (timeline shown in Table 3). The final visit (month 6) in the blinded treatment phase served as the baseline visit of the open-label phase if the participant chose to enter the open-label phase, and safety labs obtained at this time were used as baselines for the open-label phase. Similarly, the scores from the patients' neuropsychological assessments were used as both the 6-month blinded treatment phase data and the baseline open-label phase data. Study staff dispensed open-label medication at this visit. If a patient was unable to enter the open-label phase at the final visit of the blinded treatment phase, he or she could come back within 2 weeks ($\pm$2 weeks) of this visit date. Treatment, assessments, and outcomes of the open-label phase were also identical to the double-blind phase.

The next visit was considered the month 1 visit of the open-label phase. A PK sample, as well as specified safety labs, were drawn on site and sent to LabConnect, LLC. If the participant received a dose adjustment at the time of this visit, another sample could be taken at the study site or locally by a trained phlebotomist. The visits at months 3 and 6 included safety labs, developmental testing, optional electroencephalography (EEG)/auditory evoked potentials (AEP) and eye tracking procedures, physical and neurological exams, vital signs, and other study questionnaires. A PK sample could also be collected from a patient experiencing an AE or side effect. The decision to obtain this sample was made by the site physician.

If unforeseen circumstances (i.e. unexpected personal reasons) prevented the patient from complying with the established visit schedule, the site could re-schedule the visit within 14 days of the expected visit date. The reason(s) for any visit or treatment delays were documented in the case report forms for the appropriate visit. All on-site visits (months 1, 3, and 6) were required to be completed by the participant; any missed visits would warrant termination from the study.

At the final visit of the open-label phase, or the month 6 visit, participants exited the trial.

Follow-up phase

All patients had a follow-up phone call scheduled 28 days ($\pm$14 days) after the last dose of the study treatment to follow for AEs and severe adverse events (SAEs) that may have occurred after discontinuation from the study treatment.

Everolimus administration and treatment course

In this study, everolimus and placebo were formulated as identical tablets of 2.5 mg or 5 mg strength, blister-packed under aluminum foil in units of 10 tablets. Medication labels complied with U.S. legal requirements for investigational drug products, were printed in English, and included expiration date and storage conditions. For the duration of the trial, everolimus and placebo were supplied to the research pharmacies at each site directly from Novartis.

Everolimus and/or placebo were self-administered (by the patient or patient's parent/guardian). Everolimus was administered orally once daily at the same time every day, consistently with a light, low-fat meal. Everolimus or placebo tablets were to be opened only at the time of administration because the drug is both hygroscopic and light sensitive. The extent of absorption of everolimus through topical exposure is not known; therefore, patients/caregivers were advised to avoid contact with the everolimus or placebo tablets and to wash their hands thoroughly before and after administration.

The average starting dose was 4.5 mg/m^2/day of trial therapy (either everolimus or placebo), rounded to the nearest 2.5 mg dose. The patient's body surface area was calculated based on an accurate height and weight measurement performed according to institutional guidelines. Leftover study medication and all used blister packs were collected at each study visit, and drug was accounted for at this time. If 2.5 mg tablets become unavailable during the trial, patients prescribed a dose including 2.5 mg tablets were instructed to alternate between a higher and lower dose. The site physician instructed the patient to take the higher dose on Monday, Wednesday, and Friday and the lower dose on Tuesday, Thursday, Saturday, and Sunday. For example, a patient prescribed a dose of 7.5 mg would take 10 mg on Monday, Wednesday, and Friday, and 5 mg on Tuesday, Thursday, Saturday, and Sunday.

A maximum volume of 3 ml of blood was drawn for trough everolimus PK levels when necessary. The blood draws were timed to occur at 24 $\pm$ 4 h after ingestion of the drug. Blood collection for PK samples could be conducted during a scheduled visit on-site or via a designated and trained remote healthcare service. Everolimus levels were measured at ARUP Laboratories in Salt Lake City, UT, in conjunction with LabConnect, LLC. When necessary, a kit for remote collection of blood samples could be sent to the participant and returned per the shipping instructions by express mail.

Dose adjustments were permitted based on safety findings and PK levels. The site PI initiated any safety level related dose adjustments according to the protocol, in collaboration with the Medical Review Officer as appropriate. PK-based dose adjustments were initiated by an unblinded physician. All dose adjustments were made using the 2.5 mg tablets (i.e. increase by 2.5 mg, decrease by 2.5 mg, or maintain dose). If a dose adjustment was made for safety purposes, the unblinded physicians and Medical Review Officer were made aware of any dose modifications as soon as possible.

Appendix 2. Electroencephalogram, auditory evoked potential, and other measures

Resting state electroencephalogram (EEG) and auditory evoked potential (AEP) procedures were optional. If the parents/patients opted in on the consent form, resting state EEG and AEP data collection occurred at baseline, 3-month, and 6-month visits. Resting State EEG Procedures: Resting state EEG has been used increasingly to quantitatively characterize and track outcomes in a range of neuropsychiatric populations, including idiopathic ASD [26]. EEG leads were placed on the participant's head, and EEG data were collected while participants were presented with non-social, abstract

moving images in random order. The resting state EEG paradigm takes about 5 min. AEP Procedures: AEP data were collected using the same system as the resting state EEG, with clicks as the auditory input. Subjects passively listened to 150 sets of two 5 ms broadband noise bursts (65 dB) separated by an inter-stimulus interval of 500 ms with inter-set intervals of 4000 ms (total duration approximately 12 min). Offline, data were average referenced and filtered for time-frequency analyses.

Along with the EEG/AEP data, photographs were taken of the electrode placements on the participant's head. These photos allowed the research staff to view electrode placement and make informed decisions concerning poor data in relation to incorrect electrode placement. Participants were also videotaped during the EEG and AEP procedures. By videotaping participants, researchers were able to determine whether abnormal EEG data may be due to the participant's behavior.

Other measures were collected on dermoscopy, eye tracking, vital signs, body surface area, laboratory evaluations, hepatitis and HIV screening, hematology tests, coagulation, serum pregnancy and hormone testing, biochemistry and lipid profile, HbA1c, urinalysis assessment, everolimus levels in blood, blood sample for *PTEN* associated proteins, blood sample for insulin like growth factor binding protein 2 (IGFBP-2), microbiome and mycobiome sample, medical records, and clinical reports.

In analyzing the resting state EEG and AEP data, we will focus on the following: baseline EEG gamma power, AEP N1 habituation, and AEP variability (phase-locking factor, also called intertrial phase coherence). To prepare EEG and AEP data for analysis, the pipeline imports EEG from various acquisition formats and electrode layouts into a standardized format. Data will first be processed via PREP, which removes 60 Hz (Hz) line noise, identifies and interpolates bad channels, and references to average. Data will then be high pass filtered at 1 Hz, and will undergo artifact removal via MARA, an automated independent component-based supervised machine learning algorithm that handles artifacts including ocular artifact, muscle artifact, and loose electrodes. For analysis of resting EEG data, segments of high-amplitude artifact (>150 μV) will be removed, and the remaining data will be segmented into 2-s epochs. Data will undergo a Laplacian transform, as this has been shown to reduce sensitivity of the EEG signal to contamination by myogenic activity, which is otherwise particularly prominent in the gamma band.[59,60] Spectral analysis will be performed on each epoch with multitaper methods using 3 tapers, in order to determine the average power spectrum across all epochs. This will allow for determination of power in the gamma frequency band, particularly the 62–90 Hz range.

For analysis of AEP data, data will be epoched into 2000 ms trials (− 500–1500 ms), and baseline corrected using the 500 ms period prior to the first auditory stimulus in each trial. Any trial with amplitude exceeding 150 μV will be removed. N1 will be defined as the most negative-going waveform deflection between 50 and 150 ms post-stimulus, over the central region. (If N1 cannot be adequately identified in this population using these criteria, independent components analysis will be used to identify the N1 component instead). Habituation of the N1 amplitude will be quantified as percent change from the first to second auditory click in each trial. Intertrial phase coherence (ITPC) will be calculated using the function *newtimef*, and reported as the maximum absolute value ITPC in the 50–150 ms timeframe, in the alpha frequency band (8–12 Hz).

References

[1] L. Yehia, J. Ngeow, C. Eng, PTEN-opathies: from biological insights to evidence-based precision medicine, J. Clin. Invest. 129 (2019) 452–464, https://doi.org/10.1172/JCI121277.

[2] T.W. Frazier, R. Embacher, A.K. Tilot, K. Koenig, J. Mester, C. Eng, Molecular and phenotypic abnormalities in individuals with germline heterozygous PTEN mutations and autism, Mol. Psychiatr. 20 (2015) 1132–1138, https://doi.org/10.1038/mp.2014.125.

[3] L. Yehia, E. Keel, C. Eng, The clinical spectrum of PTEN mutations, Annu. Rev. Med. 71 (2020) 103–116, https://doi.org/10.1146/annurev-med-052218-125823.

[4] M.G. Butler, M.J. Dasouki, X.-P. Zhou, Z. Talebizadeh, M. Brown, T.N. Takahashi, J.H. Miles, C.H. Wang, R. Stratton, R. Pilarski, C. Eng, Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations, J. Med. Genet. 42 (2005) 318–321, https://doi.org/10.1136/jmg.2004.024646.

[5] G.E. Herman, E. Butter, B. Enrile, M. Pastore, T.W. Prior, A. Sommer, Increasing knowledge of *PTEN* germline mutations: two additional patients with autism and macrocephaly, Am. J. Med. Genet. 143A (2007) 589–593, https://doi.org/10.1002/ajmg.a.31619.

[6] R.M. Busch, J.S. Chapin, J. Mester, L. Ferguson, J.S. Haut, T.W. Frazier, C. Eng, Cognitive characteristics of PTEN hamartoma tumor syndromes, Genet. Med. 15 (2013) 548–553, https://doi.org/10.1038/gim.2013.1.

[7] C.H. Kwon, X. Zhu, J. Zhang, S.J. Baker, MTor is required for hypertrophy of Pten-deficient neuronal soma in vivo, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 12923–12928, https://doi.org/10.1073/pnas.2132711100.

[8] C.H. Kwon, X. Zhu, J. Zhang, L.L. Knoop, R. Tharp, R.J. Smeyne, C.G. Eberhart, P. C. Burger, S.J. Baker, Pten regulates neuronal soma size: a mouse model of Lhermitte-Duclos disease, Nat. Genet. 29 (2001) 404–411, https://doi.org/10.1038/ng781.

[9] D.T. Page, O.J. Kuti, C. Prestia, M. Sur, Haploinsufficiency for Pten and Serotonin transporter cooperatively influences brain size and social behavior, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 1989–1994, https://doi.org/10.1073/pnas.0804428106.

[10] J.L. Mester, A.K. Tilot, L.A. Rybicki, T.W. Frazier, C. Eng, Analysis of prevalence and degree of macrocephaly in patients with germline PTEN mutations and of brain weight in Pten knock-in murine model, Eur. J. Hum. Genet. 19 (2011) 763–768, https://doi.org/10.1038/ejhg.2011.20.

[11] A.K. Tilot, M.K. Gaugler, Q. Yu, T. Romigh, W. Yu, R.H. Miller, T.W. Frazier 2nd, C. Eng, Germline disruption of Pten localization causes enhanced sex-dependent social motivation and increased glial production, Hum. Mol. Genet. 23 (2014) 3212–3227, https://doi.org/10.1093/hmg/ddu031.

[12] A.K. Tilot, G. Bebek, F. Niazi, J.B. Altemus, T. Romigh, T.W. Frazier, C. Eng, Neural transcriptome of constitutional Pten dysfunction in mice and its relevance to human idiopathic autism spectrum disorder, Mol. Psychiatr. 21 (2016) 118–125, https://doi.org/10.1038/mp.2015.17.

[13] A.K. Tilot, T.W. Frazier 2nd, C. Eng, Balancing proliferation and connectivity in PTEN-associated autism spectrum disorder, Neurotherapeutics 12 (2015) 609–619, https://doi.org/10.1007/s13311-015-0356-8.

[14] A. Amiri, W. Cho, J. Zhou, S. Birnbaum, C. Sinton, R. McKay, L. Parada, Pten deletion in adult hippocampal neural stem/progenitor cells causes cellular abnormalities and alters neurogenesis, J. Neurosci. 32 (2012) 5880–5890, https://doi.org/10.1523/JNEUROSCI.5462-11.2012.

[15] J. Zhou, J. Blundell, S. Ogawa, C.H. Kwon, W. Zhang, C. Sinton, C.M. Powell, L. F. Parada, Pharmacological inhibition of mTORC1 suppresses anatomical, cellular, and behavioral abnormalities in neural-specific Pten knock-out mice, J. Neurosci. 29 (2009) 1773–1783, https://doi.org/10.1523/JNEUROSCI.5685-08.2009.

[16] C.H. Kwon, B.W. Luikart, C.M. Powell, J. Zhou, S.A. Matheny, W. Zhang, Y. Li, S. Baker, L.F. Parada, Pten regulates neuronal arborization and social interaction in mice, Neuron 50 (2006) 377–388, https://doi.org/10.1016/j.neuron.2006.03.023.

[17] A. Boulay, H.A. Lane, The mammalian target of rapamycin kinase and tumor growth inhibition, Recent Results Canc. Res. 172 (2007) 99–124, https://doi.org/10.1007/978-3-540-31209-3_7.

[18] D.A. Krueger, A. Sadhwani, A.W. Byars, P.J. de Vries, D.N. Franz, V.H. Whittemore, R. Filip-Dhima, D. Murray, K. Kapur, M. Sahin, Everolimus for treatment of tuberous sclerosis complex-associated neuropsychiatric disorders, Ann Clin Transl Neurol 4 (2017) 877–887, https://doi.org/10.1002/acn3.494.

[19] G.H. Roid, M. Pomplun, The stanford-binet intelligence scales, in: D.P. Flanagan, P. L. Harrison (Eds.), Contemporary Intellectual Assessment: Theories, Tests, and Issues, fifth ed., The Guilford Press, New York, NY, 2012, pp. 249–268.

[20] W.J. Chen, C.K. Hsiao, L.L. Hsiao, H.G. Hwu, Performance of the continuous performance test among community samples, Schizophr. Bull. 24 (1998) 163–174, https://doi.org/10.1093/oxfordjournals.schbul.a033308.

[21] J. Desrosiers, R. Hebert, G. Bravo, E. Dutil, The Purdue Pegboard Test: normative data for people aged 60 and over, Disabil. Rehabil. 17 (1995) 217–224, https://doi.org/10.3109/09638289509166638.

[22] K. Posner, D. Brent, C. Lucas, M. Gould, B. Stanley, G. Brown, P. Fisher, J. Zelazny, A. Burke, M.J. Oquendo, J. Mann, Columbia-suicide Severity Rating Scale (C-SSRS), Columbia University Medical Center, New York, NY, 2008.

[23] C.A. Garvey, D. Gross, L. Freeman, Assessing psychotropic medication side effects among children. A reliability study, J. Child Adolesc. Psychiatr. Ment. Health Nurs. 4 (1991) 127–131, https://doi.org/10.1111/j.1744-6171.1991.tb00509.x.

[24] R.M. Busch, S. Srivastava, O. Hogue, T.W. Frazier, P. Klaas, A. Hardan, J. A. Martinez-Agosto, M. Sahin, C. Eng, Developmental Synaptopathies Consortium, Neurobehavioral phenotype of autism spectrum disorder associated with germline heterozygous mutations in PTEN, Transl. Psychiatry 9 (2019) 253, https://doi.org/10.1038/s41398-019-0588-1.

[25] T.W. Frazier, Autism spectrum disorder associated with germline heterozygous PTEN mutations, Cold Spring Harb. Perspect. Med. 9 (2019), a037002, https://doi.org/10.1101/cshperspect.a037002.

[26] J. Wang, J. Barstein, L.E. Ethridge, M.W. Mosconi, Y. Takarae, J.A. Sweeney, Resting state EEG abnormalities in autism spectrum disorders, J. Neurodev. Disord. 5 (2013) 24, https://doi.org/10.1186/1866-1955-5-24.
[27] C. G, Assessing reliable neuropsychological change, in: F. R (Ed.), Prediction in Forensic and Neuropsychology: New Approaches to Psychometrically Sound Assessment, Erlbaum, Mahwah, NJ, 2002.
[28] R. Martin, S. Sawrie, F. Gilliam, M. Mackey, E. Faught, R. Knowlton, R. Kuzniekcy, Determining reliable cognitive change after epilepsy surgery: development of reliable change indices and standardized regression-based change norms for the WMS-III and WAIS-III, Epilepsia 43 (2002) 1551–1558, https://doi.org/10.1046/j.1528-1157.2002.23602.x.
[29] E. Berry-Kravis, V. Des Portes, R. Hagerman, S. Jacquemont, P. Charles, J. Visootsak, M. Brinkman, K. Rerat, B. Koumaras, L. Zhu, G.M. Barth, Mavoglurant in fragile X syndrome: results of two randomized, double-blind, placebo-controlled trials, Sci. Transl. Med. 8 (321) (2016 Jan 13) 321ra5.
[30] D.G. Glaze, J.L. Neul, W.E. Kaufmann, E. Berry-Kravis, S. Condon, G. Stoms, S. Oosterholt, O. Della Pasqua, L. Glass, N.E. Jones, A.K. Percy, Double-blind, randomized, placebo-controlled study of trofinetide in pediatric Rett syndrome, Neurology 92 (16) (2019 Apr 16) e1912–e1925.