Published in final edited form as:

Transgenic Res. 2010 October ; 19(5): 829–840. doi:10.1007/s11248-010-9361-1.

Functional evaluation of therapeutic response for a mouse model of medulloblastoma

Aislynn K. Samano,
Greehey Children's Cancer Research Institute, University of Texas Health Science Center, 8403 Floyd Curl Drive, MC7784, San Antonio, TX 78229-3900, USA

Department of Cellular & Structural Biology, University of Texas Health Science Center, San Antonio, TX 78229, USA

Sachiko Ohshima-Hosoyama,
Greehey Children's Cancer Research Institute, University of Texas Health Science Center, 8403 Floyd Curl Drive, MC7784, San Antonio, TX 78229-3900, USA

Thomas G. Whitney,
Department of Mechanical Engineering, University of Texas at San Antonio, San Antonio, TX 78249, USA

Suresh I. Prajapati,
Greehey Children's Cancer Research Institute, University of Texas Health Science Center, 8403 Floyd Curl Drive, MC7784, San Antonio, TX 78229-3900, USA

Aoife Kilcoyne,
Greehey Children's Cancer Research Institute, University of Texas Health Science Center, 8403 Floyd Curl Drive, MC7784, San Antonio, TX 78229-3900, USA

Eri Taniguchi,
Greehey Children's Cancer Research Institute, University of Texas Health Science Center, 8403 Floyd Curl Drive, MC7784, San Antonio, TX 78229-3900, USA

William W. Morgan,
Department of Cellular & Structural Biology, University of Texas Health Science Center, San Antonio, TX 78229, USA

Laura D. Nelon,
Greehey Children's Cancer Research Institute, University of Texas Health Science Center, 8403 Floyd Curl Drive, MC7784, San Antonio, TX 78229-3900, USA

Ai-Ling Lin,
Research Imaging Institute, University of Texas Health Science Center, San Antonio, TX 78229, USA

Osamu Togao,

Correspondence to: Charles Keller, kellerc2@uthscsa.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s11248-010-9361-1) contains supplementary material, which is available to authorized users.

Advanced Medical and Research Center, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

Inkyung Jung,
Department of Biostatistics, University of Texas Health Science Center, San Antonio, TX 78229, USA

Brian P. Rubin,
Department of Anatomic Pathology, Cleveland Clinic, Taussig Cancer Center and the Lerner Research Institute, Cleveland, OH, USA

Brent M. Nowak,
Department of Mechanical Engineering, University of Texas at San Antonio, San Antonio, TX 78249, USA

Timothy Q. Duong, and
Research Imaging Institute, University of Texas Health Science Center, San Antonio, TX 78229, USA

Charles Keller
Greehey Children's Cancer Research Institute, University of Texas Health Science Center, 8403 Floyd Curl Drive, MC7784, San Antonio, TX 78229-3900, USA

Department of Cellular & Structural Biology, University of Texas Health Science Center, San Antonio, TX 78229, USA

Department of Pediatrics, University of Texas Health Science Center, San Antonio, TX 78229, USA

Charles Keller: kellerc2@uthscsa.edu

Abstract

Medulloblastoma is an aggressive childhood cerebellar tumor. We recently reported a mouse model with conditional deletion of *Patched1* gene that recapitulates many characteristics of the human medulloblastoma. Qualitative symptoms observed in the mouse model include irregular stride length, impaired cranial nerve function and decreased motor coordination and performance. In our current study, several quantitative behavioral assays including a mouse rotarod, a forced air challenge, a screen inversion test, a horizontal wire test, and stride length analysis were evaluated to determine the most sensitive and cost-effective functional assay for impaired neuromotor behavior associated with disease progression. Magnetic resonance imaging (MRI) was used to confirm and monitor tumor growth and as an anatomical biomarker for therapeutic response. Wild type mice or medulloblastoma-prone, conditional *Patched1* knockout mice were observed by behavioral assays and MRI from postnatal weeks 3–6. Bortezomib treatment was administered during this period and therapeutic response was assessed using cerebellar volumes at the end of treatment. Of the behavioral tests assessed in this study, stride length analysis was best able to detect differences between tumor-prone mice and wild type mice as early as postnatal day 37 (P = 0.003). Significant differences between stride lengths of bortezomib treated and control tumor-bearing mice could be detected as early as postnatal day 42 (P = 0.020). Cerebellar volumes measured by MRI at the end of treatment validated the therapeutic effects seen by behavioral tests (P = 0.03). These findings suggest that stride length analysis may serve as one of the more

sensitive and cost-effective method for assessing new therapeutic compounds in this and other preclinical model of brain tumors.

Keywords

Medulloblastoma; Genetically-engineered mouse model; Brain tumor; MRI; Cerebellum

Introduction

Medulloblastoma is the most common brain tumor of children between 1 and 10 years of age (Polkinghorn and Tarbell 2007). This cerebellar neoplasm accounts for 30% of all pediatric central nervous system tumors, more frequently affecting children under the age of 5 (Gilbertson 2004). The cerebellum is responsible for the regulation and coordination of complex voluntary muscular movement as well as the maintenance of posture and balance. Unsteady gait presents in 40% of cases as the predominant symptom, especially with lateral hemisphere involvement (Crawford et al. 2007). Children suspected to have medulloblastoma are assessed by x-ray computed tomography (CT) and magnetic resonance imaging (MRI) scans with typical findings of a cerebellar mass; subsequent tumor biopsy and histology are performed to confirm the diagnosis (Hamilton et al. 1995).

Five year overall survival for medulloblastoma without gross leptomeningeal metastasis can be as high as 85% using a multimodality regimen of surgical resection, chemotherapy, and craniospinal radiation (Lafay-Cousin and Strothers 2009; Polkinghorn and Tarbell 2007; St Clair et al. 2004). However, depending upon the age of the child, radiation treatment can result in cognitive impairment, psychiatric disorders, endocrine disorders, and/or growth retardation. Thus, the major challenge is to avoid radiation and replace current treatments with novel therapeutic agents that cross the blood–brain-barrier to target tumor cell-specific signaling pathways while avoiding toxicity to normal tissue.

Early phase drug development is dependent upon the ability to accurately model human disease by research techniques such as in vitro cell culture experiments and in vivo use of animal models (preclinical studies). However, in vitro experiments fall short in reproducing tumor processes such as tumor-host interaction as well as tumor initiation, formation, or progression. Using in vivo mouse models, either via xenograft or genetically engineered mice (GEM), provides a more accurate modeling of human cancer (Fomchenko and Holland 2006). Pharmaceutical companies often use subcutaneous xenograft tumor models because of the reproducibility of tumor formation, rapid tumor growth and high penetrance. However, xenograft tumor models do not faithfully reproduce the complete etiology, pathobiology, and biochemistry of the human disease. On the other hand, GEMs represent a new generation of models that have physiologically-accurate tumor cell microenvironments including dendritic immune cell function and an (initially) intact blood–brain barrier (Fomchenko and Holland 2006; Hann and Balmain 2001; Sakakibara et al. 1996). Therefore, GEMs provide a better preclinical therapeutic platform (Fomchenko and Holland 2006). The use of GEMs also allows for the monitoring of tumor growth in live animals through the use of in vivo imaging technologies such as MRI and bioluminescence, but the former is not

cost-effective and the latter is often useful only when tumors are so large as to split the cranial sutures (Keller laboratory, unpublished data). Thus, although GEMs provide a superior tumor model, academic and pharmaceutical laboratories shy away from these models because cost-effective, quantitative methods to evaluate drug efficacy are not currently available.

Our laboratory has successfully generated a preclinical GEM model of medulloblastoma using a conditional knockout of one of the two *Patched1 (Ptc1)* genes of cerebellar granular neuronal stem cells. Tumor incidence is 100% by 10 weeks of life. Furthermore, bortezomib (Velcade™), a 26S proteasome inhibitor, was found to significantly decrease tumor growth in vitro and in vivo, and to down regulate the sonic hedgehog pathway (Taniguchi et al. 2009). Herein we apply bortezomib treatment in our mouse model to identify a cost-effective behavioral assay for measuring medulloblastoma progression and treatment response.

Methods

Mice and genotyping

Medulloblastoma-prone (MB) mice were generated and genotyped as previously described (Taniguchi et al. 2009). MB mice genotypes were *Ptc1*$^{F1–2m/WT}$ *Trp53*$^{F2–10/F2–10}$ *Pax7*$^{ICNm/WT}$. Wild type (WT) control mice genotypes were *PTC1*$^{WT/WT}$ *Trp53*$^{F2–10/F2–10}$ *Pax7*$^{WT/WT}$. The F2-10 allele for *Trp53* is essentially normal when Pax7-Cre allele (Pax7^{ICNm}) is absent.

Therapeutic administration

Bortezomib (Velcade™, hereafter abbreviated as VLCD) treatment (1 mg/kg) was administered to MB mice or WT mice on postnatal day 21 (P21) by intraperitoneal injection once a week for 3 weeks. Mice were sacrificed on P43 and cerebellums were processed as previously described (Taniguchi et al. 2009). Dimethyl sulfoxide (DMSO) was used to suspend bortezomib and was used as the vehicle control treatment on MB mice or WT mice. The mice were divided into four groups, MB treated (MB-VLCD), MB non-treated (MB-DMSO), WT treated (WT-VLCD), and WT non-treated (WT-DMSO). The number of mice per group varied for each experiment, as described in the results.

Mouse rotarod

Each mouse was tested twice a week from P23 to P43, or until severe ataxia developed. The *TSE Rotarod System Advanced*© mouse rotarod measured the change in motor performance by investigating the mouse's ability to run on an accelerating rod (Hamm et al. 1994). The mouse was placed on a rotating rod that accelerated from 5 rotations per minute (rpm) to 40 rpm over 90 s. The mouse's time on the rod was monitored, 5 runs per test and a mean of the best 3 times was recorded. Mean time less than 90 s resulted in failure of the test. Results were recorded on a pass or fail basis.

Forced air challenge

Each mouse was tested twice a week from P23 to P43, or until severe ataxia developed. A device was built using a Nerf™ rocket launcher and plastic plumber tubing (1/2″ diameter, 1″ elbow tube) to deliver 60 psi of forced air in one-second. Because cranial nerve (CN) VII can be impinged by lateral cerebellar enlargement in the mouse (Supplementary Fig. 1), CN VII function could be examined by the forced air challenge in keeping with manual methods previously described (Arenkiel et al. 2004). Failure of nerve function was detected if the mouse could not close its eyes, lay its ears back, or move its whiskers in response to the forced air. This event was recorded by a single examiner. Results were recorded on a *pass* or *fail* basis.

180° Horizontal screen

Each mouse was tested twice a week from P23 to P43 or until severe ataxia developed. A custom device was built with an invertible 4″ × 4″ mesh screen as previously described (Coughenour et al. 1977). The mouse was placed on top of the screen and then the screen was inverted carefully with a handle putting the mouse upside down on the screen. Failure was recorded if the mouse could not climb back over to the top within 30 s. Results were recorded on a *pass* or *fail* basis.

Horizontal wire

Each mouse was tested twice a week from P23 to P43, or until severe ataxia developed. A horizontal wire device was built using two wooden planks (12″ × 2″ × ½″) with a horizontally strung wire (1 mm diameter) approximately 20 cm from the base. This apparatus was used to investigate motor coordination of the hind limbs. Mice were lifted by the tail base and allowed to grasp the wire with their forepaws, and then released (Facklam et al. 1992). If the mouse could not bring at least one hind leg to the wire within 3 s, failure of coordination was recorded. Results were recorded on a *pass* or *fail* basis.

Weight

Each mouse was tested twice a week from P23 to P43, or until severe ataxia developed. A *Sartorius BL600* digital scale (Sartorius, Gottingen, Germany) was used to record the mouse's weight. Weight was documented in grams for each mouse and a mean for each group was calculated.

Stride length analysis

Each mouse was tested twice a week from P23 to P43, or until severe ataxia developed. Stride length was measured by examining hindlimb stride length on a sheet of paper (Waza et al. 2005). Food coloring was used on the mouse's hind paws and the mouse was allowed to walk on a strip of paper leaving paw imprints on the paper. The sequential stride lengths of left hindleg stride length and right hindleg stride length were combined and a minimum of 10 strides were recorded using the visual machine system (see supplemental methods). A mean of the stride lengths (left paw to left paw and right paw to right paw) per mouse was recorded and a mean per group was calculated.

Magnetic resonance imaging

MRI scans were performed every 7 days from P25 to P40, and on P43. All MRI scans were performed on a 7 Telsa Bruker Biospec horizontal bore scanner (Bruker, Bellerica, MA) with a BG6S gradient insert that can produce maximal gradient amplitude of 100 Gauss/cm (Bruker, Bellerica, MA). A custom-designed 10-mm-inner-diameter circular surface coil was used for radiofrequency transmission and signal reception. Automated shimming was applied over the sensitivity volume of the surface coil. Following the scout scan, multi-slice, contiguous sagittal images were acquired with rapid acquisition and relaxation enhancement (RARE) sequence with the following parameters: repetition time TR = 2571 ms; effective echo time TE = 60 ms; RARE factor = 8; slice thickness = 0.4 mm; field of view (FOV) = 15.6 × 15.6 mm; in-plane resolution = 140 × 128; number of slices = 20; number of averages = 16. Cerebellar volumes were manually calculated for each slice by a neuroradiologist (O.T.). The slice volumes were then combined for a total volume value.

Statistical analysis

The nonparametric results from the mouse rotarod, forced air challenge, horizontal screen, and horizontal wire were analyzed using a Fisher's Exact test to determine if any of the assays had more failing mice than the others. The number of mice that failed each assay was compared between each of the four mice groups for each day. An ANOVA with random subject (mouse) effect test was used to analyze weights. Genders were analyzed separately for weights because a significant difference was found between male and female mice on each day. These tests were performed using *SAS® 9.1.3* software. A two-way ANOVA with a general linear model was used to determine a significant difference, if any, between stride lengths across days and across groups. Original data values were used for the two-way ANOVA. A one-way ANOVA test was used to compare and identify any significant differences in stride lengths between the groups for each day. Unpaired *t*-tests were performed to compare treated and untreated tumor mice on P37 and P42 to confirm a significant difference. A Student Newman-Keuls test was performed to find which group, if any, had a significantly different stride length measurement. Original data values were used for the two-way ANOVA. Means of each mouse's stride length was first calculated, and then means of each group were used for the one-way ANOVA test and *t*-test. There was no significant difference between male and female stride lengths on any given day between any given groups, thus genders were combined due to low sample numbers but were maintained at equal male:female ratios in each group. A linear regression was used to correlate cerebellar volumes with stride lengths. A one-way ANOVA test was used to compare the cerebellar volumes at the end of treatment on P43 between WT-DMSO, MB-VLCD and MB mice. These tests were performed using *Graph Pad Prism®* software and *Minitab®* software. When no significant difference was noted or when there was low sample numbers, a power analysis curve was performed using *Minitab®* software.

Results

Sensitive detection of tumor growth in a preclinical mouse model has continued to be a challenge for drug development laboratories but we hypothesized physiological/behavioral testing may be most cost-effective. Some of the known physiological disruptions caused by

cerebellar tumor growth seen in our mouse model included qualitatively unsteady gait, irregular stride length, poor balance, impaired coordination, and decreased motor performance. In order to test for these functional symptoms of disease progression, a variety of behavioral assays were implemented (refer to Supplementary Table 1 for a list of all mice used and mice genders). The *TSE Rotarod System Advanced*© mouse rotarod (RR) was used to test motor performance (Fig. 1a). The forced air challenge (FAC) apparatus delivered a forced puff of air into a mouse's face in order to test CN VII function (Fig. 1b). The 180° horizontal screen (HS) was employed to test the mouse's coordination of limbs and ability to climb over to the top of the screen (Fig. 1c). The horizontal wire (HzW) apparatus was used to assess the mouse's balance and limb coordination by testing its ability to grasp the wire with the hind paws (Fig. 1d). A digital scale was used to monitor weight of all of the mice (Fig. 1e). Painted hind paws prints (Fig. 1f) were used to measure and record stride lengths for stride length analysis using the visual machine system.

Mouse rotarod, forced air challenge, horizontal screen, and horizontal wire are not reproducible and therefore cannot detect treatment effects

From postnatal day 23 (P23) to postnatal day 42 (P42), the mice were tested twice a week on a pass-fail basis for the mouse rotarod performance, forced air challenge, horizontal screen, and horizontal wire tests. These results were then organized by groups for a Fisher Exact test for each day. Due to small sample numbers, genders were combined. Mouse cohorts were MB-DMSO, MB-VLCD, WT-DMSO and WT-VLCD as described in methods. The number of mice that failed in each group for each test was plotted over time (P23 to P42) into separate graphs. Sample number varied among groups for each day and the numbers have been provided below each behavioral assay graph. For motor performance on the mouse rotarod, there was no significant difference in number of mice that failed on any given day (Fig. 2a). However, the MB-DMSO group had more mice fail on postnatal days 35 and 37 than any other group. On P42, both the MB-DMSO and MB-VLCD mice had an equal number of failures. Similarly, there was no significant difference between the numbers of failing mice for any group using the forced air challenge (Fig. 2b). The MB-DMSO group had one mouse show cranial nerve dysfunction on P37 and two mice on P42. The MB-VLCD group had two mice with cranial nerve dysfunction on P42 as well. More mice failed the horizontal screen test than any other test. However, the results were not consistent and there were more failures among wild type mice compared to tumor-prone mice (Fig. 2c). Although each group had mice fail on different days, there was only a significant difference in the number of mice that failed in the MB-VLCD group compared to all the other groups on P28 ($P \leq 0.003$). The difference found could most likely be attributed to low sample numbers on P28 since 4 of 4 MB-VLCD mice failed. The horizontal wire test also found no significant difference in the numbers of mice failing for each day (Fig. 2d). However, the MB-DMSO mice did have more failures than any other group on P37 and P42. Although the behavioral assays were able to detect symptoms in the MB-DMSO mice, the tests were not able to reproduce consistent symptom detection over time, nor were the tests accurate in detecting symptoms only in tumor prone mice.

Body weights cannot detect treatment effects of bortezomib

Body weight in grams of the groups of mice (MB-DMSO, MB-VLCD, WT-DMSO, and WT-VLCD) were monitored from P23 to P42, twice a week, using a digital scale. Male and female mice were separated due to a significant difference between the gender's weight (P = 0.001) on any given day (Supplementary Figs. 2, 3 and 4). Each gender's weight by group mean was first analyzed using a one-way ANOVA for each day. No significant difference was found between groups for either gender. Genders were then combined for a repeated measures ANOVA with random subject (mouse) test between groups over time and no significant differences were found between weights (P = 0.67). For male mice, the body weights showed no significant difference and power analysis curve was performed (Supplementary Fig. 2). The results demonstrated a low power value possibly due to the low sample number in each group (sample number was reduced when gender was divided). The female mice demonstrated the same results (Supplementary Fig. 3). Weight gain (grams) over time was also analyzed by gender and no significant difference was found.

Stride length analysis differentiates treated and untreated tumor-bearing mice

For the previously described cohorts, the distance between the sequential left hindleg stride length and right hindleg stride length (combined) were measured twice a week from P23 to P42. The means of each group were plotted and the sample number for each day per group has been provided (Fig. 3a). A two-way ANOVA with a general linear model was used to analyze the original stride lengths of each mouse per group across all days. By this analysis, a significant difference was found only between the MB-DMSO mice stride lengths and the WT-VLCD mice on P42 ($P \leq 0.01$). A one-way ANOVA test was then performed separately for each day in order to detect a difference in mean stride lengths between the groups of mice (Fig. 3b). A Student Newman-Keuls test was performed for each day in order to identify the group that differed (Supplementary Table 2). A significant difference in stride lengths was observed between the MB-DMSO stride lengths and all the wild type stride lengths on P37 ($P \leq 0.003$) and P42 ($P \leq 0.002$). The WT-DMSO mice, WT-VLCD mice, and MB-VLCD mice showed no significant difference in stride lengths from each other on any day. However, in the one-way ANOVA test, the MB-VLCD mice showed a significant difference in stride lengths from the MB-DMSO mice on P42 ($P \leq 0.020$). This result was confirmed with an unpaired t-test. Over time, the MB-DMSO mice showed a decrease in stride lengths of the hind paws, while both WT mice groups did not. The MB-VLCD mice showed a slight decrease in stride length distance over time but was not significantly different than the WT mice groups. Large standard deviations among mice in each group is most likely attributed to differences in inter-individual activity levels, low sample numbers and an inability to control the speed when mice were released to run across the paper. In tumor-bearing mice, variability could also be attributed to tumor size and location. A power analysis was performed for each day. As sample size increased over days for each group, the power of the stride length analysis to detect irregularities increased as well. On P37 and P42 (n = 8), the power of the test to detect stride length change was near 1.0. Thus, a sample size of 8 or more mice per group is best-suited to this assay.

Qualitative correlation between increasing cerebellar tumor volumes decreasing stride lengths

WT-DMSO cerebellums (n = 2) and a MB-DMSO cerebellums (n = 2) of live mice were scanned on P25 and P32 (Fig. 4a and b). WT-DMSO cerebellums (n = 2) and MB-DMSO cerebellums (n = 3) were scanned on P32 and P40 (Fig. 4c and d). On P32, the MRI scan revealed thickening of the cerebellar folia layers in the MB-DMSO mouse but no clearly defined tumor mass. On P40, the MRI scan showed the presence of a tumor mass. Cerebellar volumes were calculated from the MRI scans for P25, P32, and P40. A difference between cerebellar volumes of WT-DMSO mice and MB-DMSO mice was detected on P40. MB-DMSO mice had higher cerebellar volumes than the cerebellums of WT-DMSO mice (Fig. 4e). As previous results showed, stride lengths of MB-DMSO mice were significantly different than those of the WT-DMSO mice (P = 0.005) on P37. This symptom of a decreased stride length appeared prior to tumor mass development on P40 as indicated by cerebellar volumes calculated from the MRI scans. Even though the sample number is small for each group, a correlation analysis shows a trend that as cerebellar volumes increased beyond 100 mm^3, stride lengths decreased (Fig. 4f).

End of study cerebellar volumes on postnatal day 43 reflect decreasing cerebellar tumor volume with bortezomib treatment

WT-DMSO mice, MB-DMSO mice, and MB-VLCD mice were scanned on postnatal day 43 to compare cerebellar volumes at the end of treatment. WT-DMSO mice (n = 2) showed no tumor mass (Fig. 5a). MB-VLCD mice (n = 4, 2 not shown) showed enlarged cerebellums, with slight thickening of cerebellar folia (Fig. 5c, yellow arrowheads). MB-DMSO (n = 4, 2 not shown) showed enlarged cerebellums with overt tumor masses (Fig. 5b, yellow asterisks). A one-way ANOVA was performed to detect any difference in cerebellar volumes between the three groups of mice (Fig. 5d). The results indicated a significant difference in volumes between the three groups ($P \leq 0.03$). A Student Newman-Keuls test was then performed and indicated that the MB-DMSO mice had significantly different cerebellar volumes than the MB-VLCD mice and the WT-DMSO mice but there was no significant difference between the MB-VLCD mice and the WT-DMSO mice. A power analysis (P = 1.0) was performed to confirm the test's ability to detect a significant difference (Fig. 5e). Therefore, tumor volumes on P43 were able detect tumor growth in mice cerebellums, and a strong therapeutic effect between treated and non-treated MB mice was observed.

Discussion

In our previous studies we examined the preclinical efficacy of bortezomib by treating mice with fulminant medulloblastoma tumors, and we established that bortezomib extended survival in these mice (Taniguchi et al. 2009). A practical limitation of that study design was that mice were often nearly too moribund to tolerate therapy. For the current study, we have changed our approach to more closely match a time to progression test of therapeutic efficacy. Therapy begins at a point in the animal's life after which tumors are known to have already initiated (P21), and the goal is to assess functionally and/or anatomically whether progression is inhibited by therapy over a period of 3 weeks. In order to functionally assess

and measure neurological deficits associated with disease progression in medulloblastoma-prone mice, a variety of cost-effective and reproducible behavioral assays such as the Mouse Rotarod, Forced Air Challenge, Horizontal screen and Horizontal Wire were implemented. We qualitatively examined whether the results of the behavioral tests coincide with an increase in cerebellar volumes and MRI detection of a tumor mass, and we determined if the time of symptom detection corresponded with cerebellar tumor volumes at the end of treatment. The results suggest that the behavioral assays, though inexpensive, were not effective in reproducibly detecting onset of medulloblastoma tumor growth with the cohort sizes used in this study. Many of the behavioral assays were able to detect symptoms but with little reproducibility or accuracy for the cohort sizes employed here. The mouse rotarod was able to detect decreased motor performance and the forced air challenge was able to detect cranial nerve VII dysfunction in cancer prone mice but not until P42. The horizontal screen and the horizontal wire were not able to distinguish between wild type mice and cancer prone mice. Due to small sample numbers, a change in body weight showed no significant difference between cancer prone mice and wild type mice as demonstrated by the power analysis. However, we believe gender could also be an important confounding factor in assessing tumor progression using weights.

Despite negative results for the above-mentioned functional assays, stride length analysis was in fact effective in identifying a change in stride lengths associated with tumor growth as well as in assessing therapeutic effect. Stride length analysis was able to detect a decrease in stride length at P37 prior to MRI detection (P40). We observed a trend that stride lengths decrease as cerebellar volumes increase beyond 100 mm^3. With the cohort sizes used, gender did not have an effect on stride lengths, but a judicious approach might be to use equal male:female ratios between groups, which is the approach we took for this study. Using the means per group for a one-way ANOVA, stride length analysis was able to detect differences between cancer-prone mice and wild type mice as early as P37 with reproducibility for subsequent days. Analyzing stride lengths was effective and inexpensive. We also noted that the stride length analysis changes on P37 occurred prior to uniform tumor mass presence on MRI scans at P40. The power analysis, however, showed that sample number had a significant effect on the power of the test to detect a significant difference. Fortunately, in our studies, stride length analysis was effective in detecting tumor growth ($P = 0.03$) as well as in monitoring therapeutic effect ($P = 0.020$) with the sample sizes used here. These promising results thus affirm the utility of stride length for assessing preclinical therapeutic efficacy.

By a more conventional method of anatomical assessment, MRI scans at the end of treatment revealed a significant therapeutic effect on cerebellar volumes. Our results of image segmentation show that measuring total cerebellar volume (rather than tumor burden) is a valid surrogate for tumor volume in GEMs, particularly when tumors are multi-focal or the external granular layer is uniformly thickened. By our measures of segmented cerebellar volumes, WT mice and MB treated mice showed no significant difference in cerebellar volumes at the end of treatment but differed considerably from the MB non-treated cerebellar volumes. Thus, analyzing cerebellar volumes using MRI segmentation was a very

accurate technique to determine treatment efficacy, although not a functional evaluation method.

Anatomical imaging by MRI was also helpful to define in part the cause(s) for performance variability and lack of sensitivity of many of the behavioral tests. The advantage of transgenic models of cancer is that the tumors arise in situ and that immunocompetence is preserved (as is the initial blood brain barrier), but the acknowledged limitation in the field is that exact timing of onset and the location in which the tumor arises can vary slightly. From our experience with this study and a preceding pilot study, we have appreciated by MRI that the tumor size and location can vary from mouse to mouse at the earliest stage of onset. Fortunately, less variability in tumor burden was present in treated mice at day 42, which is the study endpoint we would recommend with this genetically engineered model.

In the future, a more sensitive measuring system and comprehensive analysis of mice dynamic gait (by which one could maintain the mice at a constant speed and a more consistent gait) might improve functional detection of neurological deficit over the simplified metric of stride lengths alone. Gait linearity, gait width, and heel-toe pressure are parameters to explore for tumor detection prior to P37. Additional methods testing sensory perception (*i.e.*, balance in response to visual stimulus) could also be explored. The results of this study are thus very encouraging for the use of mouse stride length or gait analysis as a functional adjunct or primary outcome measure in preclinical mouse models of cerebellar tumors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by The National Brain Tumor Foundation and the St. Baldrick's Foundation. We thank Kristine S. Vogel, Wouter Koek, Benjamin R. Arenkiel, and Linda M. McManus for ideas, advice and guidance on these studies.

References

Arenkiel BR, Tvrdik P, Gaufo GO, Capecchi MR. Hoxb1 functions in both motoneurons and in tissues of the periphery to establish and maintain the proper neuronal circuitry. Genes Dev. 2004; 18:1539–1552. [PubMed: 15198977]

Coughenour LL, McLean JR, Parker RB. A new device for the rapid measurement of impaired motor function in mice. Pharmacol Biochem Behav. 1977; 6:351–353. [PubMed: 857257]

Crawford JR, MacDonald TJ, Packer RJ. Medulloblastoma in childhood: new biological advances. Lancet Neurol. 2007; 6:1073–1085. [PubMed: 18031705]

Facklam M, Schoch P, Bonetti EP, Jenck F, Martin JR, Moreau JL, et al. Relationship between benzodiazepine receptor occupancy and functional effects in vivo of four ligands of differing intrinsic efficacies. J Pharmacol Exp Ther. 1992; 261:1113–1121. [PubMed: 1318371]

Fomchenko EI, Holland EC. Mouse models of brain tumors and their applications in preclinical trials. Clin Cancer Res. 2006; 12:5288–5297. [PubMed: 17000661]

Gilbertson RJ. Medulloblastoma: signaling a change in treatment. Lancet Oncol. 2004; 5:209–218. [PubMed: 15050952]

Hamilton SR, Liu B, Parsons RE, Papadopoulos N, Jen J, Powell SM, et al. The molecular basis of Turcot's syndrome. N Engl J Med. 1995; 332:839–847. [PubMed: 7661930]

Hamm RJ, Pike BR, O'Dell DM, Lyeth BG, Jenkins LW. The Rotarod test: an evaluation of its effectiveness in assessing motor deficits following traumatic brain injury. J Neurotrauma. 1994; 11:187–196. [PubMed: 7932797]

Hann B, Balmain A. Building 'validated' mouse models of human cancer. Curr Opin Cell Biol. 2001; 13:778–784. [PubMed: 11698196]

Lafay-Cousin L, Strother D. Current treatment approaches for infants with malignant central nervous system tumors. Oncologist. 2009; 14(4):433–444. [PubMed: 19342475]

Polkinghorn WR, Tarbell NJ. Medulloblastoma: tumorigenesis, current clinical paradigm, and efforts to improve risk stratification. Nat Clin Pract Oncol. 2007; 4:295–304. [PubMed: 17464337]

Sakakibara T, Xu Y, Bumpers HL, Chen FA, Bankert RB, Arredondo MA, et al. Growth and metastasis of surgical specimens of human breast carcinomas in SCID mice. Cancer J Sci Am. 1996; 2:291–300. [PubMed: 9166547]

St Clair WH, Adams JA, Bues M, Fullerton BC, La Shell S, Kooy HM, et al. Advantage of protons compared to conventional X-ray or IMRT in the treatment of a pediatric patient with medulloblastoma. Int J Radiat Oncol Biol Phys. 2004; 58:727–734. [PubMed: 14967427]

Taniguchi E, Cho MJ, Arenkiel BR, Hansen MS, Rivera OJ, McCleish AT, et al. Bortezomib reverses a posttranslational mechanism of tumor genesis for patched1 haploinsufficiency in medulloblastoma. Pediatr Blood Cancer. 2009; 52(2):136–144. [PubMed: 19213072]

Waza M, Adachi H, Katsuno M, Minamiyama M, Sang C, Tanaka F, et al. 17-AAG, an Hsp90 inhibitor, ameliorates polyglutamine-mediated motor neuron degeneration. Nat Med. 2005; 11:1088–1095. [PubMed: 16155577]

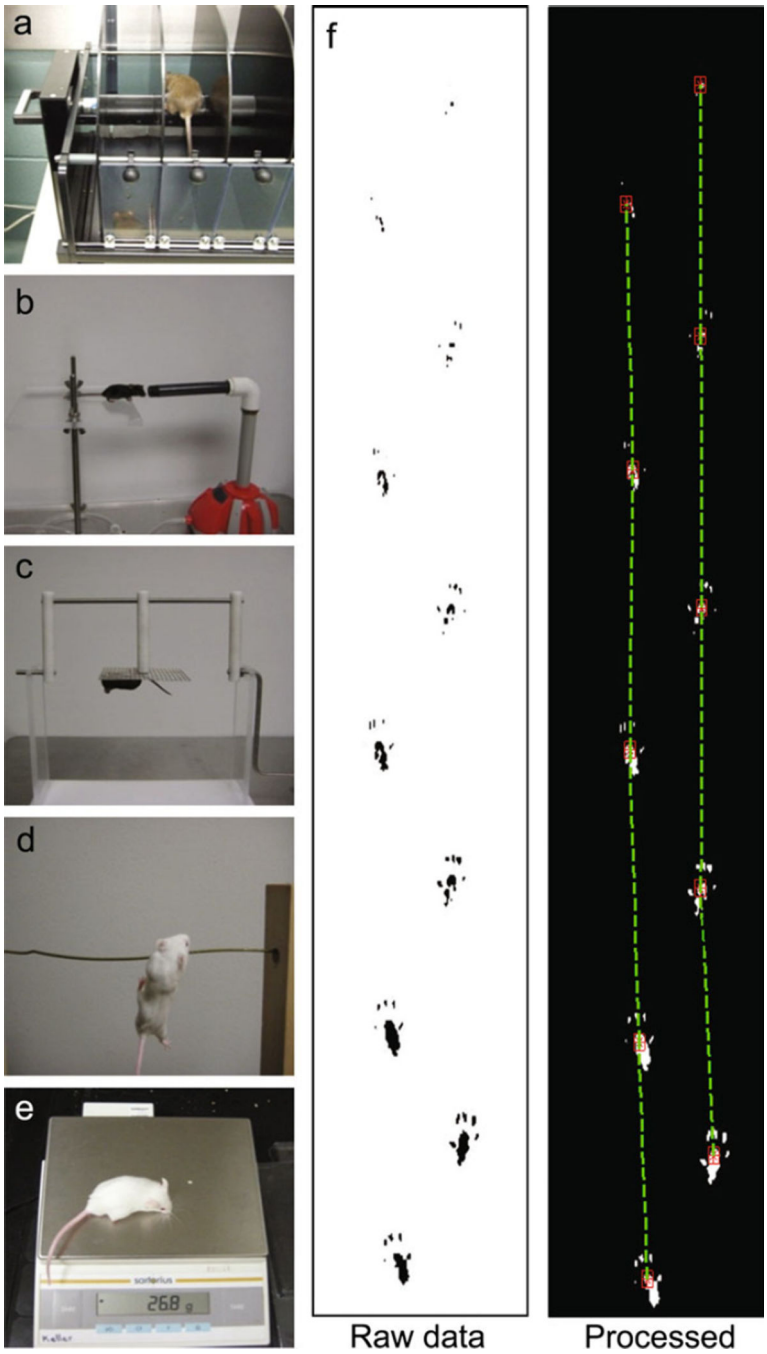

Fig. 1.

Behavioral assays used to evaluate cranial nerve function, motor coordination, motor performance, weight, and stride length. **a** *TSE Rotarod System Advanced*© mouse rotarod, **b** forced air challenge, **c** horizontal screen, **d** horizontal wire, **e** digital scale, **f** hind stride length raw data (*left*) and processed data using visual machine system (*right*)

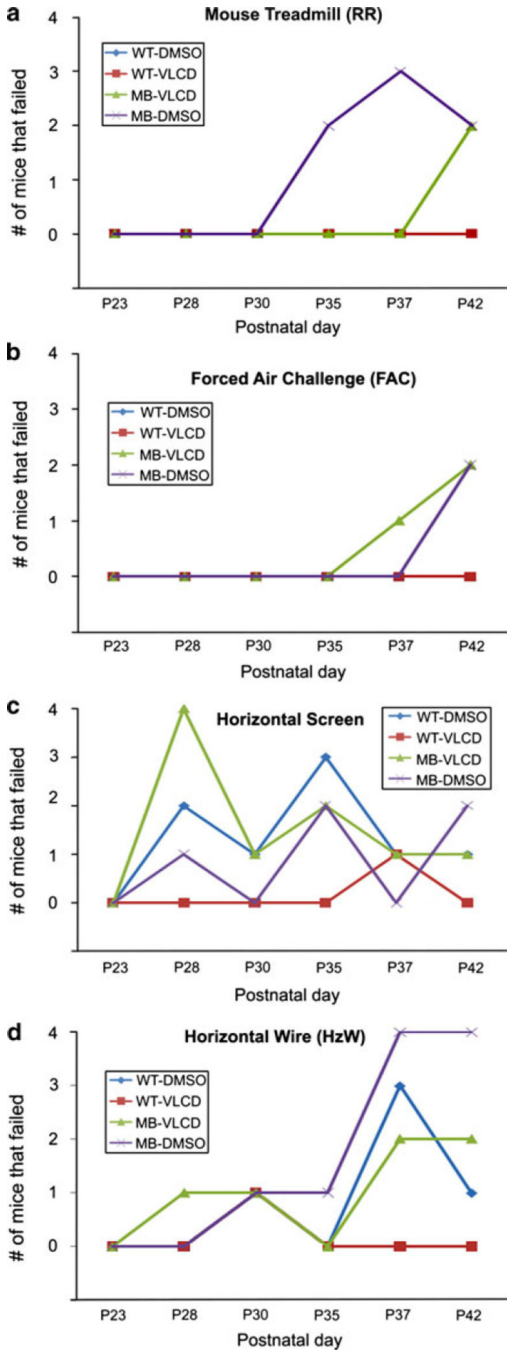

Fig. 2.

Experimental test results. **a** Rotarod results, graph shows number of mice in each group that failed the mouse rotarod test per day. MB-DMSO had more failing on days 35 and 37 but had the same number of MB-VLCD mice failing on day 37. A Fishers Exact test was used to identify any significant differences between numbers of failing mice per group. **b** Forced Air Challenge Results, graph shows number of mice in each group that failed the forced air challenge per day; MB-VLCD had more failing on day 37 but not on day 42. A Fishers Exact test was used to identify any significant differences between numbers of failing mice

per group. **c** Horizontal screen Results, graph shows number of mice in each group that failed the forced air challenge per day. On P28, all groups except WT-VLCD had at least 1 failing mouse. MB-VLCD had significantly more mice (*P = 0.003) fail than any other group however; the trend was not consistent for subsequent days. For postnatal days 30, 35, 37, and 42 all groups continue to have some failing mice, including the wild type mice, but no group had significantly more failings after P28. A Fishers Exact test was used to identify any significant differences between numbers of failing mice per group. **d** Horizontal Wire Results. Graph shows number of mice in each group that failed the forced air challenge per day. All groups had failing mice starting on P28. A Fishers Exact test was used to identify any significant differences between numbers of failing mice per group. No group had significantly more failing mice than any other group on any given day

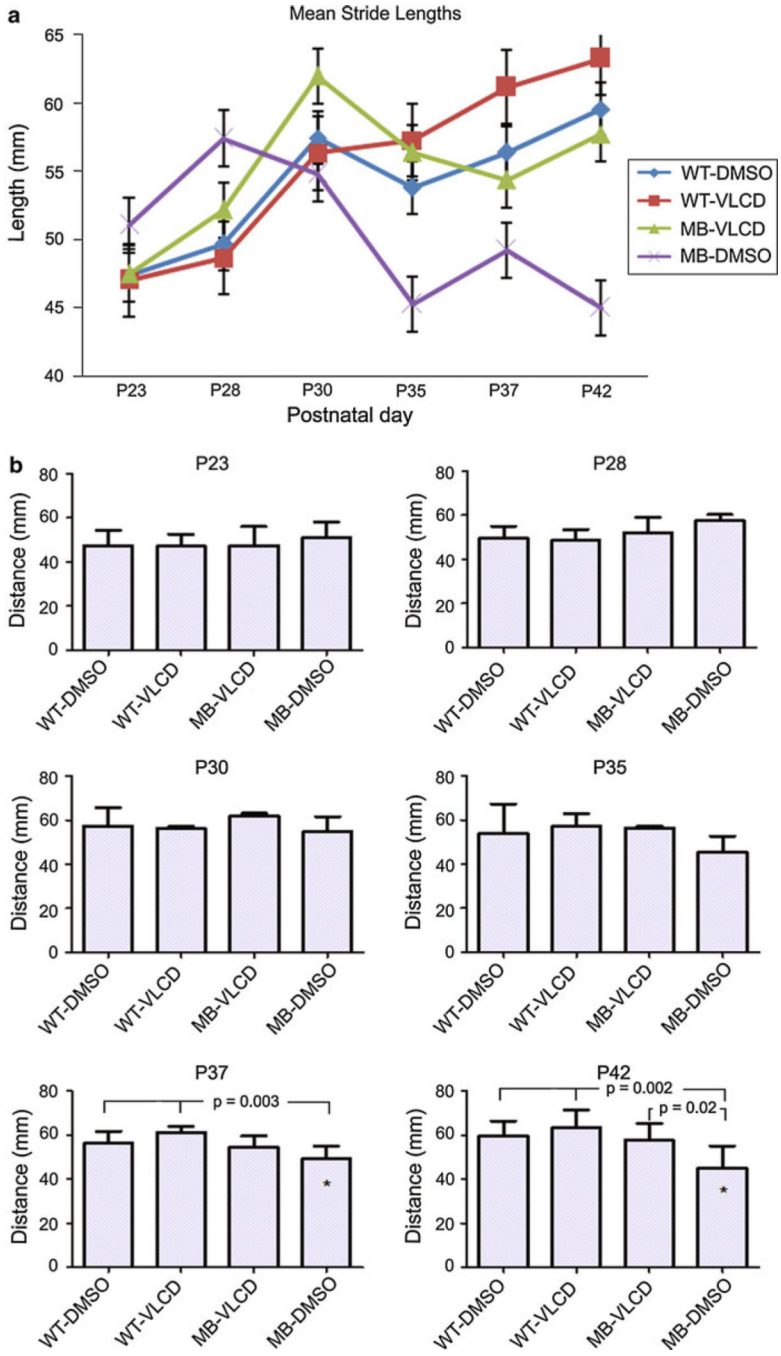

Fig. 3.

Stride Length Analysis. **a** The mean stride lengths for each group on each day are plotted. A two-way ANOVA with a general linear model was performed and found a significant difference in stride lengths between the MB-DMSO groups and the wild type mice on P42 (**P* = 0.01). Error bars, s.e.m. **b** A one-way ANOVA with a Student Newman-Keuls test was performed for each day to determine if there was a significant difference between the stride lengths of the mice in each group on any given day. On P37 and P42, MB-DMSO was found to have a stride length significantly different (*P* = 0.003, 0.002, respectively) than

both wild type groups. On P42, MB-VLCD and MB-DMSO groups showed a significant differences in stride length ($P \leq 0.020$)

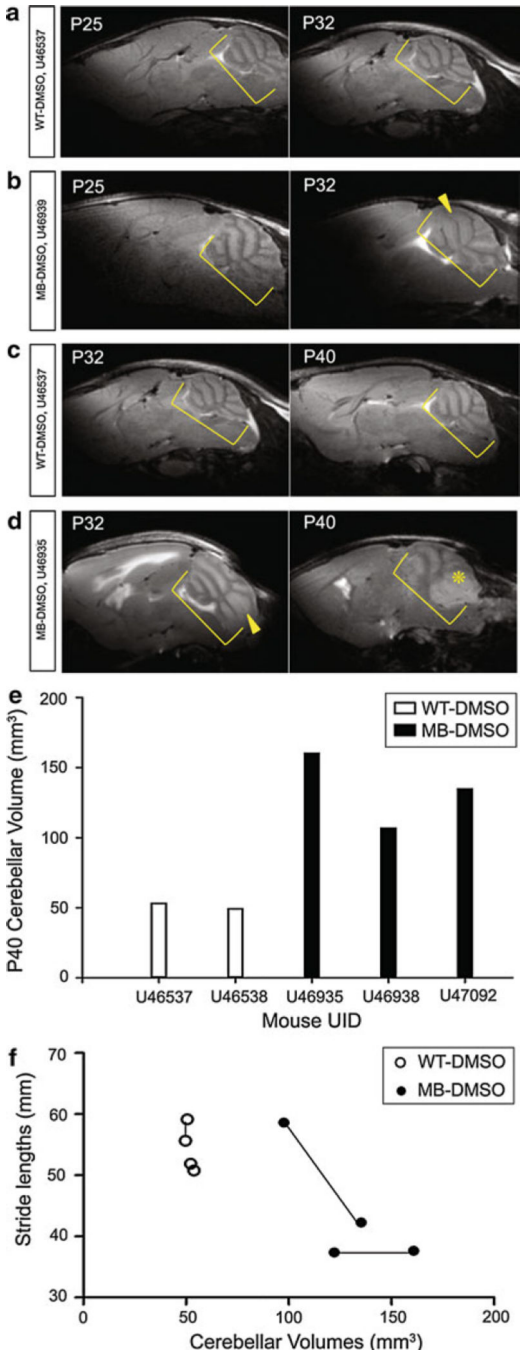

Fig. 4.

Wild type and Medulloblastoma-prone Cerebellum MRI Scans during Progression. *Yellow brackets* denote cerebellums, with layer thickening (*arrowhead*) and tumor mass (*asterisk*). **a** WT-DMSO (*top*) P25 (*left*) and P32 (*right*) on same mouse, U46537. **b** MB-DMSO (*bottom*) P25 (*left*) and P32 (*right*) on same mouse, U46539. **c** WT-DMSO (*top*) P32 (*left*) and P40 (*right*) on same mouse, U46937. **d** MB-DMSO (*bottom*) P32 (*left*) and P40 (*right*) on same mouse, U46935. **e** WT-DMSO and MB-DMSO cerebellar volumes calculated from MRI scans on P40. **f** Correlation curve between cerebellar volumes and stride lengths of MB

mice (n = 2) and WT mice (n = 2). Each set of 2 connected points indicates one mouse from P32 to P40. Curve is only suggestive due to small sample number

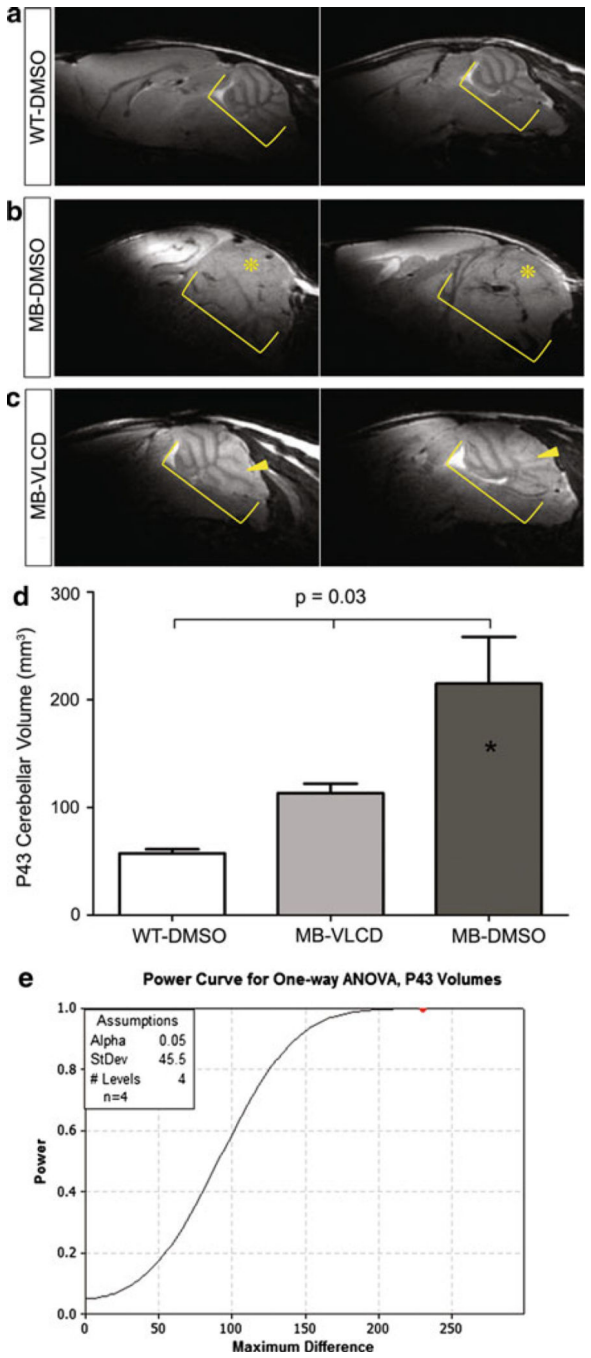

Fig. 5.

Wild type and Medulloblastoma-prone Cerebellum MRI Scans at End of Treatment. *Yellow brackets* denote cerebellum. **a** WT-DMSO cerebellums on P43. U46359 (*left*) and U46362 (*right*), **b** MB-DMSO cerebellums on P43. Scan shows enlarged cerebellum with folia layer thickening and tumor mass (*asterisk*). U46337 (*left*) and U46333 (*right*), **c** MB-VLCD cerebellums on P43. *Yellow arrows* show tumor development. U46330 (*left*) and U46327 (*right*), **d** One-way ANOVA with Student Newman-Keuls results for difference in cerebellar volumes between WT-DMSO, MB-VLCD, and MB-DMSO (MRI segmentation). There was

no significant difference between MB-VLCD mice and WT-DMSO cerebellar volumes. There was a significant difference between MB-DMSO and both other groups ($P \leq 0.03$). **e** Power analysis shows a power of ≥0.80 to detect a difference of 0.05 with 4 mice/group