SUPPLEMENTARY MATERIAL

A comparative evaluation of treatments with 17β-estradiol and its brain-selective prodrug in a double-transgenic mouse model of Alzheimer's disease

Anna E. Tschiffely[a,b,‡], Rosemary A. Schuh[c,d], Katalin Prokai-Tatrai[e,f], Laszlo Prokai[e,f,*] and Mary Ann Ottinger[b,§,*]

[1]Neuroscience and Cognitive Science Graduate Program, University of Maryland College Park, MD 20742, USA
[2]Department of Animal and Avian Sciences, University of Maryland College Park, MD 20742, USA
[3]Research Service, VAMHCS, Baltimore, MD 210201
[4]Department of Neurology, University of Maryland School of Medicine, Baltimore, MD 21201, USA
[5]Center for Neuroscience Discovery, Institute for Healthy Aging, University of North Texas Health Science Center, Fort Worth, TX 76107, USA
[6]AgyPharma LLC, Mansfield, TX 76063, USA

[‡]Current Address:
Naval Medical Research Center
503 Robert Grant Ave
Silver Spring, MD 20910

[§]Current Address:
Department of Biology and Biochemistry
Division of Research
University of Houston
Houston, TX 77204-2015

*Correspondence: Division of Research, University of Houston, Houston, TX 77204-2015, USA (M.A. Ottinger) and the Center for Neuroscience Discovery, Institute for Healthy Aging, University of North Texas Health Science Center, Fort Worth, TX 76107, USA (L. Prokai) maotting@central.uh.edu (M.A. Ottinger) and Laszlo.Prokai@unthsc.edu (L. Prokai). These authors share senior authorship.

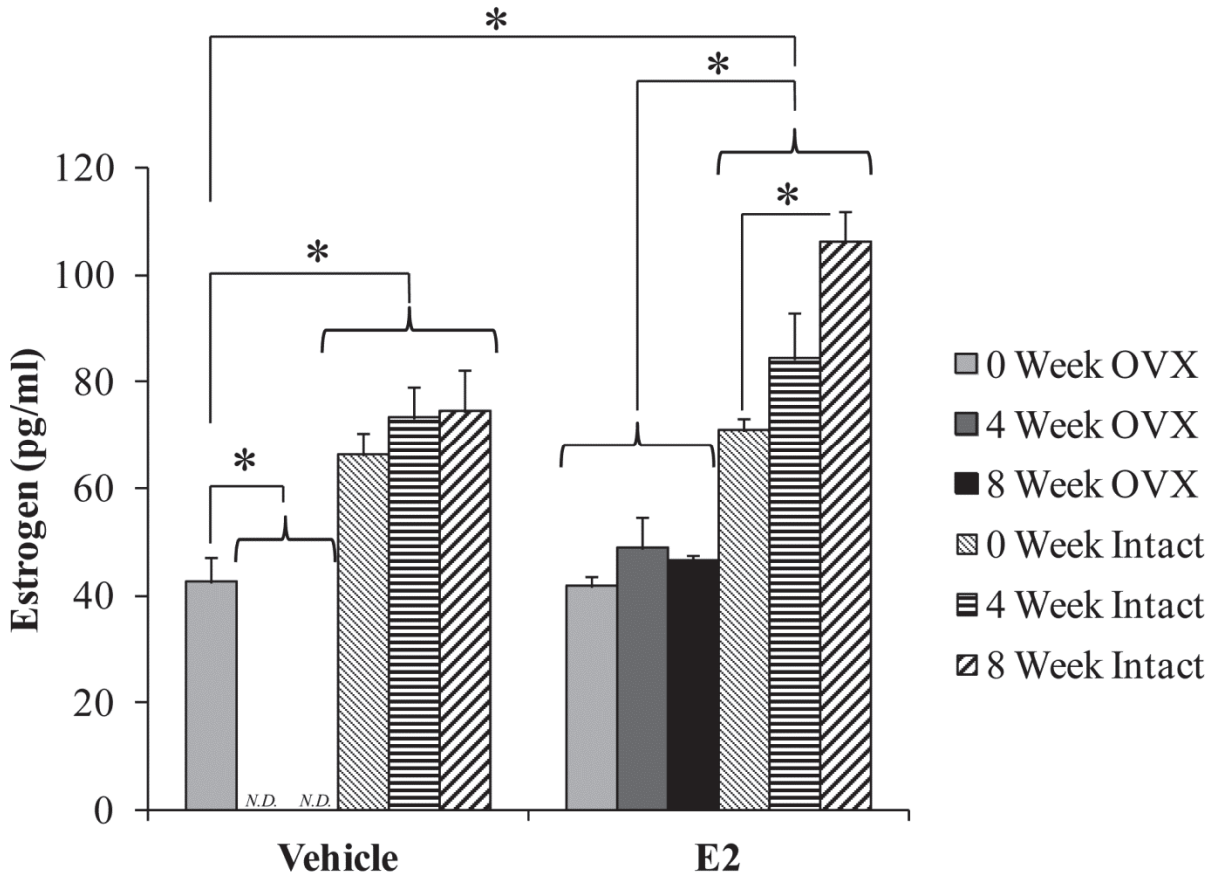
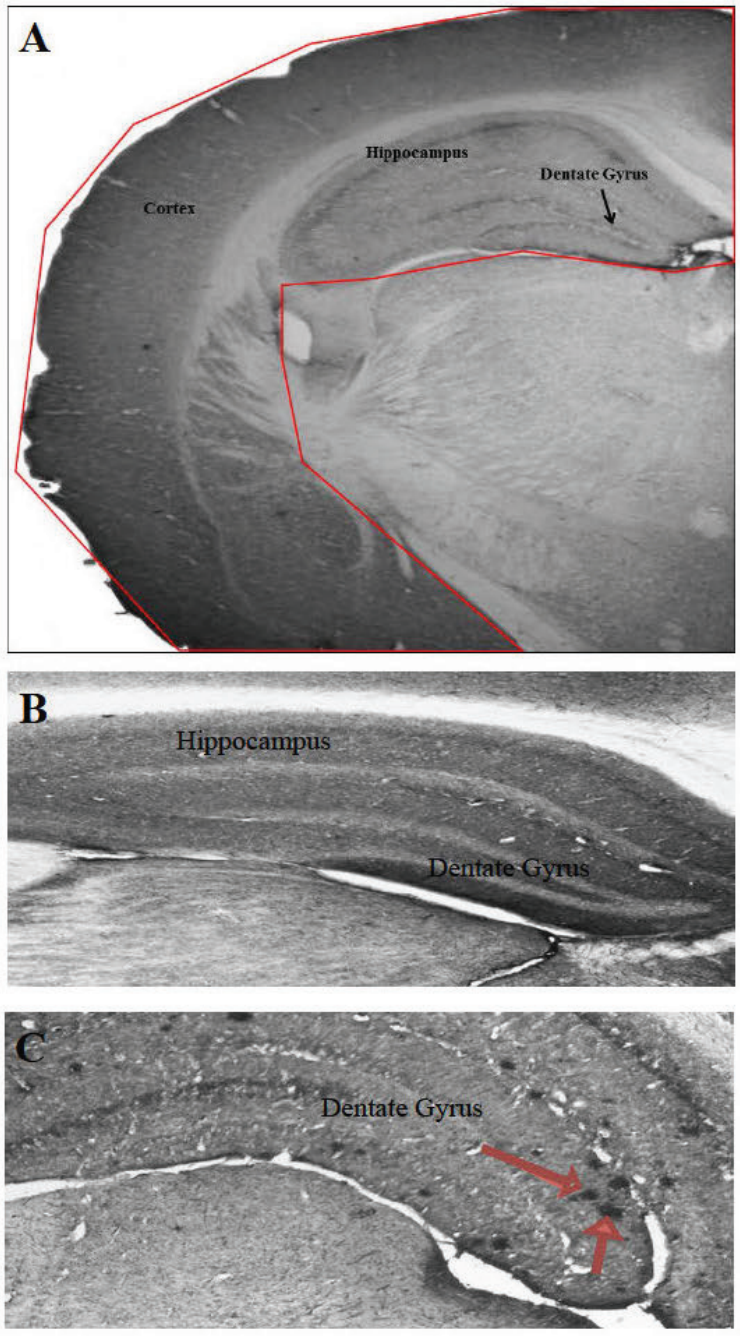


Fig. S1. EIA results show successful ovariectomy and re-establishment of serum estrogen levels in OVX female mice through delivering E2 via Alzet osmotic pumps. In intact animals, there was a statistically significant increase of serum estrogen levels after an 8-week continuous delivery of E2. The employed EIA cross-reacted with estrone (12%) and estriol (0.30%), and we noted significant cross-reactivity with DHED; therefore, assay results were only considered for the vehicle and E2 groups. For OVX mice receiving vehicle, serum estrogen below detectable levels at 4 and 8 weeks after starting the experiments (indicated as *N.D.*, not detected). *$p<0.05$



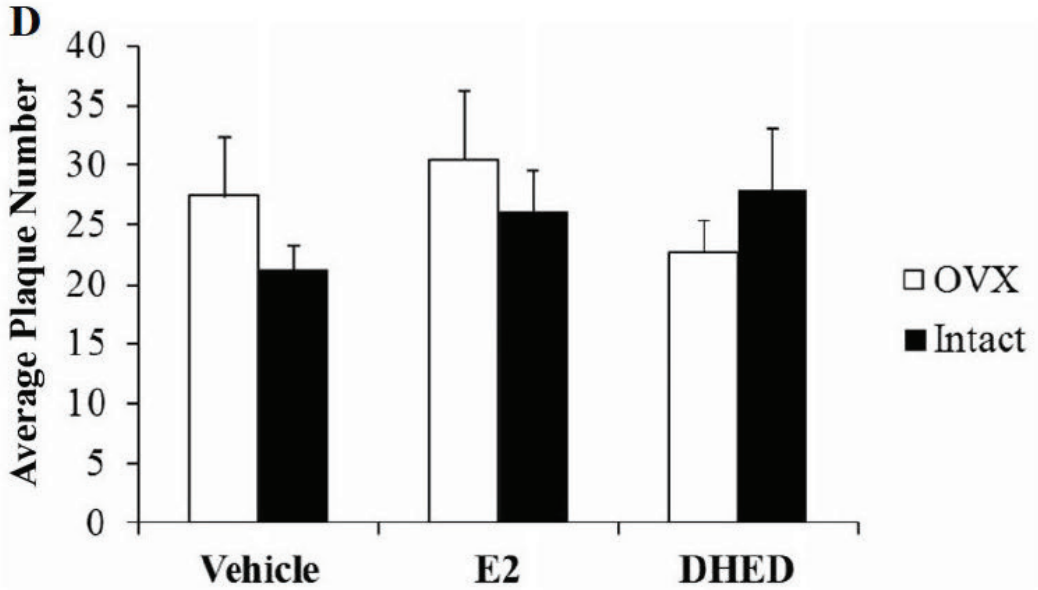
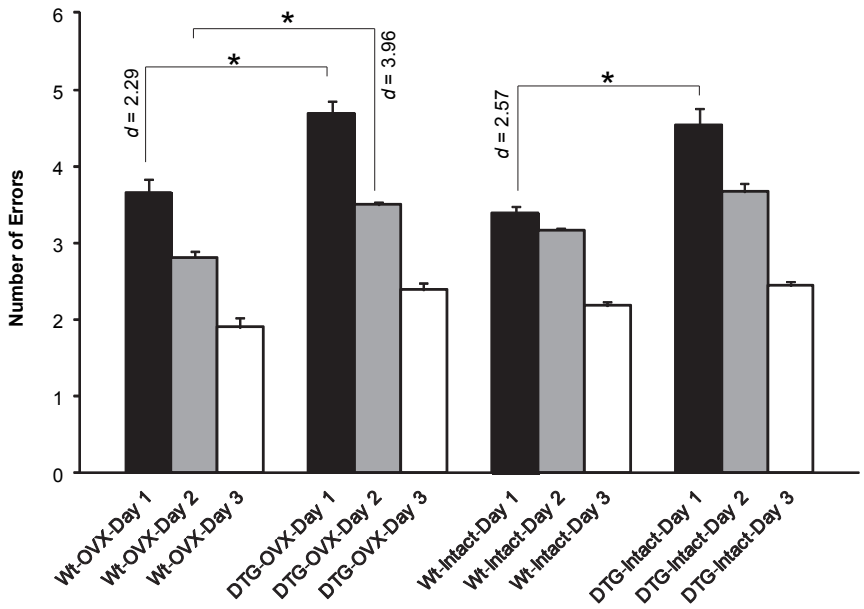


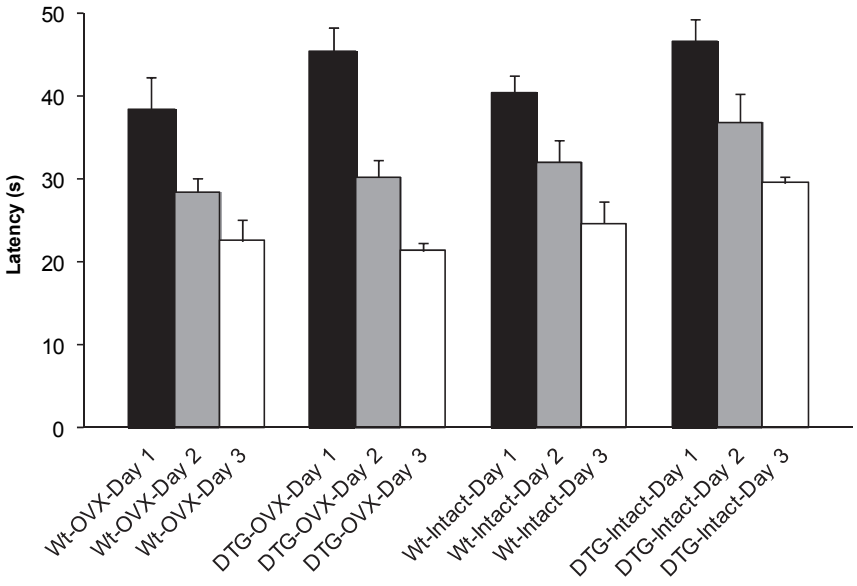
Fig. S2. Semi-quantitative measurements of plaque load in the hippocampus and surrounding cortex in DTG mice. (**A**) The area within the red boundaries was considered for counting plaques; a representative hippocampal/dentate gyrus area of (**B**) non-transgenic and (**C**) DTG female mice with red arrows pointing to plaques; (**D**) average number of plaques (±SEM) counted in the experimental groups involved in our study.

A



Genotype	Genotype	Difference	*p*	95% Confidence Interval	
				Lower	Upper
APPswe/PS1dE9 (DTG)	Wild-type (Wt)	0.686	<0.001	0.550	0.821

B



Genotype	Genotype	Difference (s)	*p*	95% Confidence Interval (s)	
				Lower	Upper
APPswe/PS1dE9 (DTG)	Wild-type (Wt)	3.9	0.008	1.1	6.8

Fig. S3 (preceding page)**.** Statistically significant behavioral deficit upon RAWM testing at 8 moths of age manifested in female APPswe/PS1dE9 double-transgenic (DTG) mice compared to wild-type (Wt) females. (**A**) By counting errors, statistically significant main effect of the APPswe/PS1dE9 transgene introduction ($F_{(1,72)}$=101.6; p<0.001; η_p^2=0.585), as well as genotype x trial ($F_{(2,72)}$=9.44; p<0.001; η_p^2=0.208) and ovarian status x trial ($F_{(2,72)}$=4.52; p=0.014; η_p^2=0.112) interactions were shown by factorial analysis—in addition to detecting the obvious main effect of trials ($F_{(2,72)}$=244.8; p<0.001; η_p^2=0.872). *Post hoc* Tukey tests revealed statistically significant differences (*p<0.05) in the number of errors between both OVX and intact DTG *versus* Wt animals (Effect sizes were shown by Cohen's *d* values). Studies involving 6 treatment groups would reach statistical power of 0.8 (α=0.05; minimum detectable difference of 0.686; standard deviation of residuals: 0.31, estimated from this experiment) with N=7/group. (**B**) Albeit measuring latencies revealed main effects of the AD transgenes, ovarian status and trials without interaction of these factors, *post hoc* Tukey tests did not reveal statistically significant differences with N=7/group. Also, statistical power of 0.8 (α=0.05; minimum detectable difference of 3.9 s; standard deviation of residuals: 6.6 s, estimated from this experiment) would be reached only with a large number of animals (N>70/group). Therefore, we evaluated number of errors made by the DTG animals (N≥7/treatment group) in the RAWM testing during the study reported here.