ARTICLES
https://doi.org/10.1038/s41591-021-01321-w

nature medicine

Check for updates

Inhibition of phosphodiesterase-4D in adults with fragile X syndrome: a randomized, placebo-controlled, phase 2 clinical trial

Elizabeth M. Berry-Kravis[1]✉, Mark D. Harnett[2], Scott A. Reines[2], Melody A. Reese[3], Lauren E. Ethridge[3,4], Abigail H. Outterson[1], Claire Michalak[1], Jeremiah Furman[1] and Mark E. Gurney[2]✉

The goal of this study was to determine whether a phosphodiesterase-4D (PDE4D) allosteric inhibitor (BPN14770) would improve cognitive function and behavioral outcomes in patients with fragile X syndrome (FXS). This phase 2 trial was a 24-week randomized, placebo-controlled, two-way crossover study in 30 adult male patients (age 18–41 years) with FXS. Participants received oral doses of BPN14770 25 mg twice daily or placebo. Primary outcomes were prespecified as safety and tolerability with secondary efficacy outcomes of cognitive performance, caregiver rating scales and physician rating scales (ClinicalTrials.gov identifier: NCT03569631). The study met the primary outcome measure since BPN14770 was well tolerated with no meaningful differences between the active and placebo treatment arms. The study also met key secondary efficacy measures of cognition and daily function. Cognitive benefit was demonstrated using the National Institutes of Health Toolbox Cognition Battery assessments of Oral Reading Recognition (least squares mean difference +2.81, $P = 0.0157$), Picture Vocabulary (+5.81, $P = 0.0342$) and Cognition Crystallized Composite score (+5.31, $P = 0.0018$). Benefit as assessed by visual analog caregiver rating scales was judged to be clinically meaningful for language (+14.04, $P = 0.0051$) and daily functioning (+14.53, $P = 0.0017$). Results from this study using direct, computer-based assessment of cognitive performance by adult males with FXS indicate significant cognitive improvement in domains related to language with corresponding improvement in caregiver scales rating language and daily functioning.

FXS is a serious genetic, neurodevelopmental disorder with severe impact on intellect, behavior and daily function. Children with FXS fall behind neurotypical children with respect to intellect and behavioral function from early in development[1]. FXS affects 1 in 4,000 US men and 1 in 8,000 US women or upwards of 60,000 people in the United States[2]. The majority of males with FXS present with moderate-to-severe intellectual disability with an IQ on average in the range of 40–45 and an average mental age of 5–6 years[2,3]. Some men with FXS acquire only rudimentary language skills and more than 50% have autism spectrum disorder (ASD). The most severely affected men with FXS may be unable to learn to use the toilet independently. FXS has severe impact on patients and on families, caregivers and the community who provide lifelong care.

The disorder is due to expansion of a CGG repeat sequence in the promoter region of the *FMR1* gene that encodes the fragile X mental retardation 1 protein (FMRP)[4,5]. Expansion of the CGG repeat beyond 200 repeats leads to methylation and full or partial silencing of *FMR1* depending on the degree of methylation. The gene is fully methylated in up to 80% of cases of FXS, which leads to an absence of FMRP. *FMR1* is expressed in most tissues in the body but the brain is the tissue most severely impacted by the absence or reduction of FMRP[2]. Since the *FMR1* gene is located on the X chromosome, males are affected more than females. Females show a broader range of disability due to mosaicism caused by X chromosome inactivation during embryonic development. Due to the randomness of X chromosome inactivation in tissue progenitor cells, brain development and function can be relatively normal in females if the mutant X chromosome is predominantly inactivated, while other females may be severely affected if the normal X chromosome is predominantly inactivated. About a third of females inheriting an *FMR1* full mutation develop intellectual disability, typically in the mild range, but disability can be moderate or severe. Another third are diagnosed with a less severe learning disability and roughly a third have normal cognition but may have emotional or behavioral issues. Average IQ in females with FXS is in the 75–80 range and 20% have ASD[2,3].

Although supportive behavioral treatment with medications addressing symptoms is available, currently there are no specific therapeutic options that target the underlying disorder or treat intellectual disability in FXS and exploration of multiple disease-targeting therapeutic strategies has not yet met with success[3]. Negative clinical trial results may reflect issues with trial design or clinical development strategy although much has been learned[6–8].

A consistent observation across patients and FXS animal models is an alteration in cyclic AMP (cAMP) metabolism. An initial report[9] of lower basal and stimulated levels of cAMP in platelets from patients with FXS compared to patients with other types of intellectual disability, such as trisomy 21, was confirmed in subsequent studies of human lymphoblastoid cell lines and fibroblasts derived from patients with FXS[10] and later with human progenitor cells carrying a mutant FXS chromosome that had been induced to form neuronal cells in cell culture[11]. Transfection of neural cells

[1]Department of Pediatrics, Neurological Sciences, and Biochemistry, Rush University Medical Center, Chicago, IL, USA. [2]Tetra Therapeutics, Grand Rapids, MI, USA. [3]Department of Psychology, University of Oklahoma, Norman, OK, USA. [4]Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA. ✉e-mail: Elizabeth_Berry-Kravis@rush.edu; mark@tetratherapeutics.com

with an *FMR1* complementary DNA (cDNA) construct and overexpression of FMRP led to increased cAMP production[12]. With the identification and molecular cloning of *FMR1*, the observation of altered cAMP metabolism was extended to a *Drosophila* FXS model carrying a deletion of the *Dfmr1* gene and to a mouse FXS model (*Fmr1*$^{-/-}$ mice) that does not express FMRP[11,13]. In *Drosophila*, deletion of *Dfmr1* impairs olfactory learning, social interaction and courting[14]. In mice, deletion of the *Fmr1* gene causes hyperactivity and impairs social interaction[15]. In both the *Drosophila* and mouse models of FXS, cAMP is decreased in the brain and behavioral deficits are reversed by genetic or pharmacological manipulations that restore cAMP levels[13,14,16].

In the brain, associative signaling during memory formation through the *N*-methyl-D-aspartate receptor (NMDAR), by increasing intracellular Ca^{++}, increases intracellular cAMP through the activation of calcium/calmodulin-dependent adenylate cyclases (ADCY1 and ADCY8)[17–20]. This links Ca^{++} influx through the NMDAR to the early stages of learning and memory which depend on cAMP signaling through protein kinase A (PKA) and phosphorylation of the cAMP response element-binding protein (CREB)[20–23]. Associative signaling through the NMDAR receptor occurs in the cortex and hippocampus at excitatory synapses on the dendritic spines of pyramidal neurons. Dendritic spines are abnormal in the human FXS brain and in the mouse *Fmr1*$^{-/-}$ model[24–26]. Rather than maturing normally, dendritic spines remain long, thin and filopodia-like.

cAMP-specific 3′,5′-cyclic PDE4D is a key modulator of cAMP levels relevant to learning and memory. The importance of PDE4D in the human brain is underscored by the discovery that ultrarare missense mutations in *PDE4D* cause acrodysostosis type 2 with or without hormone insensitivity (ACRDYS2), a neurodevelopmental disorder associated with intellectual disability (IQ ranging from 24 to 85)[27–30]. Positron emission tomography imaging of PDE4D in the primate brain with a selective, radiolabeled ligand shows the highest levels of expression in the prefrontal cortex and hippocampus[31], two regions of the brain that are important for intellect and where dendritic spine pathology is present in patients and in the *Fmr1*$^{-/-}$ mouse model[24–26].

BPN14770 is a first-in-class PDE4D allosteric inhibitor that is selective for the dimeric, PKA-activated form of the enzyme[32,33]. In adult male *Fmr1*$^{-/-}$ mice, dosing with BPN14770 for 2 weeks ameliorates multiple behavioral phenotypes, for example, hyperactivity and decreased social interaction, as well as benefiting natural behaviors such as nest building and marble burying[16]. Behavioral improvement is associated with structural changes in dendritic spines on pyramidal cells in the prefrontal cortex. Indeed, knock down of PDE4D messenger RNA in wild-type mice shows an even stronger effect on dendritic spine maturation[34]. The behavioral improvement persists for up to 2 weeks after the last dose of the drug or up to 28 times the plasma half-life of BPN14770 in mice ($t_{½}$ = 10–12 h).

In this article, we report the findings from an exploratory, phase 2, randomized, two-period crossover study of BPN14770 in 30 adult male patients with FXS. This study included only male patients because this was expected to reduce variability in clinical severity and drug response and allow for better detection of efficacy in a small exploratory clinical trial. The crossover design facilitated participant recruitment since every participant was exposed to the active drug and this also allowed us to explore the potential for a carryover effect as suggested by the preclinical model. Indeed, a substantial carryover effect was observed, which limited the prespecified statistical analysis to period 1 only.

Results

Study design and BPN14770 dose selection. This was a randomized, double-blind, placebo-controlled, two-period crossover study. The study enrolled 30 male participants aged 18–41 years with FXS due to an *FMR1* full mutation (>200 CGG repeats). Each treatment period in the two-period crossover was 12 weeks in duration with no washout between periods. Participants were randomized to the treatment sequence BPN14770 crossing over to placebo (that is, treatment sequence A-B) or placebo crossing over to BPN14770 (that is, treatment sequence B-A). Participants, parents/caregivers and clinician raters were blinded to treatment. Clinical safety and efficacy end points were assessed at baseline and during the week 6 and week 12 clinic visits within each period. Participants received twice daily oral doses of 25 mg of BPN14770 or matching placebo. The BPN14770 25 mg twice-daily dose was selected to provide sufficient plasma exposure to explore potential efficacy (C_{min} approximately 100 ng ml^{-1}) but remain below the threshold for nausea and emesis established in phase 1 ascending dose studies in healthy individuals (C_{max} approximately 1,500 ng ml^{-1}).

Participants. A total of 30 individuals were screened at a single center with no screen failures; all 30 individuals were randomized and all completed the 2 periods; thus, the analysis populations (safety, intent to treat (ITT) and completers) were all the same, that is, included all 30 individuals. Extended Data Fig. 1 provides the CONSORT diagram of individual participation and completion. The clinical portion of the study was conducted from 9 July 2018 when the first individual was enrolled to 31 July 2020 when the last study visit with the last individual occurred.

Demographic and other baseline characteristics for the study population are summarized in Extended Data Fig. 2. Participants enrolled averaged 31.6 years of age. All were male with the majority white (86.7%) and non-Hispanic/non-Latino (93.3%) with a median body mass index of 27.3 kg m^{-2}. The median Stanford–Binet *z*-deviation full-scale IQ[35] was 42.6 and ranged from 24.6 to 66.2, which indicated that participants were mildly to severely intellectually impaired. At baseline, 36.7% of participants were assessed as mildly ill, 43.3% as moderately ill and 20% as markedly ill based on the Clinical Global Impression-Severity scale (CGI-S). There were no notable demographic or medical history differences between participants randomized to the treatment sequences A-B or B-A. Comorbidities and medical histories were consistent with those expected for men with FXS, but some participants had more remarkable medical histories (for example, scoliosis, cholecystectomy, atrophic kidney, seizures, Gilbert syndrome); none had serious or chronic unmanaged conditions.

Before enrollment and throughout the study, 87% of participants were on concomitant medications typical for this population, with the most common (>10% of participants) being analgesics, antibacterials, antidiarrheals, antihistamines, metformin (6 participants; 3 in the treatment sequence A-B and 3 in the B-A treatment sequence), lipid-modifying agents, mineral supplements, nasal preparations, acamprosate (4 participants, 2 in treatment sequence A-B and 2 in the B-A treatment sequence), psychoanaleptics (14 participants) including citalopram, escitalopram, fluoxetine and methylphenidate hydrochloride, psycholeptics (14 participants) including aripiprazole, buspirone, risperidone and olanzapine (Anatomic Therapeutic Chemical level 2 classification of the World Health Organization Drug Dictionary). There was no systematic difference in the use of medications between the two treatment sequence groups.

Participants demonstrated good compliance with their study medication, with most compliance percentage rates (based on expected versus returned pill counts) showing well over 80% compliance across all visits. Notably, the pharmacokinetic concentrations collected at week 12 within each drug period showed that all participants had plasma BPN14770 concentrations within the range expected, which provided additional assurance of compliance with study medication.

Table 1 | Incidence of TEAEs sorted in descending incidence in the BPN14770 25-mg group (safety population)

Medical Dictionary for Regulatory Activities preferred term	BPN14770 25 mg twice daily, *n* = 30, *n* (%)	Placebo, *n* = 30, *n* (%)
At least one TEAE	11 (36.7)	8 (26.7)
Vomiting	3 (10.0)	2 (6.7)
Upper respiratory tract infection	2 (6.7)	3 (10.0)
Abdominal pain	1 (3.3)	0 (0)
Dyspepsia	1 (3.3)	0 (0)
Infective bursitis	1 (3.3)	0 (0)
Streptococcal pharyngitis	1 (3.3)	0 (0)
Sinusitis	1 (3.3)	0 (0)
ALT increased	1 (3.3)	0 (0)
AST increased	1 (3.3)	0 (0)
Blood alkaline phosphatase increased	1 (3.3)	0 (0)
Rotator cuff syndrome	1 (3.3)	0 (0)
Headache	1 (3.3)	0 (0)
Insomnia	1 (3.3)	0 (0)
Anal incontinence	0 (0)	1 (3)
Diarrhea	0 (0)	2 (6.7)
Tooth infection	0 (0)	1 (3.3)
Bruxism	0 (0)	1 (3.3)
Enuresis	0 (0)	1 (3.3)
Urinary incontinence	0 (0)	1 (3.3)
Hiccups	0 (0)	1 (3.3)
Blister	0 (0)	1 (3.3)

Safety and tolerability of BPN14770. The study met the primary end point of BPN14770 safety and tolerability. All participants enrolled in the study completed the study; there were no discontinuations (Extended Data Fig. 1). The most common treatment-emergent adverse events (TEAEs) were vomiting and upper respiratory tract infection, with no meaningful differences between the treatment arms (Table 1). All other events reported in the BPN14770 arm occurred in only one patient each, with no tolerability issues identified in any patient. There was no TEAE judged by the investigator to be at least possibly related to treatment with BPN14770.

There was one serious adverse event (SAE), severe septic olecranon bursitis, that the investigator and medical monitor assessed as an intercurrent illness unrelated to the study drug. There were no clinically important abnormalities in clinical laboratory values except for the individual experiencing severe septic olecranon bursitis who had increases in alanine aminotransferase (ALT), aspartate aminotransferase (AST) and alkaline phosphatase, which returned to baseline after the resolution of the SAE. There were no overall changes in electrocardiograms, no increases from baseline in QTcF or QTcB interval and no values over 500 ms were observed at any time during the study. There was no change in suicidal ideation or in clinically notable self-injurious behaviors.

Persistence of effect and carryover from period 1 to period 2. Statistical analyses of secondary outcome measures were performed taking into account the crossover design characteristic of the study, that is, including data from both period 1 and period 2. However, according to the a priori statistical analysis plan included in the study protocol, a predefined analysis of period 1 only (the first 12 weeks of the study) was to be performed if a meaningful carryover effect into period 2 was observed. Preclinical data suggested persistence of drug effect may extend well beyond the last dose of the study medication[16]; thus, this preplanned aspect of the analysis was included in the original protocol and statistical analysis plan. Furthermore, since measurable but very low BPN14770 plasma concentrations were evident as long as 12 weeks after the last dose from period 1 to period 2 in most participants, it was concluded that the period 1 only outcomes were the most appropriate results from which to form the basis for conclusions regarding the outcomes of this study.

Cognitive benefit of BPN14770. The study met multiple secondary efficacy measures. Two direct measures of cognitive performance were completed by participants on an iPad or computer, the National Institutes of Health-Toolbox Cognition Battery (NIH-TCB)[36] and the Test of Attentional Performance for Children (KiTAP)[37]. The NIH-TCB is a battery of extensively validated, computer-administered cognitive tests with utility across childhood and adolescence, early adulthood and old age. The NIH-TCB assessments have been adapted for participants with intellectual disability and have been designed to minimize the floor and ceiling effects that are often present in testing batteries designed for the general population[38,39]. The NIH-TCB tests selected for the study were Cognition Crystallized Composite, Picture Vocabulary, Oral Reading Recognition, Pattern Comparison Processing Speed and Picture Sequence Memory. The Cognition Crystallized Composite score combines the Picture Vocabulary and Oral Reading scores and has been scaled to correlate with IQ[39].

Figure 1a and Table 2 provide the summary of the NIH-TCB primary uncorrected standard scores analyses through period 1 only. Higher scores for the NIH-TCB indicate better function. These results revealed the statistically significant benefit of treatment with BPN14770 over placebo in the Cognition Crystalized Composite (least squares (LS) mean difference +5.31, $P = 0.0018$), Picture Vocabulary (+5.81, $P = 0.0342$) and Oral Reading Recognition (+2.81, $P = 0.0157$). Furthermore, while not statistically significant, both Picture Sequence Memory and Pattern Comparison Processing Speed each demonstrated strong numerical trends favoring BPN14770 treatment over placebo.

The KiTAP assessed attention and inhibition, measures of executive function, as the participants explored an enchanted castle presented on a computer. Go/No-Go errors were statistically significantly reduced in the BPN14770 arm versus placebo (LS mean difference −2.78, $P = 0.0425$). Alertness mean reaction time improved (decreased) in the BPN14770 arm versus placebo, although this finding was not statistically significant (Fig. 1 and Table 2).

Efficacy outcomes measured by parent/caregivers. The study also met secondary efficacy measures assessed by parent/caregivers. Visual Analog Scales (VAS) were constructed for three domains: daily functioning; anxiety/irritability; and language, using patient-specific behavioral anchors selected by the parents/caregiver rater. As examples, behavioral anchors selected for daily functioning were 'cleaning up after himself/shaving', 'fewer reminders needed for daily activity' and 'staying on task to leave house with fewer prompts'. For language, examples of behavioral anchors were 'perseverating on certain topics/appropriate conversation', 'flow of conversation- varied topics' and 'longer, clearer sentences'. Each of the VAS outcomes for the period 1 only analysis favored the BPN14770 arm over placebo with statistical significance that was judged to be clinically meaningful for language (LS mean difference +14.04, $P = 0.0051$) and daily functioning (+14.53, $P = 0.0017$) on

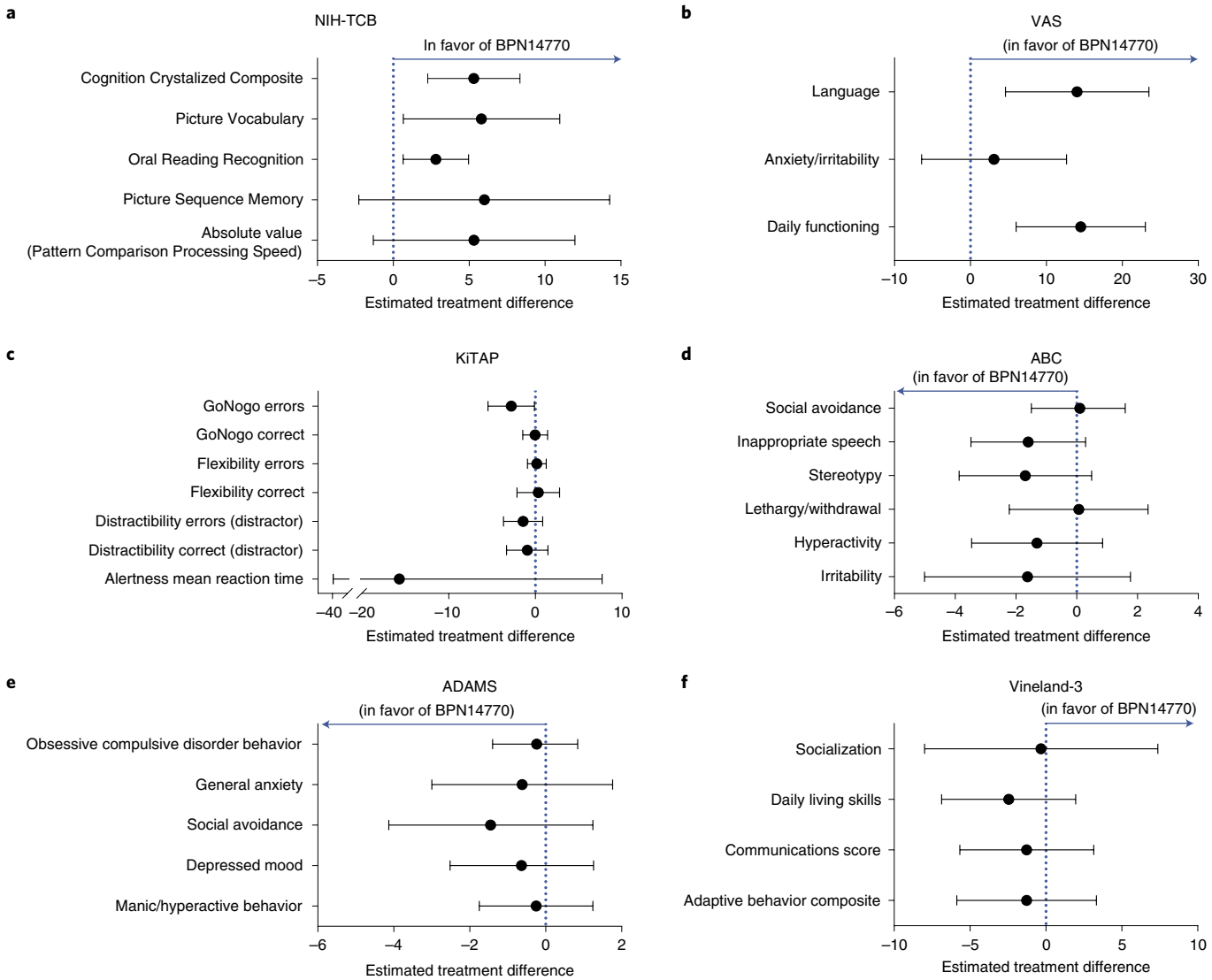


Fig. 1 | Forest plots of clinical outcomes. a–f, Forest plots for the clinical outcomes as assessed using the NIH-TCB (**a**), VAS parent/caregiver rating scales using patient-specific behavioral anchors (**b**), KiTAP (**c**), ABC (**d**), ADAMS (**e**) and Vineland-3 assessment of adaptive behaviors (**f**). Data are presented for period 1 only showing the LS mean difference (±95% CI) between BPN14770 and placebo on change from baseline through week 12 ($n=15$ BPN14770 and 15 placebo participants, except $n=14$ placebo participants for Vineland-3). The LS mean difference and 95% CI were obtained from the MMRM model and reflect the overall treatment effect during period 1 (including data from both weeks 6 and 12) from a model that included treatment, visit, predose baseline covariate and baseline Stanford–Binet full IQ score as fixed effects, with participant as a random effect.

a scale from 0 = worst behavior to 100 = best behavior (Fig. 1b and Table 2).

Other outcome measures scored by parent/caregivers showed numerical trends in favor of BPN14770 over placebo. The Aberrant Behavior Checklist (ABC) showed a numerical trend for improvement in inappropriate speech, stereotypy, hyperactivity and irritability (Fig. 1d), while the Anxiety, Depression, and Mood Scale (ADAMS) showed a numerical trend for improvement in all subscales (Fig. 1e).

Efficacy outcomes measured by physician raters. There was no change in the CGI-S in either treatment sequence. Although not statistically significant, the CGI-Improvement (CGI-I), when dichotomized as 'improved' versus 'not improved', demonstrated a numerically better distribution for improvement for the BPN14770 arm, with 26.7% of BPN14770-treated participants and 6.7% of placebo-treated participants improving from baseline according to clinician/investigator assessment (Table 2). The Vineland-3 adaptive behavior scores were numerically slightly lower on BPN14770 relative to placebo, but none of the changes in Vineland-3 subscale scores were statistically significant (Fig. 1f).

Persistence of effect after last dose of BPN14770. In general, there was a persistence of effect in the crossover from BPN14770 to placebo, while participants crossing over from placebo to BPN14770 showed improvement when exposed to BPN14770. The LS mean change from baseline values through week 24 for the NIH-TCB (Fig. 2a–c) and VAS caregiver rating scales (Fig. 2d–f) is shown by treatment sequence group. Of note was the marked persistence in improvement on the VAS scores through the full 12 weeks of period 2 for those participants in the BPN14770 to placebo treatment sequence group and the notable improvements in the placebo to BPN14770 treatment sequence group during period 2 after participants crossed over from placebo to treatment with BPN14770

Table 2 | Period 1–only domain and overall end point change from baseline through week 12

Assessment	Domain	Placebo *n*	BPN14770 25 mg *n*	Placebo LS mean	BPN14770 25 mg LS mean	Pairwise comparison (BPN14770 minus placebo)			
						LS mean difference	95% CI	*P* versus placebo	Favors BPN14770 (+)
NIH-TCB	Cognition Crystallized Composite	15	15	−3.118	2.172	5.29	2.16, 8.42	**0.0018**	+
	Picture Vocabulary	15	15	−3.212	2.574	5.79	0.47, 11.11	**0.0342**	+
	Oral Reading Recognition	15	15	−1.77	1.025	2.80	0.57, 5.02	**0.015**7	+
	Picture Sequence Memory	15	15	1.404	7.391	5.99	−2.65, 14.63	0.1638	+
	Pattern Comparison Processing Speed	15	15	6.7	1.307	−5.39	−12.24, 1.45	0.1172	+
KiTAP	Go/No-Go errors	15	15	1.26	−1.53	−2.78	−5.47, −0.10	**0.0425**	+
	Go/No-Go correct	15	15	0.56	0.54	−0.02	−1.45, 1.41	0.9756	
	Flexibility errors	15	15	−0.08	0.08	0.17	−0.91, 1.25	0.7543	
	Flexibility correct	15	15	0.01	0.35	0.34	−2.11, 2.79	0.7777	
	Distractibility errors (distractor)	15	15	0.71	−0.71	−1.42	−3.67, 0.84	0.2070	+
	Distractibility correct (distractor)	15	15	0.32	−0.62	−0.93	−3.32, 1.45	0.4281	
	Alertness mean reaction time (ms)	15	15	111.4	−45.1	−156.5	−390.4, 77.29	0.1805	+
VAS	Daily functioning	15	15	−3.51	11.01	14.53	6.01, 23.04	**0.0017**	+
	Anxiety/irritability	15	15	5.96	9.07	3.10	−6.44, 12.65	0.5091	+
	Irritability/language	15	15	3.86	17.91	14.04	4.60, 23.49	**0.0051**	+
ABC	Irritability	15	15	0.91	−0.71	−1.62	−5.01, 1.77	0.3350	+
	Lethargy/withdrawal	15	15	0.8	0.87	0.07	−2.22, 2.35	0.9522	
	Stereotypy	15	15	0.88	−0.81	−1.69	−3.87, 0.49	0.1234	+
	Hyperactivity	15	15	1.19	−0.12	−1.31	−3.46, 0.85	0.2237	+
	Inappropriate speech	15	15	0.56	−1.03	−1.60	−3.48, 0.29	0.0937	+
	Social avoidance	15	15	0.08	0.18	0.10	−1.49, 1.69	0.9003	
ADAMS	Manic/hyperactive disorder	15	15	−1.11	−1.36	−0.25	−1.75, 1.24	0.7288	+
	Depressed mood	15	15	0.15	−0.49	−0.64	−2.53, 1.26	0.4944	+
	Social avoidance	15	15	−0.14	−1.59	−1.45	−4.13, 1.24	0.2786	+
	General anxiety	15	15	−0.79	−1.41	−0.62	−3.00, 1.76	0.5970	+
	Obsessive compulsive disorder behavior	15	15	0.04	−0.24	−0.28	−1.40, 0.84	0.6124	+
Vineland-3	Adaptive behavior composite	14	15	−3.2	−4.48	−1.27	−5.87, 3.33	0.5742	
	Communication standard score	14	15	−1.59	−2.85	−1.26	−5.67, 3.15	0.5619	
	Daily living skills standard score	14	15	−2.01	−4.46	−2.45	−6.87, 1.96	0.2636	
	Socialization standard score	14	15	−5.48	−5.81	−0.33	−8.03, 7.37	0.9305	
CGI	Severity	15	15	−0.06	−0.04	0.02	−0.10, 0.14	0.7271	
	Improvement	15	15	3.91	3.76	−0.15	−0.40, 0.12	0.2238	+

Data shown are the clinical assessment tool, domain, number of individuals for each group, LS mean for each group and pairwise comparison between BPN14770 and placebo. All tests were two-sided with no adjustments for multiplicity. Statistics were estimated based on an ANCOVA model with fixed terms for treatment, baseline domain score and baseline Stanford–Binet full IQ covariates. Data were missing in the Vineland-3 for 1 placebo-treated individual at week 12, period 1. See the relevant sections above for details on the outcomes for each end point. Bold font indicates statistical significance ($P = 0.05$) in favor of BPN14770.

(Fig. 2d–f). By the end of the 24-week study, participants in the BPN14770 to placebo treatment sequence group displayed a nearly 25-point improvement in language from baseline. The period 2 only analysis of BPN14770 versus placebo did not show statistical significance due to the persistence of effect in the period 2 placebo group.

Biomarker assessments. Exploratory biomarkers showed numerical superiority for BPN14770 over placebo but did not reach statistical significance (Fig. 3). For the event-related potential (ERP) analysis, there was a trend for reduced N1 amplitude scores to the first paired stimulus (LS mean difference +0.39 μV; 95% confidence

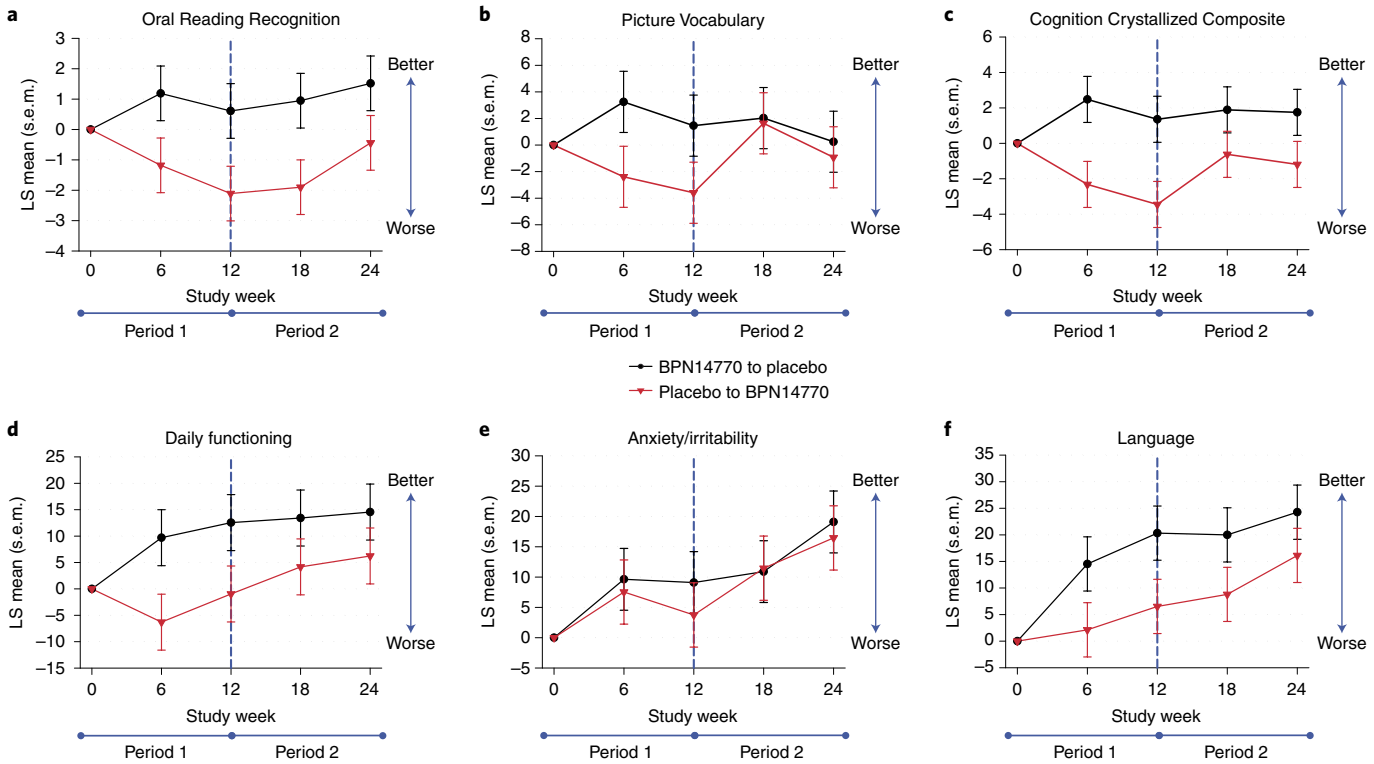


Fig. 2 | LS mean change from baseline values through week 24 for the NIH-TCB and VAS caregiver rating scales. a–c, NIH-TCB by sequence LS mean (±s.e.m.) change from baseline through week 24: Oral Reading Recognition (**a**); Picture Vocabulary (**b**); Cognition Crystallized Composite (**c**). **d–f**, VAS domain by sequence LS mean (±s.e.m.) change from baseline through week 24: daily functioning (**d**); anxiety/irritability (**e**); language (**f**). The persistence of effect for the sequence BPN14770 period 1 to placebo period 2 is shown ($n=15$ BPN14770 and 15 placebo participants). The vertical arrows indicate the direction of change (better or worse). Baseline is the last non-missing value before the first dose of the study medication for period 1, day 1 predose. LS means (±s.e.m.) were obtained from a mixed effects model. The change from baseline was used in the model as the dependent variable. The model included sequence, visit, predose baseline, baseline Stanford–Binet full IQ score and sequence by visit interaction as the fixed effects.

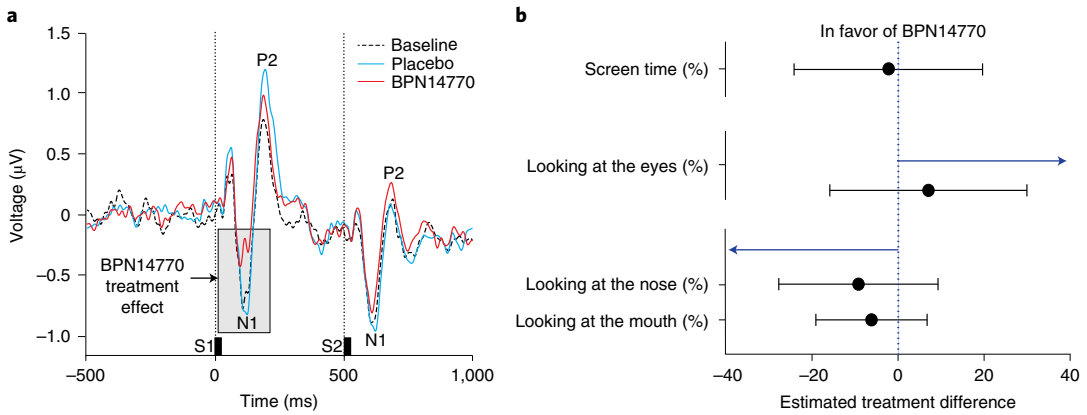


Fig. 3 | Exploratory biomarkers showing the numerical superiority of BPN14770 over placebo. a, ERP assessment of N1 amplitude habituation. Averaged waveforms for baseline, placebo and 25 mg of BPN14770 (ITT population) showed a numerical trend for reducing N1 amplitude to the first paired auditory stimulus (+0.29 μV, two-sided $P=0.0681$, $n=22$ BPN14770 and 23 placebo participants). ANCOVA model with treatment as a fixed term and baseline ERP and baseline IQ as covariates. S1-first paired auditory stimulus, S2-second paired auditory stimulus. The N1 and P2 waves in the ERP are labeled. **b**, Social gaze behavior assessed by eye tracking with faces showed a numerical trend for increased looking at the eyes and decreased looking at the nose and mouth. LS means (±95% CI) were obtained from a mixed effects model ($n=10$ BPN14770 and 7 placebo participants). The ANCOVA model used treatment as a fixed term and baseline looking time and baseline IQ as covariates.

interval (CI) −0.07, 0.85 μV; $P=0.0884$, $n=12$ BPN14770 and 9 placebo participants). Interestingly, the ITT analysis showed the same numerical trend for reduced N1 amplitude (LS mean difference +0.29 μV; 95% CI −0.02–0.60 μV; $P=0.0681$; $n=22$ BPN14770 and 23 placebo participants), which indicated less potential for carryover of effect in the ERP analysis from period

Table 3 | Calculation of effect size based on Cohen's *d*

Parameter	Baseline range	Treatment difference between BPN14770 and placebo	s.d. of differences	*P*	Cohen's *d*
NIH-TCB Cognition Crystallized Composite	28–81	5.3	5.9	0.0009	0.9
VAS daily functioning	1–96	14.5	16.4	0.0009	0.9
VAS language	1–93	14	19	0.004	0.7

P values were estimated from the MMRM with fixed terms for treatment, baseline domain score and baseline Stanford–Binet full IQ covariates. All tests were two-sided with no adjustments for multiplicity.

1 to period 2 in the treatment sequence A-B (BPN14770 to placebo) (Fig. 3a). N1 habituation to the second paired stimulus, P2 amplitude, and P2 latency after a standard or oddball stimulus were unchanged. For eye tracking, an assessment of social gaze behavior[40], there were trends for increased time spent looking at eyes and decreased time spent looking at noses and mouths; however, only 10 BPN14770 and 7 placebo participants completed the eye tracking test at baseline and at the period 1, week 12 visit (Fig. 3b). The differential change from looking at the nose/mouth to time spent looking at the eyes was +14.74%.

Pharmacokinetic assessment. Blood samples drawn at week 12 of each period showed that all study participants had plasma levels of BPN14770 within the expected range ($535 \pm 74\,ng\,ml^{-1}$; mean ± s.e.m.). Pharmacokinetic sampling in the treatment sequence of period 1 BPN14770 to period 2 placebo revealed low ($0.5–18\,ng\,ml^{-1}$) but measurable levels of BPN14770 in the period 2 plasma samples up to 12 weeks after the last dose.

Discussion

There are no approved pharmacological treatments for FXS and thus no demonstrated precedent efficacy end points that have been validated from a clinical and regulatory viewpoint. Thus, the primary outcome measures of this study were safety and tolerability. Prespecified secondary efficacy outcome measures prioritized cognitive performance, followed by subjective caregiver and physician rating scales. Each end point was assessed independently, while overall conclusions are based on a preponderance of evidence and clinical relevance of outcomes. This was an exploratory, hypothesis-generating study, so there was no statistical correction for multiple comparisons. This approach allowed for examination of a wide range of outcomes to assess where, in the multimodal spectrum of symptoms in FXS, BPN14770 might have clinical benefit.

This study met the primary objectives by showing that treatment with 25 mg of BPN14770 twice daily for 12 weeks was well tolerated in a population of males with FXS and aged in their mid-30s. All participants completed the study with no discontinuations. There were no deaths and no TEAEs resulting in discontinuations. One SAE (severe septic olecranon bursitis) occurred during BPN14770 treatment but was judged by the investigator to be unrelated to the study treatment. No other significant TEAEs occurred during this study. Notably, effects on the gastrointestinal tract (nausea, diarrhea and vomiting) were minimal. Marketed PDE4 inhibitors (apremilast and roflumilast) have dose-limiting tolerability due to nausea, vomiting and diarrhea, all of which would be unacceptable in a patient with intellectual disability due to increased burden of care. The TEAEs seen in this study of BPN14770 were vomiting and upper respiratory tract infection; however, there were no meaningful differences between the treatment arms. For example, there were three isolated incidents of vomiting on BPN14770 and two incidents of vomiting on placebo. There was no evidence that participants with intellectual disability and autism spectrum disorder were more sensitive to gastrointestinal disturbance, nor concern that lack of tolerability might allow participants to differentiate active from placebo arms of the crossover study. Given the very good tolerability of BPN14770 in this study, higher doses could be explored in future clinical studies.

Despite the small size of the study, there was a statistically significant improvement in language as assessed directly using the tablet-based assessment tools of the NIH-TCB and as assessed subjectively by the parent/caregiver using a VAS. The improvement in the language and daily functioning domains of the VAS was judged to be clinically meaningful. Moreover, the preponderance of other clinical assessments (KiTAP, ABC, ADAMS and CGI-I) showed a numerical trend for improvement by BPN14770. The improvement in language function and communication would be particularly beneficial in participants with FXS and is an area of special need identified by FXS patient advocacy groups[2,41].

Given the potential mechanism of action of BPN14770 on the rewiring of excitatory cortical connections on dendritic spines, we prioritized cognitive assessments performed by the patients as the key secondary efficacy measure with caregiver and physician rating scales as additional secondary efficacy measures. The study met the key secondary outcome of cognitive benefit as assessed using the NIH-TCB. The Cognition Crystallized Composite score has been scaled to correlate with IQ, which indicates that BPN14770 is acting to improve intellect in patients with intellectual disability in whom the typical IQ range is 40–45. Given that the range of observed Cognition Crystallized Composite score values was 53 (minimum to maximum) at baseline, this reflects a 10% change on the scale range observed which would generally be a meaningful treatment benefit. Given the low *P* values, the effect size as assessed by Cohen's *d* is very high. In his original paper, Cohen suggested that $d = 0.2$ should be considered a small effect size, 0.5 should represent a medium effect size and 0.8 should indicate a large effect size. Table 3 shows that the Cohen's *d* values for the NIH-TCB Cognition Crystallized Composite score and supportive caregiver assessments using a VAS for language and daily functioning ranged from 0.7 to 0.9.

The analysis of secondary efficacy outcomes followed a prespecified statistical analysis plan; however, a limitation is that no correction was made for multiplicity of testing due to the exploratory nature of the trial. Also, the decrement in the placebo group on some of the NIH-TCB outcome measures may have contributed to the findings. Nevertheless, the assumption in a randomized clinical trial is that differences between the active drug and placebo are attributable to the drug. Further studies in larger cohorts are planned and will help to evaluate the potential treatment benefit of BPN14770 in a broader fashion.

There was significant persistence of effect in the treatment sequence period 1 BPN14770 to period 2 placebo. The improvement in language and daily function was maintained or continued to improve 12 weeks after the last dose of BPN14770. The persistence of effect could be due to carryover of BPN14770 due to extended washout or due to more enduring effects on brain function as suggested by the preclinical studies in *Fmr1*$^{-/-}$ mice[16].

A limitation of the study is that no washout was planned between period 1 and period 2. The specific reason for omitting a washout between periods was logistical, in that the double-blind portion of the study was already 24 weeks long and inclusion of the additional time and visits required for a washout was judged to be prohibitive for this target population (for both patients and caregivers). The use

of a crossover design was driven by consideration that in this rare syndrome, enrollment would be limited by patient availability and we were seeking to obtain safety and efficacy information on every possible patient. Finally, we felt that recruitment would be facilitated if all participants had the opportunity to be exposed to the active drug as well as placebo.

The ERP and eye tracking biomarker assessments also favored BPN14770. Abnormally large ERP amplitudes, dissociable from habituation deficits, are an established biomarker in FXS[42]; therefore, the trend toward decreased N1 ERP amplitudes with BPN14770 is also likely a trend for improvement. Eye gaze behavior is altered in FXS consistent with other social behaviors that are altered in ASD. Looking at the eyes when presented with pictures of faces was shown previously to be modulated in patients with FXS exposed to mavoglurant[40]. We also observed a numerical trend for looking at the eyes compared to the nose or mouth in patients treated with BPN14770 compared to placebo, although group size was insufficient for statistical significance as not all patients completed the assessment.

The results of this trial represent an important step toward validating a scientifically based approach to the treatment of FXS. By inhibiting the enzymatic activity of PDE4D, BPN14770 increases the levels of cAMP in the brain and other tissues[43], potentially overcoming the deficits of lower basal and stimulated levels of cAMP thought to be fundamental to this condition[9–11]. More broadly, this mechanism of drug action may be applicable to other neurodevelopmental disorders involving intellectual disability including ACRDYS2 due to activating mutations in PDE4D[44], nonsyndromic intellectual disability due to *CC2D1A* mutation[45], Rubinstein–Taybi syndrome[46], Menke–Hennekam syndrome[47], Phelan–McDermid syndrome[48] and others.

The overall pattern of benefit seen in this randomized, placebo-controlled, exploratory study of BPN14770 in FXS, including improved language function on cognitive tests and parent/caregiver perception of benefit for language and daily function, as well as the directionality of change toward improvement on multiple other clinical measures and biomarkers, is consistent with an effect of BPN14770 on cortical circuitry and addresses a core deficit in patients with FXS. Together with evidence of good tolerability, the preliminary indication of cognitive efficacy justifies expanded clinical trials of BPN14770 in FXS. Although this study was too small in size and scope, as a phase 2 trial, to attempt to delineate responder characteristics, future larger studies could incorporate methodology to assess whether parameters that may underly variability in FXS, such as *FMR1* mosaicism status, FMRP levels and comorbid diagnosis of ASD, have an impact on response to BPN14770 (ref. [49]). Future clinical trials in FXS should also include children to explore benefit during neurodevelopment and women, who comprise a genetically heterogeneous population where both the mutant and wild-type *FMR1* genes are present and expressed.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-021-01321-w.

Received: 24 November 2020; Accepted: 15 March 2021;
Published online: 29 April 2021

References

1. Klaiman, C. et al. Longitudinal profiles of adaptive behavior in fragile X syndrome. *Pediatrics* **134**, 315–324 (2014).
2. Hagerman, R. J. et al. Fragile X syndrome. *Nat. Rev. Dis. Primers* **3**, 17065 (2017).
3. Berry-Kravis, E. M. et al. Drug development for neurodevelopmental disorders: lessons learned from fragile X syndrome. *Nat. Rev. Drug Discov.* **17**, 280–299 (2018).
4. Pieretti, M. et al. Absence of expression of the *FMR-1* gene in fragile X syndrome. *Cell* **66**, 817–822 (1991).
5. Verkerk, A. J. et al. Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).
6. Berry-Kravis, E. M. et al. Effects of STX209 (arbaclofen) on neurobehavioral function in children and adults with fragile X syndrome: a randomized, controlled, phase 2 trial. *Sci. Transl. Med.* **4**, 152ra127 (2012).
7. Berry-Kravis, E. et al. Mavoglurant in fragile X syndrome: results of two randomized, double-blind, placebo-controlled trials. *Sci. Transl. Med.* **8**, 321ra325 (2016).
8. Berry-Kravis, E. et al. Arbaclofen in fragile X syndrome: results of phase 3 trials. *J. Neurodev. Disord.* **9**, 3 (2017).
9. Berry-Kravis, E. & Huttenlocher, P. R. Cyclic AMP metabolism in fragile X syndrome. *Ann. Neurol.* **31**, 22–26 (1992).
10. Berry-Kravis, E., Hicar, M. & Ciurlionis, R. Reduced cyclic AMP production in fragile X syndrome: cytogenetic and molecular correlations. *Pediatr. Res.* **38**, 638–643 (1995).
11. Kelley, D. J. et al. The cyclic AMP cascade is altered in the fragile X nervous system. *PLoS ONE* **2**, e931 (2007).
12. Berry-Kravis, E. & Ciurlionis, R. Overexpression of fragile X gene (*FMR-1*) transcripts increases cAMP production in neural cells. *J. Neurosci. Res.* **51**, 41–48 (1998).
13. Choi, C. H. et al. PDE-4 inhibition rescues aberrant synaptic plasticity in *Drosophila* and mouse models of fragile X syndrome. *J. Neurosci.* **35**, 396–408 (2015).
14. Kanellopoulos, A. K., Semelidou, O., Kotini, A. G., Anezaki, M. & Skoulakis, E. M. C. Learning and memory deficits consequent to reduction of the fragile X mental retardation protein result from metabotropic glutamate receptor-mediated inhibition of cAMP signaling in *Drosophila*. *J. Neurosci.* **32**, 13111–13124 (2012).
15. *Fmr1* knockout mice: a model to study fragile X mental retardation. The Dutch-Belgian Fragile X Consortium. *Cell* **78**, 23–33 (1994).
16. Gurney, M. E., Cogram, P., Deacon, R. M., Rex, C. & Tranfaglia, M. Multiple behavior phenotypes of the fragile-X syndrome mouse model respond to chronic inhibition of phosphodiesterase-4D (PDE4D). *Sci. Rep.* **7**, 14653 (2017).
17. Barad, M., Bourtchouladze, R., Winder, D. G., Golan, H. & Kandel, E. Rolipram, a type IV-specific phosphodiesterase inhibitor, facilitates the establishment of long-lasting long-term potentiation and improves memory. *Proc. Natl Acad. Sci. USA* **95**, 15020–15025 (1998).
18. Byers, D., Davis, R. L. & Kiger, J. A. Jr. Defect in cyclic AMP phosphodiesterase due to the *dunce* mutation of learning in *Drosophila melanogaster*. *Nature* **289**, 79–81 (1981).
19. Levin, L. R. et al. The *Drosophila* learning and memory gene *rutabaga* encodes a $Ca2^{+}$/calmodulin-responsive adenylyl cyclase. *Cell* **68**, 479–489 (1992).
20. Wong, S. T. et al. Calcium-stimulated adenylyl cyclase activity is critical for hippocampus-dependent long-term memory and late phase LTP. *Neuron* **23**, 787–798 (1999).
21. Bourtchuladze, R. et al. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* **79**, 59–68 (1994).
22. Frey, U., Huang, Y. Y. & Kandel, E. R. Effects of cAMP simulate a late stage of LTP in hippocampal CA1 neurons. *Science* **260**, 1661–1664 (1993).
23. Kandel, E. R. The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. *Mol. Brain* **5**, 14 (2012).
24. Comery, T. A. et al. Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. *Proc. Natl Acad. Sci. USA* **94**, 5401–5404 (1997).
25. Irwin, S. A., Galvez, R. & Greenough, W. T. Dendritic spine structural anomalies in fragile-X mental retardation syndrome. *Cereb. Cortex* **10**, 1038–1044 (2000).
26. He, C. X. & Portera-Cailliau, C. The trouble with spines in fragile X syndrome: density, maturity and plasticity. *Neuroscience* **251**, 120–128 (2013).
27. Linglart, A. et al. *PRKAR1A* and *PDE4D* mutations cause acrodysostosis but two distinct syndromes with or without GPCR-signaling hormone resistance. *J. Clin. Endocrinol. Metab.* **97**, E2328–E2338 (2012).
28. Lee, H. et al. Exome sequencing identifies *PDE4D* mutations in acrodysostosis. *Am. J. Hum. Genet.* **90**, 746–751 (2012).
29. Michot, C. et al. Exome sequencing identifies *PDE4D* mutations as another cause of acrodysostosis. *Am. J. Hum. Genet.* **90**, 740–745 (2012).
30. Lynch, D. C. et al. Identification of novel mutations confirms *PDE4D* as a major gene causing acrodysostosis. *Hum. Mutat.* **34**, 97–102 (2013).

31. Wakabayashi, Y. et al. Discovery, radiolabeling, and evaluation of subtype-selective inhibitors for positron emission tomography imaging of brain phosphodiesterase-4D. *ACS Chem. Neurosci.* **11**, 1311–1323 (2020).
32. Burgin, A. B. et al. Design of phosphodiesterase type 4D (PDE4D) allosteric modulators for cognition with improved safety. *Nat. Biotechnol.* **28**, 63–70 (2010).
33. Gurney, M. E. et al. Design and synthesis of selective phosphodiesterase 4D (PDE4D) allosteric inhibitors for the treatment of fragile X syndrome and other brain disorders. *J. Med. Chem.* **62**, 4884–4901 (2019).
34. Baumgartel, K. et al. PDE4D regulates spine plasticity and memory in the retrosplenial cortex. *Sci. Rep.* **8**, 3895 (2018).
35. Sansone, S. M. et al. Improving IQ measurement in intellectual disabilities using true deviation from population norms. *J. Neurodev. Disord.* **6**, 16 (2014).
36. Weintraub, S. et al. Cognition assessment using the NIH Toolbox. *Neurology* **80**, S54–S64 (2013).
37. Knox, A. et al. Feasibility, reliability, and clinical validity of the Test of Attentional Performance for Children (KiTAP) in Fragile X syndrome (FXS). *J. Neurodev. Disord.* **4**, 2 (2012).
38. Shields, R. H. et al. Validation of the NIH Toolbox Cognitive Battery in intellectual disability. *Neurology* **94**, e1229–e1240 (2020).
39. Hessl, D. et al. The NIH Toolbox Cognitive Battery for intellectual disabilities: three preliminary studies and future directions. *J. Neurodev. Disord.* **8**, 35 (2016).
40. Hessl, D. et al. Effects of mavoglurant on visual attention and pupil reactivity while viewing photographs of faces in Fragile X Syndrome. *PLoS ONE* **14**, e0209984 (2019).
41. Riley, C., Mailick, M., Berry-Kravis, E. & Bolen, J. The future of fragile X syndrome: CDC stakeholder meeting summary. *Pediatrics* **139**, S147–S152 (2017).
42. Ethridge, L. E. et al. Auditory EEG biomarkers in Fragile X Syndrome: clinical relevance. *Front. Integr. Neurosci.* **13**, 60 (2019).
43. Zhang, C. et al. Memory enhancing effects of BPN14770, an allosteric inhibitor of phosphodiesterase-4D, in wild-type and humanized mice. *Neuropsychopharmacology* **43**, 2299–2309 (2018).
44. Briet, C. et al. Mutations causing acrodysostosis-2 facilitate activation of phosphodiesterase 4D3. *Hum. Mol. Genet.* **26**, 3883–3894 (2017).
45. Al-Tawashi, A. & Gehring, C. Phosphodiesterase activity is regulated by CC2D1A that is implicated in non-syndromic intellectual disability. *Cell Commun. Signal.* **11**, 47 (2013).
46. Bourtchouladze, R. et al. A mouse model of Rubinstein–Taybi syndrome: defective long-term memory is ameliorated by inhibitors of phosphodiesterase 4. *Proc. Natl Acad. Sci. USA* **100**, 10518–10522 (2003).
47. Menke, L. A. et al. Further delineation of an entity caused by *CREBBP* and *EP300* mutations but not resembling Rubinstein–Taybi syndrome. *Am. J. Med. Genet. A* **176**, 862–876 (2018).
48. Yi, F. et al. Autism-associated SHANK3 haploinsufficiency causes *I*h channelopathy in human neurons. *Science* **352**, aaf2669 (2016).
49. Budimirovic, D. B. et al. A genotype-phenotype study of high-resolution *FMR1* nucleic acid and protein analyses in fragile X patients with neurobehavioral assessments. *Brain Sci.* **10**, 694 (2020).

Methods

Study design and ethical approval. This was a single-center, phase 2, randomized, double-blind, placebo-controlled, two-period crossover study to obtain preliminary assessment of the efficacy and safety of BPN14770 in adult males with FXS (ClinicalTrials.gov identifier: NCT03569631).

The study protocol, participant information and consent form, investigational brochure, available safety information, participant recruitment procedures (for example, advertisements), information about payments and compensation available to participants and documentation evidencing the investigator's qualifications were submitted to the institutional review board at the clinical site, Rush University Medical Center, for ethical review and approval according to local regulations, before the study started. The study was performed in accordance with ethical principles according to the Declaration of Helsinki (Fortaleza, October 2013), seventh revision, 64th World Medical Association General Assembly Meeting and are consistent with the International Conference on Harmonization/good clinical practice, applicable regulatory requirements and the sponsor or its delegate's policy on bioethics. Participant recruitment was managed by the clinical site and was supplemented through the efforts of FXS patient advocacy groups.

Participant eligibility. Eligible participants included men aged 18–45 years inclusive with a diagnosis of FXS and a molecular genetic confirmation of an *FMR1* full mutation (≥200 CGG repetitions). The study enrolled only men since they have the full *FMR1* mutation in all of their cells while women are genetically mosaic due to X chromosome inactivation during early embryonic development and therefore are more diverse in clinical impairment. Participants were allowed current treatment with no more than three prescribed psychotropic medications. Antiepileptic medications were permitted and were not counted as psychotropic medications if they were used for the treatment of seizures. Antiepileptics for other indications, such as the treatment of mood disorders, counted toward the limit of permitted medications. Permitted concomitant psychotropic medications must have been at a stable dose and dosing regimen for at least two weeks before screening and must have remained stable during the period between screening and the commencement of study medication.

Study conduct. The study consisted of a screening period of up to 28 d before initial study drug administration, followed by two 12-week double-blind treatment periods. The screening and baseline visits occurred at the same time, provided the results of safety labs were obtained. No washout period was utilized between double-blind treatment periods; instead, efficacy and biomarker assessments during the second double-blind period (period 2) were obtained after a minimum of six weeks to allow for drug washout. A final follow-up phone contact for safety was planned one week after the conclusion of period 2.

Randomization and masking. Eligible participants were randomized to treatment sequence in a blinded, balanced (1:1) fashion, to receive either 25 mg of BPN14770 twice daily or matching placebo during period 1, followed by the opposite treatment during period 2. Thus, 15 participants were included in each treatment sequence. A computer-generated randomization scheme was prepared before the study started. The randomization was not stratified by any covariate or other baseline characteristic.

Participants and study staff were blinded to treatment assignments. Only the research pharmacy was unblinded to treatment assignments to provide the correct capsules in the blinded bottles to be distributed to participants as a 32-count bottle. The study remained blinded until all participants completed all study procedures and the clinical database was locked.

BPN14770 treatment. BPN14770 25 mg or placebo was administered orally twice daily for 12 weeks. One capsule was to be taken orally in the morning and one capsule was to be taken orally at night, with at least 120 ml (4 ounces) of liquid. Doses should have been taken at least 6 h apart and at least 30 min before or 1 h after meals. BPN14770 or placebo were provided as a hard capsule in an excipient blend. Participants were permitted to open the capsules and sprinkle the contents on food (applesauce) if they were unable to swallow the intact capsule. To monitor compliance, participants were asked to keep a daily record of all doses taken or missed and return all unused study drug at each study visit and at the end of the study.

Primary and secondary efficacy measures. Primary outcome measures included the incidence of TEAEs. Secondary outcome measures included: (1) the NIH-TCB: (a) Cognition Crystallized Composite; (b) Picture Vocabulary; (c) Oral Reading Recognition; (d) Picture Sequence Memory; and (e) Pattern Comparison Processing Speed; (2) KiTAP; (3) VAS using patient-specific behavioral anchors to assess domains of (a) language, (b) anxiety/irritability and (c) daily function; (4) ABC; (5) ADAMS; (6) CGI-S; (7) CGI-I; (8) Vineland-3 assessment of adaptive behavior; (9) ERPs; and (10) assessment of social gaze by eye tracking.

To quantify cognitive ability at baseline, the Stanford–Binet 5 was administered to all participants and scored by the *z*-deviation method published by Sansone et al.[35] to give accurate IQ values not impacted by floor effects expected in individuals with intellectual disability. Cognitive performance throughout the study was assessed using tests that could be performed by the participant on a iPad or computer. The NIH-TCB, KiTAP, VAS, CGI-S and CGI-I were assessed at the week 6 and week 12 study visits of each period. The ABC, ADAMS, Vineland-3, ERP and eye tracking were assessed only at the week 12 study visit at the end of each period.

The NIH-TCB tests were administered on an iPad and adapted for participants with intellectual disability (NIH-TCB v.2.0).[38] The KiTAP is an executive function test administered on a computer, for which multiple subtests have been shown to be feasible and reproducible in FXS[37]. The predefined end points selected for the study were alertness mean reaction time, distractibility correct (distractor), distractibility errors (distractor), flexibility correct, flexibility errors, Go/No-Go correct, and Go/No-Go errors. The caregiver VAS used patient-specific behavioral anchors (0 = worst behavior, 100 = best behavior) and assessed domains of language, anxiety/irritability and daily function. The ABC is a 58-item parent/caregiver rating scale that has been subjected to factor analysis for FXS; in the resulting scoring modification, it was used to assess behaviors across six dimensions or subscales: irritability; hyperactivity; lethargy/withdrawal; stereotypy; inappropriate speech; and social avoidance. The ADAMS is a 28-item behavior-based informant instrument rated by the parent/caregiver and has been used to assess anxiety, depression and mood disorders in individuals with intellectual disability.

Biomarker assessments. Biomarker assessments included measurement of ERPs and eye tracking. Assessment of the ERP measured auditory habituation to trains of repeated stimuli, resting state alpha and gamma power and auditory oddball responses[42]. An electroencephalogram (EEG) was continuously recorded and digitized at 512 Hz, with a 5th order Bessel anti-aliasing filter at 200 Hz, using a 32-channel BioSemi ActiveTwo system (BioSemi). All sensors were referenced to a monopolar reference feedback loop connecting a driven passive sensor and a common-mode-sense active sensor, both located on the posterior scalp. Participants completed three tasks: auditory habituation, which consisted of 150 paired 50-ms white-noise bursts 500 ms apart, with each pair separated by a 4,000 ms interstimulus interval; auditory oddball, which consisted of 432 'standard' tones (1,000 Hz; 90% of stimuli) and 48 'oddball' tones (2,000 Hz; 10% of stimuli); and resting EEG consisting of two sets of 1 min sitting quietly with eyes open followed by 30 s eyes closed. Auditory tasks were presented at 70 dB sound pressure level via headphones. Data were filtered at 0.5–100 Hz, artifacts were removed using independent component analysis and referenced to average reference. For auditory habituation and oddball, N1 ERP amplitudes were calculated as the most negative peak across frontocentral electrodes between 50 and 200 ms poststimulus. For resting EEG, the alpha power was calculated separately for eyes open and eyes closed conditions using a 1-s moving window fast Fourier transform with 50% overlap across occipital electrodes. Gamma power was calculated for eyes open only across all electrodes using the same fast Fourier transform technique.

Assessment of social gaze behavior by eye tracking to faces was in a quiet room with the lights turned off. The eye tracker (Tobii Eye Tracker Technology) was calibrated for each participant at the beginning of each session. After calibration, participants viewed pictures of faces (calm, fearful or happy) shown on the screen as described by Hessl[40]. Each assessment began with the presentation of a scrambled face image for 1 s followed immediately by its matched face image for 3 s. An intertrial interval containing a uniform gray screen was shown for 0.5, 1 or 2 s, determined randomly. The order of face presentation was pseudorandomized and each eye tracking session lasted approximately 6 min. Measurements included looking time to the eye region of interest and number of fixations to the eye, nose or mouth, as well as pupil dilatation by pupillometry.

Safety outcomes. Safety assessments throughout the study included collection of spontaneously reported adverse events, clinical laboratory tests, vital signs, physical and neurological examination, suicidality monitoring using the Columbia-Suicide Severity Rating Scale, concomitant medications and 12-lead electrocardiogram.

Pharmacokinetic assessments. Pharmacokinetic samples were collected at a single time point at screening and at the end of period 1 and period 2 (week 12) to confirm that the study drug was present when expected and to estimate plasma exposure. Participants were asked to record the time of day of the last dose taken before the blood draw. Pharmacokinetic outcomes were the concentration of BPN14770 at the time point sampled.

Sample size calculations. Given the exploratory nature of this study, a definitive sample size calculation was not possible. However, for the sake of providing an evaluation of the potential power to detect a meaningful difference between treatments, for a crossover study having no carryover effects with 80% power, a 5% alpha level and a common s.d. of 3 units, a sample size of 30 evaluable participants (15 per sequence group) will allow detection of an effect size of approximately 2.25 units on the ABC total score outcome measure. If carryover effects or excessive dropouts allow only for a period 1 analysis, a sample size of 15 per group during the first period would allow for detection of an effect size of 3.19 units.

Statistical methods. The statistical analysis plan was defined and documented before the database was locked. The primary analysis method for all efficacy end

points was the assessment of changes from baseline as the dependent variable in a mixed model for repeated measures (MMRM). The model included sequence, period, treatment, visit, predose baseline score of the end point being assessed, baseline Stanford–Binet full IQ score and visit by treatment interaction as fixed effects, with participant within sequence as a random effect. This analysis was planned to be performed for both the ITT and completer populations. Both the week 6 and week 12 data were included in the models where appropriate. Note that some end points were only collected at a single time point, that is, week 12, and thus the repeated measures model essentially becomes an analysis of covariance (ANCOVA) model.

As prespecified, if the crossover analysis demonstrated significant carryover from period 1 into period 2, the analysis would focus on the period 1–only analysis. This was in fact the case. Thus, the focus for overall conclusions as to treatment effect are based primarily on the period 1–only assessments. Furthermore, since there were multiple domains for each measurement tool, forest plots were employed to demonstrate the LS mean differences (and 95% CIs about those differences) between the BPN14770 and placebo arms. All tests were two-sided with no adjustments for multiplicity.

For the period 1 data only, the change from baseline to week 12 was analyzed using an ANCOVA model with fixed terms for treatment and baseline Stanford–Binet full IQ covariates for the ITT population for the primary uncorrected standard scores. Primary inferences were drawn from treatment differences for the LS mean changes from baseline derived from the MMRM models at week 12. As additional supportive information, treatment differences for week 6 were derived using the MMRM model. LS means for each visit and s.e.m. of the LS means are displayed graphically. As stated in the prespecified statistical analysis plan, this was a study to obtain preliminary assessment of the effects of BPN14770 in participants with FXS. The study examined the following hypotheses (for BPN14770 25 mg versus placebo): H_o: there is NO difference between treatment with BPN14770 25 mg and treatment with placebo with respect to changes over 12 weeks in end points associated with FXS; H_a: there IS a difference between treatment with BPN14770 25 mg and treatment with placebo with respect to changes over 12 weeks in end points associated with FXS.

Because this was an initial small study in FXS, the end points were not prioritized from a statistical perspective. Each end point was assessed independently, while overall conclusions are based on a preponderance of evidence and clinical relevance of outcomes. All statistical tests were two-sided. This was an exploratory, hypothesis-generating study, so there was no statistical correction for multiple comparisons; rather, review of the preponderance of evidence was used to determine potential treatment benefit. This approach allowed for examination of a wide range of outcomes to inform as to which specific domains BPN14770 might benefit. Additional post hoc analyses were added (after unblinding) that assessed the period 2–only outcomes, using the week 12 values as 'baseline' from which to assess changes. This analysis was performed in response to the carryover effect that was evident after unblinding the treatment allocations codes.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All requests for raw and analyzed data will be promptly reviewed by the study sponsor, Tetra Therapeutics, and by the clinical site, Rush University Medical Center, to verify if the request is subject to any intellectual property or confidentiality obligations. Patient-related data not included in the paper were generated as part of a clinical trial and may be subject to patient confidentiality as required by the Health Insurance Portability and Accountability Act. Any data that can be shared will be released via a material transfer agreement. Requests for data may be made to Elizabeth_Berry-Kravis@rush.edu or to info@tetratherapeutics.com.

Acknowledgements

We thank the patients and their families for participating in the clinical trial. Direct clinical costs were funded by the FRAXA Research Foundation. Access to and training on the NIH-TCB was obtained in association with work on HD076189 (E.M.B.K.). Tetra Therapeutics provided drug product and funded trial administration and independent data analysis.

Author contributions

E.M.B.-K. designed and conducted the clinical study. M.D.H. led the biostatistical analysis. S.A.R. designed the clinical study and served as medical monitor. L.E.E. and M.A.R. analyzed the EEG data. A.H.O., C.M. and J.F. conducted the clinical study. M.E.G. contributed to the clinical protocol and drafted the manuscript.

Competing interests

E.M.B.-K., L.E.E., M.A.R., A.O., C.M. and J.F. declare no competing interests. M.D.H. and S.D.R. are paid consultants to Tetra Therapeutics. M.E.G. is an employee of Tetra Therapeutics, which is a wholly owned subsidiary of Shionogi & Company that has a financial interest in BPN14770.

Additional information

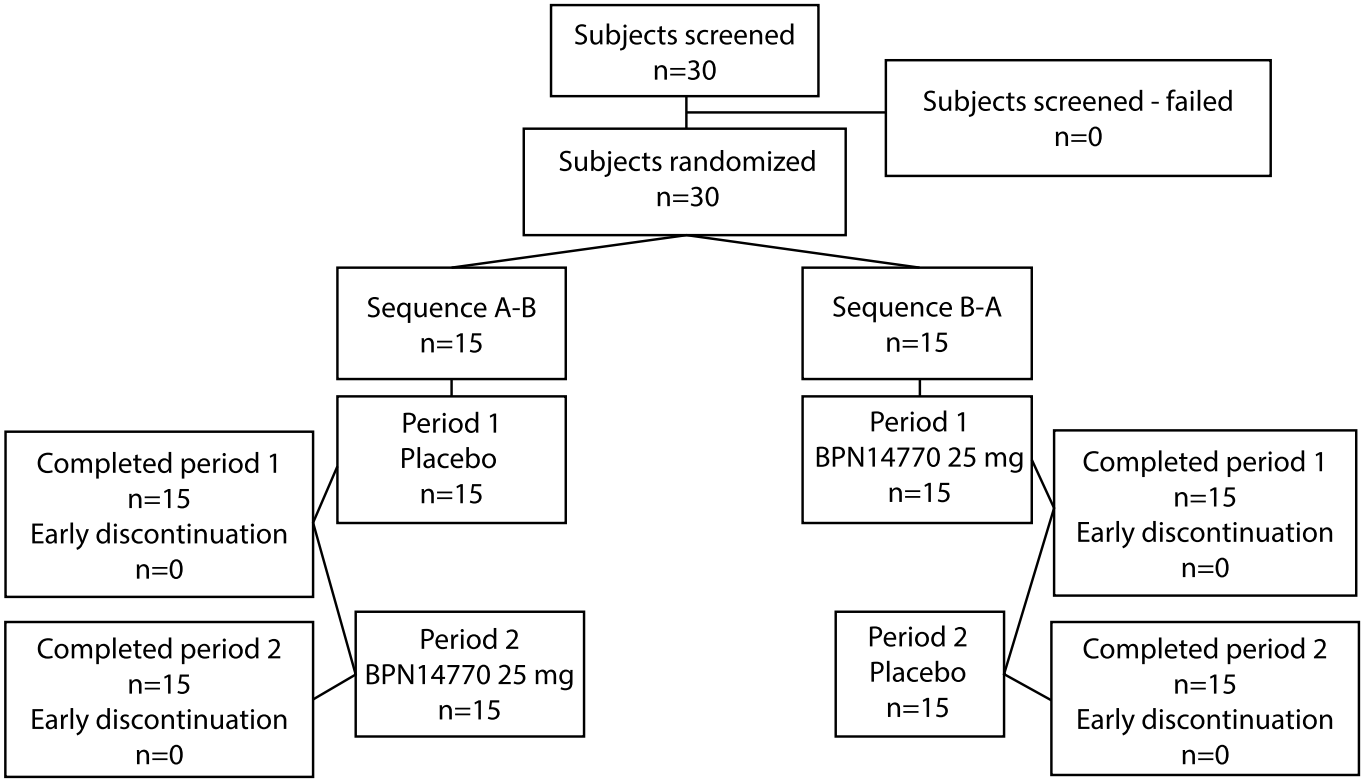
Extended data is available for this paper at https://doi.org/10.1038/s41591-021-01321-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-021-01321-w.

Correspondence and requests for materials should be addressed to E.M.B.-K. or M.E.G.

Reprints and permissions information is available at www.nature.com/reprints.

Peer reviewer information *Nature Medicine* thanks Allan Reiss, Dejan Budimirovic and Blythe Durbin-Johnson for their contribution to the peer review of this work. Jerome Staal was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.



Extended Data Fig. 1 | CONSORT flow diagram for the Phase 2 trial. A total of 30 subjects were screened, all met criteria for entry, 15 were randomized to the A-B (drug-placebo) treatment sequence and 15 were randomized to the B-A (placebo-drug) treatment sequence. There were no discontinuations so the Safety, ITT and Completer populations were identical ($n = 30$ subjects).

Category	Sequence A-B N=15 n (%)	Sequence B-A N=15 n (%)	All Subjects N=30 n (%)
Age (years)			
Mean (SD)	31.2 (6.82)	32.1 (8.00)	31.6 (7.32)
Median	33	35	33.5
Min, max	18 to 39	19 to 41	18 to 41
Gender n (%)			
Male	15 (100.0)	15 (100.0)	30 (100.0)
Female	0	0	0
Race n (%)			
American Indian or Alaska Native	0	1 (6.7)	1 (3.3)
Asian	0	0	0
Black or African American	1 (6.7)	2 (13.3)	3 (10.0)
Native Hawaiian or Other Pacific Islander	0	0	0
White	14 (93.3)	12 (80.0)	26 (86.7)
Other	0	0	0
Ethnicity n (%)			
Hispanic or Latino	1 (6.7)	1 (6.7)	2 (6.7)
Non-Hispanic/Non-Latino	14 (933)	14 (93.3)	28 (93.3)
Height (cm)			
Mean (SD)	174.6 (6.32)	178.7 (6.52)	176.6 (6.64)
Median	176	180	178.3
Min, Max	163 to 185	166 to 187	163 to 187
Weight (kg)			
Mean (SD)	88.3 (12.55)	82.1 (16.10)	85.2 (14.52)
Median	89.9	81.2	87.5
Min, Max	64.7 to 109.2	54.6 to 110.8	54.6 to 110.8
Body mass index (BMI) (kg/m^2)			
Mean (SD)	28.9 (3.15)	25.7 (4.72)	27.3 (4.26)
Median	28.7	26.7	28.0
Min, max	22 to 33	18 to 34	18 to 34
Stanford-Binet Full IQ			
Mean (SD)	40.8 (9.78)	44.7 (12.42)	42.8 (11.16)
Median	41.6	43.5	42.6
Min, max	24.6 to 52.9	25.7 to 66.2	24.6 to 66.2
CGI-S			
Mean (SD)	3.8 (0.68)	3.9 (0.83)	3.8 (0.75)
Median	4	4	4
Min, max	3 to 5	3 to 5	3 to 5

Extended Data Fig. 2 | Subject demographics and baseline characteristics for the treatment Sequence A-B (BPN14770 to Placebo) and treatment Sequence B-A (Placebo to BPN14770). Abbreviations: max=maximum; min=minimum; SD=standard deviation.

nature research

Corresponding author(s): Mark Gurney

Last updated by author(s): 12/4/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Cognitive assessment was with the NIH Toolbox Version 2.0. Clinical data were collected and processed using a validated EDC system. The system and procedures were designed in compliance with Title 21 of the Code of Federal Regulations (21 CFR Part 11).
Data analysis	All analyses were performed using SAS V 9.3 (SAS Institute, Inc, Cary, North Carolina, USA).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All requests for raw and analyzed data are promptly reviewed by the study Sponsor, Tetra Therapeutics, and by the clinical site, Rush University Medical Center, to verify if the request is subject to any intellectual property or confidentiality obligations. Patient-related data not included in the paper were generated as part of a clinical trial and may be subject to patient confidentiality as required by the Health Insurance Portability and Accountability Act. Any data that can be shared will be released via a Material Transfer Agreement. Requests for data may be made to Elizabeth_Berry-Kravis@rush.edu or to info@tetratherapeutics.com.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Given the exploratory nature of this study, a definitive sample size calculation was not possible. However, for the sake of providing an evaluation of the potential power to detect a meaningful difference between treatments, a crossover study having no carryover effects with 80% power, a 5% alpha level, and a common standard deviation of 3 units, a sample size of 30 evaluable subjects (15 per sequence group) will allow detection of an effect size of approximately 2.25 units on the ABC Total Score outcome measure. If carryover effects or excessive dropouts allow only for a Period 1 analysis, a sample size of 15 per group during the first period would allow for detection of an effect size of 3.19 units.
Data exclusions	A total of 30 subjects were screened at a single center with no screen-failures; all 30 subjects were randomized, all completed the 2 periods, and thus, the analysis populations (Safety, ITT, and Completers) were all the same, i.e, included all 30 subjects. Therefore no data has been excluded.
Replication	This is the first clinical trial of a PDE4D allosteric inhibitor in Fragile X syndrome. We plan to conduct future clinical trials to replicate these findings.
Randomization	Eligible subjects were randomized to treatment sequence in a blinded, balanced (1:1) fashion, to receive either BPN14770 25 mg bid or matching placebo during Period 1, followed by the opposite treatment during Period 2. Thus, 15 subjects were included in each treatment sequence. A computer-generated randomization scheme was prepared prior to the study start. The randomization was not stratified by any covariate or other baseline characteristic.
Blinding	Subjects and study staff were blinded to treatment assignments. Only the research pharmacy was unblinded to treatment assignments, in order to provide the correct capsules in the blinded bottles to be distributed to the subjects as a 32 count bottle. The study remained blinded until all subjects completed all study procedures and the clinical database was locked.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Eligible participants included men aged 18 to 45 years, inclusive with a diagnosis of FXS and a molecular genetic confirmation of an FMR1 full mutation (≥200 CGG repetitions).
Recruitment	Subject recruitment was managed by the clinical site and was supplemented through the efforts of the FRAXA Research Foundation. Patients were recruited from the clinic population of EBK and from families that contacted Rush University Medical Center based on interest in the trial after postings on clinicaltrials.gov and the FRAXA or NFXF websites. The main bias would be that we cannot enroll the patients with the most severe behaviors into trials as they cannot cooperate for blood tests or other procedures. There were no other biases and all patients interested and eligible were selected until the

	study was filled.
Ethics oversight	The study protocol, subject information and consent form, Investigational Brochure, available safety information, subject recruitment procedures (e.g., advertisements), information about payments and compensation available to the subjects and documentation evidencing the Investigator's qualifications were submitted to the Institutional Review Board (IRB) at the clinical site, Rush University Medical Center, for ethical review and approval according to local regulations, prior to the study start.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	ClinicalTrials.gov Identifier: NCT03569631
Study protocol	Uploaded with manuscript to Nature Medicine portal
Data collection	The study was conducted at Rush University Medical Center, Chicago, IL, USA. . The clinical portion of the study was conducted from July 9, 2018 when the first subject was enrolled to July 31, 2020 when the last study visit with the last subject occurred.
Outcomes	Primary and secondary outcome measures were prespecified in the Statistical Analysis Plan and registered in ClinicalTrials.gov Identifier: NCT03569631. Primary Outcome Measures : Incidence of Treatment-Emergent Adverse Events [Safety and Tolerability] [Time Frame: up to 24 weeks] Number of subjects with treatment-related adverse events as assessed by MedDRA. Secondary Outcome Measures: NIH-TCB: NIH Toolbox Cognitive Battery for Intellectual Disabilities [Time Frame: Change from Baseline to Week 6 and Week 12 of each Period for each of 7 assessments] A set of 7 cognition assessments administered on an iPad KiTAP Test of Attentional Performance [Time Frame: Change from Baseline to Week 6 and Week 12 of each Period] A set of 4 subtests designed in the form of short games CGI-S: Clinical Global Impression-Severity [Time Frame: Change from Baseline to Week 6 and Week 12 of each Period] The Clinical Global Impression - Severity scale (CGI-S) is a 7-point scale that requires the clinician to rate the severity of the patient's illness at the time of assessment, relative to the clinician's past experience with patients who have the same diagnosis. Considering total clinical experience, a patient is assessed on severity of mental illness at the time of rating 1, normal, not at all ill; 2, borderline mentally ill; 3, mildly ill; 4, moderately ill; 5, markedly ill; 6, severely ill; or 7, extremely ill. CGI-I: Clinical Global Impression-Improvement [Time Frame: Change from Baseline to Week 6 and Week 12 of each Period] The CGI is a clinician-rated scale utilizing history from the caregiver and incorporating it into a clinical rating, first for severity, and then for clinical follow-up. The CGI-S will be used at the baseline assessment to judge symptom severity as 1 = Normal, not at all ill; 2 = Borderline ill; 3 = Mildly ill; 4 = Moderately ill; 5 = Markedly ill; 6 = Severely ill; and 7 = Among the most extremely ill. The CGI-I will be used at the Week 8 and End of Treatment/Week 16 visits to judge the change in clinical impression as 1 = Very Much Improved; 2 = Much Improved; 3 = Minimally Improved; 4 = No Change; 5 = Minimally Worse; 6 = Much Worse; and 7 = Very Much Worse. VAS: Visual Analog Rating Scale [Time Frame: Change from Baseline to Week 6 and Week 12 of each Period] The VAS will be used to measure the severity of three specific behavioral symptoms targeted in this study: behavior problems, language abilities, and eating behavior. For each behavior the caregiver is instructed to mark their impression of the behavior at the baseline visit and again at the Week 8 and End of Treatment/Week 16 visits. The calculated distance in cm between the visit marks thereby demonstrates whether each behavior stayed the same, improved, or worsened during the study and by how much. The scale is from 0 cm (defined as "worst behavior") to 10 cm ("behavior not a problem"). ABC: Aberrant Behavior Checklist [Time Frame: Change from Baseline to Week 12 of each Period] The Aberrant Behavior Checklist (ABC) is a 58-item rating scale used to assess maladaptive behaviors across five original subscales: Irritability (15 items from 0-45), Social Withdrawal (16 items from 0-48), Stereotypy (7 items from 0-21), Hyperactivity (16 items from 0-48), Inappropriate Speech (4 items from 0-12). Additionally, Social Avoidance, a newly developed four-item subscale (from 0-12) of the ABC that captures core social avoidance aspects of Fragile X Syndrome is reported. All items on the ABC are rated from 0 (not at all a problem) to 3 (the problem is severe in degree). Higher scores indicate greater maladaptive behaviors. Differences between Baseline and Week 10 are used as an indicator of change. ADAMS: Anxiety, Depression, and Mood Scale [Time Frame: Change from Baseline to Week 12 of each Period] The ADAMS (Anxiety, Depression, and Mood Scale) is a 28-item behavior-based informant instrument designed to assess anxiety, depression and mood disorders in individuals with intellectual disability. Items are rated on a scale of 0 ("behavior has not occurred, or is not a problem") to 3 ("behavior occurs a lot, or is a severe problem"). The scale is composed of 5 factors which address: Manic/ Hyperactive Behavior, Depressed Mood, Social Avoidance, General Anxiety and Obsessive/Compulsive Behavior. Vineland-3 Rating Scale [Time Frame: Change from Baseline to 12 of each Period] The VABS-III is a caregiver survey interview that measures the personal and social skills of individuals from birth through adulthood. It

was designed to assess handicapped and non-handicapped persons in their personal and social functioning and is appropriate for individuals of all ages. The Adaptive Behavior Composite (ABC) score is calculated from the caregiver responses using age-adjusted scoring tables. ABC scores range from 20 to 160 and indicate low (20-70), moderately low (70-85), adequate (85-115), moderately high (115-130), or high (130-160) overall adaptive functioning.

ERP: Event Related Potentials [Time Frame: Change from Baseline to Week 12 of each Period]
An EEG (electroencephalogram) maps timing of a subject's responses to sounds heard in a headset.

Eye Tracking [Time Frame: Change from Baseline Week 12 of each Period]
Subjects will view pictures shown on the screen. Each assessment begins with presentation of a scrambled face image followed immediately by its matched face image. Measurements include looking time to the eye region of interest (ROI), and number of fixations to the eye ROI, as well as pupil dilatation by pupilometry.

Statistical methods
The Statistical Analysis Plan was defined and documented before the database was locked. The primary analysis method for all efficacy endpoints was the assessment of changes from baseline as the dependent variable in a mixed model for repeated measures (MMRM). The model included sequence, period, treatment, visit, predose baseline score of the endpoint being assessed, baseline Stanford-Binet Full IQ score, and visit by treatment interaction as fixed effects, with subject within sequence as a random effect. This analysis was planned to be performed for both the Intent To Treat (ITT) population and the Completer population. Both Week 6 and Week 12 data were included in the models, where appropriate. Note that some endpoints were only collected at a single timepoint, Week 12, and thus the repeated measures model essentially becomes an ANCOVA model.

As was pre-specified, if the crossover analysis demonstrated significant carryover from Period 1 into Period 2, the analysis would focus on the Period 1-only analysis. This was in fact the case and thus, the focus for overall conclusions as to treatment effect are based primarily on the Period 1-only assessments. Further, as there were multiple domains for each measurement tool, the use of Forest plots has been employed to demonstrate the LSMean differences (and 95% CIs about those differences) between the BPN14770 and placebo arms. All statistical tests were two-sided and no correction was made for multiplicity of testing.

For Period 1 data only, the change-from-baseline to Week 12 was analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment and baseline Stanford-Binet Full IQ covariates for the ITT population for the primary uncorrected standard scores.

Primary inferences were drawn from treatment differences for the Least Squares Mean (LS Means) changes from baseline derived from the MMRM models at Week 12. As additional supportive information, treatment differences for Week 6 were derived using the MMRM model. LSMeans for each visit and the standard errors (SEs) of the LSMeans are displayed graphically.

As stated in the pre-specified statistical analysis plan, this was a study to obtain preliminary assessment of the effects of BPN14770 in subjects with Fragile X Syndrome. The study examined the following hypothesis (for BPN14770 25 mg vs placebo):

- Ho: There is NO difference between treatment with BPN14770 25 mg and treatment with placebo with respect to changes over 12 weeks in endpoints associated with Fragile X syndrome.
- Ha: There IS a difference between treatment with BPN14770 25 mg and treatment with placebo with respect to changes over 12 weeks in endpoints associated with Fragile X syndrome.

Because this was an initial small study in FXS, the endpoints were not prioritized from a statistical perspective. Each endpoint was assessed independently, while overall conclusions are based on a preponderance of evidence and clinical relevance of outcomes. All statistical tests were two-sided. This was an exploratory, hypothesis-generating study, so there was no statistical correction for multiple comparisons; rather, review of the preponderance of evidence was used to determine potential treatment benefit. This approach allowed for examination of a wide range of outcomes to inform as to which specific domains BPN14770 might have benefit.

Additional post-hoc analyses were added (after unblinding) that assessed the Period 2 Only outcomes, using the Week 12 values as 'baseline' from which to assess changes. This analysis was performed in response to the carryover effect that was evident after unblinding the treatment allocations codes.