Tests to assess motor phenotype in mice: a user's guide

Simon P. Brooks and Stephen B. Dunnett

Abstract | The characterization of mouse models of human disease is essential for understanding the underlying pathophysiology and developing new therapeutics. Many diseases are often associated with more than one model, and so there is a need to determine which model most closely represents the disease state or is most suited to the therapeutic approach under investigation. In the case of neurological disease, motor tests provide a good read-out of neurological function. This overview of available motor tasks aims to aid researchers in making the correct choice of test when attempting to tease out a transgenic phenotype or when assessing the recovery of motor function following therapeutic intervention.

6-OHDA
A neurotoxin that is selective for catecholamine neurons and is widely used to produce an animal model of Parkinson's disease following injection into the nigrostriatal dopamine pathway.

The Brain Repair Group, School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3AX, South Wales, UK.
e-mails: brookssp@cf.ac.uk; dunnett@cf.ac.uk
doi:10.1038/nrn2652
Published online 10 June 2009

In the past decade there has been a proliferation of studies characterizing behavioural phenotypes in mice, which frequently take the background literature in rats as their starting point and adapt it to the mouse. This has been driven in part by the need for standardized phenotype screening in large-scale programmes of genetic variation, such as the Harwell mouse mutagenesis programme[1]. In this Review, we focus on the development, validation, characterization and quantitation of tests of mouse motor function, the most widely explored behavioural phenotypes in neurological research. We seek to provide a critical appraisal of the range of tests available and of their validity for different experimental applications, as well as guidance on the factors that need to be considered when selecting the most appropriate test(s) for a particular experimental programme or application. We focus on models of basal ganglia disorders — in particular, Parkinson's disease (PD) and Huntington's disease (HD) — and readers are referred to REFS 2–4 for descriptions of motor tests specifically related to the assessment of cerebellar ataxias.

A selection of the most widely used tests are illustrated in FIG. 1 and described individually in more detail below. Four aspects of testing need to be considered during experimental planning: validity, reliability, utility and sensitivity (BOX 1; TABLE 1).

We review the tests that are available for assessing motor function in mice, in order of increasing specificity and complexity. We also attempt to categorize them according to the specific aspects of motor function that are assessed, although of course many of them measure several aspects of behaviour. Most tests of motor function in mice were originally designed for rats; here we limit the discussion to tests that have been found to work reliably in mice.

General neurological tests

Generic neurological screens. Many laboratories have generated test batteries for screening motor impairments in mice, with differing degrees of coverage, complexity and validation[5–10]. Of these, the most comprehensive and systematically validated is the SHIRPA battery, which offers a collection of simple tests selected to provide a standardized, high-throughput screen for the rapid assessment of mouse phenotypes[9,11] (BOX 2). The protocol involves a three-stage screening procedure that is designed to mimic the human neurological and psychiatric diagnostic procedure. The SHIRPA test battery is an excellent tool by which to obtain an initial and relatively broad screen of mouse phenotypes. However, although it highlights features for further specific analysis, it does not provide useful individual dependent variables of sufficient sensitivity and discrimination for routine use in specific quantitative analysis.

Cylinder test. The SHIRPA battery uses defined environments for observing and recording behaviours. An observational test using a specific item of this equipment is the cylinder test (FIG. 1a; see Supplementary information S1 (movie)). This was first developed for detecting forelimb impairments in rats with unilateral 6-hydroxydopamine (6-OHDA) lesions, an animal model of PD, and has proved to be a simple and efficient test of unilateral deficits in voluntary forelimb use[12,13]. Mice are placed in a glass or perspex cylinder and filmed with video equipment either from below or from the side. Mice with unilateral

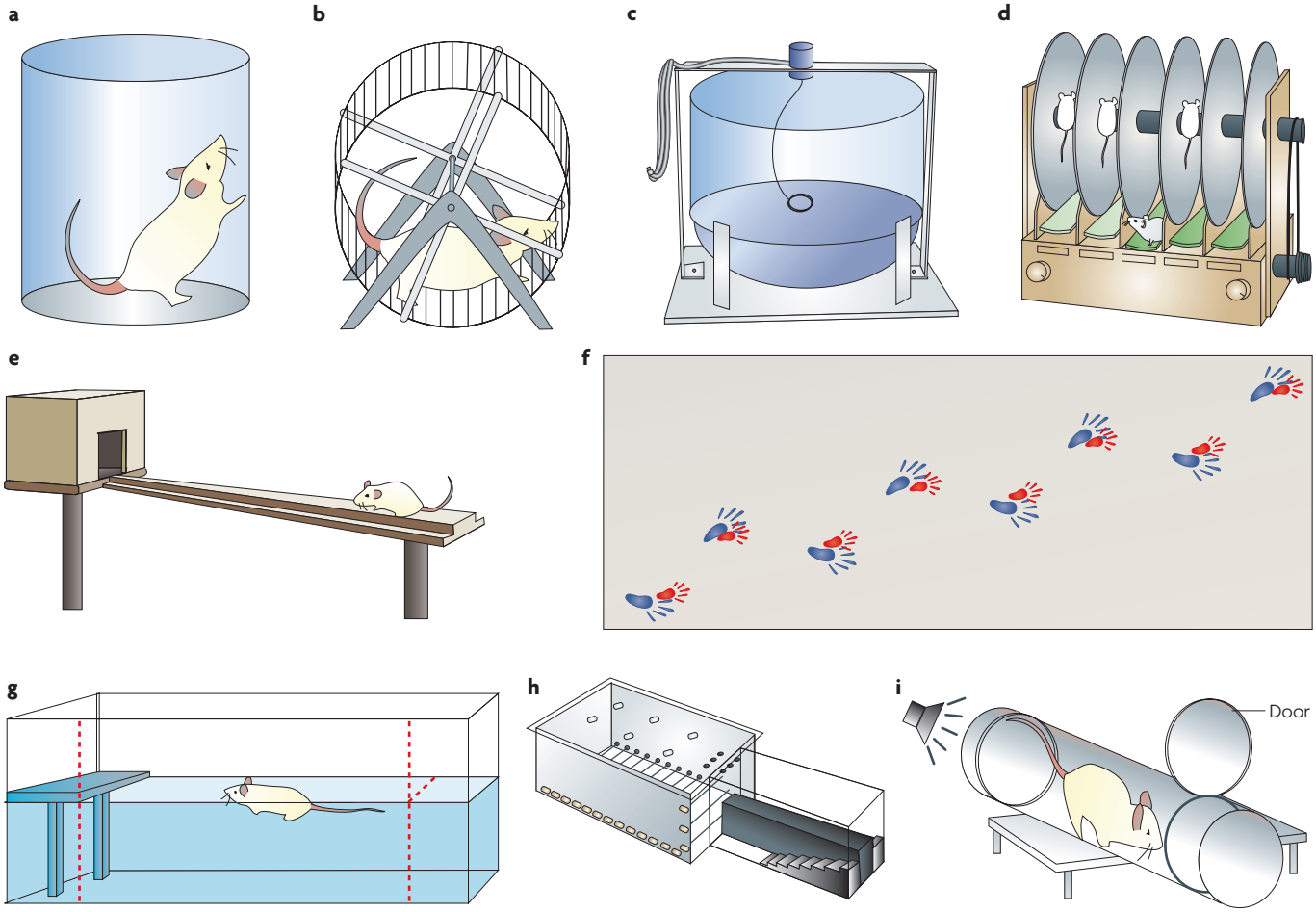

Figure 1 | **A selection of behavioural tests for mice. a** | Cylinder test. The mouse is placed in a transparent cylinder, and forepaw exploration is assessed by the number of instances the animal uses each paw to touch the cylinder. **b** | Running wheel. The running wheel is used as a measurement of voluntary home cage activity, often over a 24 h period. **c** | Rotometer. The rotometer is used to dermine the extent of lateralized lesions and the subsequent efficacy of therapeutic interventions by measuring the animals' rotational bias, typically after administration of a psychomotor stimulant drug. The test is most frequently used in conjunction with unilateral 6-hydroxydopamine lesions as a model of Parkinson's disease. **d** | Rotarod. The rotarod is a beam that can rotate at a fixed or an accelerating speed. Mice are placed on the beam and their latency to fall provides a measurement of their motor coordination. **e** | Raised beam test. Mice are placed on the beam and their ability to traverse it is considered to be an indication of their balance. Paw slips and traverse time are used as the measurements. **f** | Footprint analysis. The paws are dipped in ink or paint, so that the mice leave a trail of footprints as they walk or run along a corridor to a goal box. Measurements of stride length, base width, and fore and hind paw overlap give an indication of gait. Automated versions of the task use video processing of footage taken from below the mice. **g** | Swimming test. The mice are filmed swimming the length of a perspex tank to an escape platform. The video footage is subsequently analysed for the number of fore and hind limb strokes and the latency to reach the platform. **h** | Staircase reaching test. This is a test of manual dexterity, as the mice are required to reach down to retrieve food pellets from the steps of the staircase. They can use only the paw on the same side as the staircase to retrieve pellets, and so the test provides a simple measurement of lateralized impairments. The actual measurements are the number of pellets retrieved and the number displaced. **i** | Acoustic startle chamber. Acoustic startle and prepulse inhibition measure 'sensory–motor gating' by assessing the animals' ability to detect a warning prepulse that subsequently dampens the reflexive startle response. The animals are placed in a restraining tube housed on a force transducer that measures startle amplitude in response to the stimuli.

6-OHDA lesions exhibit a decline in use of the contralateral paw when rearing to explore the environment[14]. Paw usage can be determined by measuring the number of impaired forelimb (load-bearing) wall contacts made with the contralateral paw, as a percentage of total contacts[14], and/or the amount of time spent with paws in contact with the walls. This test is easy to carry out, takes no more than 5 minutes of observation to make an accurate assessment, and requires no training. As the primary outcome measurement involves an asymmetry between the affected and unaffected limbs, each animal provides its own control for individual differences in motivation, exploration and activation. However, therein also lies the test's major disadvantage: it is useful for detecting

Box 1 | **Criteria for test selection**

Test validity

To what extent does the test measure that which it is designed to measure? Three important types of validity need to be considered. Face validity assesses how well the behaviour in the animal matches that which is observed in humans. For example, for a dyskinesia test, the face validity will be how closely the range of abnormal movements reflects similar abnormal movements in human patients[15]. Construct validity addresses whether test performance relies on the same neurobiological mechanisms as those that underlie the disease. For example, chronic treatment with L-3,4$(OH)_2$-phenylalanine (L-DOPA) causes abnormal dyskinetic movements in both animals and humans when it is administered in the context of long-term depletion of endogenous nigrostriatal dopamine (resulting from Parkinson's disease in man[117] or from explicit lesioning in the animal[15]). Assessments of test validity must take into account the particular sensory and motor repertoires of mice, which have evolved to overcome the natural challenges that mice have to face. Finally, predictive validity reflects how well the response to an experimental manipulation, such as drug administration in humans, can reliably be predicted from the response to the same manipulation in the animal model.

The validity of behavioural tasks can often be probed pharmacologically. For example, a task that measures anxiety should be sensitive to anxiolytic and anxiogenic compounds[118], a test of locomotor activity should be sensitive to stimulant drugs[119], and a test of depression and reward should be sensitive to neuroleptic and antidepressant drugs[120].

An important component of test validity is whether the tests selected are appropriate for the particular level of motor organization and function that is under experimental manipulation.

Test reliability

A second set of criteria relates to test reliability — how reproducible are the measurements obtained from a particular test? Test–retest reliability is a measurement of the likelihood of obtaining the same result if the test is repeated. This seems to be straightforward, but it is complicated by the fact that the repetition itself can alter the outcome, most obviously owing to fatigue or the effects of learning. Inter-rater reliability relates to the reproducibility of a test for different investigators, either on separate occasions (in which case it will be confounded by test–retest reliability) or through independent ratings of the same data (for example, through separate scoring of video records). Tests based on rating scales require particular attention to standardization of raters and training, as well as explicit determination of inter-rater reliability. An important component in test design is to seek objective measurement, ideally by automating data collection.

Test sensitivity

The sensitivity of a test relates not only to the particular test demands but also to the specific brain regions that are involved. Consequently, a particular test may be highly sensitive to disturbance in one branch of the motor system but less so in another, depending on the correspondence between the particular aspects of motor control provided by the affected system and the performance demands of the test[121]. Once a test that is appropriate to the system under investigation has been selected, various strategies are available to increase its sensitivity in order to detect real differences between experimental groups. The most efficient approach is to enhance statistical power by using multivariate statistics, by including full repeated measures designs and by adopting the most powerful *post hoc* testing protocols[122]. A simple, although not always practical, additional means of enhancing power is to increase animal numbers; this always needs separate consideration of the ethical imperative of the 'three Rs' — reduction, refinement and replacement[123].

Test utility

In addition to the more formal criteria of validity and reliability, test selection is invariably influenced by a range of more practical factors (TABLE 1). Some features of test design directly affect experimental power. The sensitivity of a test to small differences in performance needs to be balanced against the variation of the experimental measurement from one sampling period to another — whether due to intrinsic variability of the animal's performance or to variability in the measurement — in order to optimize the statistical power of the test to detect clear differences between experimental conditions. Such power analyses, whether formal or implicit, affect the number of animals required to achieve statistical detection of significant differences. This has clear cost implications, not just in financial terms but equally importantly in terms of the capacity of the animal breeding programme to generate sufficient animals of each particular genotype at the appropriate ages, and in terms of animal welfare considerations.

the effects of unilateral motor impairment, as the lesion induces a marked and lateralized loss of function, but it is less suitable for measuring bilaterally symmetric dysfunction, which is often subtle and is the type of dysfunction caused by most genetic mutations.

AIMs scales. Observational scales have been well defined for rating abnormal involuntary movements (AIMs) that constitute dyskinesias associated with chronic L-3,4$(OH)_2$-phenylalanine (L-DOPA) treatment[15] or cell transplantation[16,17] in dopamine-depleted rats, and similar scales have been validated for mice[18–20]. AIMs ratings are typically conducted in a transparent open environment such as the perspex cylinders used in the cylinder test and typically record the presence, duration and severity of abnormal axial, limb, orofacial and locomotive movements in 1 minute time bins every 20 or 30 minutes over 150 minutes of drug action. From these ratings, an overall AIMs score is calculated as the area beneath the curve on a plot of AIMs score per observation period against overall observation time[18]. The AIMs scoring system in conjunction with a well-characterized lesion and a known cause of dyskinesias (for example, chronic L-DOPA treatment) provides a good functional mouse model of the human PD condition. It is the application of the rating scale to a good model of the disease that is valuable, rather than the application of a rating scale to a nonspecific pathology, which provides little information about the underlying disease state and hence potential therapeutic approaches.

Other observational measurements. Other observational techniques are frequently used as stand-alone assessments of mouse motor behaviour. These include general assessments of mice in the home cage or test arena using rating scale assessments of general motor behaviours, such as tremor, gait abnormalities, abnormal limb displacements, righting reflexes, walking and running behaviours[21,22]; assessment of specific abnormalities, such as body clasping in mice when suspended by the tail[14,23,24]; and rating scales of extensor reflexes[25].

Locomotor activity tests

Open-field test. The simplest tests of locomotor activity involve observing and recording an animal's movements around an open arena. When placed in the centre of a field, a mouse will typically run to the walled edge and then explore its way around the whole arena while remaining close to the wall. Over time, as the animal habituates to the new environment and its anxiety reduces, the mouse will increasingly venture out towards central parts of the arena before returning to the edges. Exploitation of this behavioural profile forms the basis of the study of open-field locomotor activity in mice. Open-field tests for mice can be conducted in circular[26,27] or square arenas[28–30], divided into segments or squares, and will typically record (by observation) the number of edge and centre squares entered, along with bouts of rearing, grooming and defaecation. A recent study[31] found that it was possible to separate anxiety-related activity from purely motor-related activity in

L-DOPA
The biochemical precursor to the neurotransmitter dopamine. It is used as a therapeutic drug for dopamine replacement therapy in Parkinson's disease.

Time bin
Subdivision of an experimental time period, the use of which enhances a test's sensitivity.

C57BL/6J mice by administering the anxiolytic drug chlordiazepoxide at concentrations that did not affect levels of motor activity. However, this approach is not feasible in most experimental scenarios; thus, although it provides a simple measurement of locomotor activity, the motor component in the open-field test is typically confounded by anxiety levels in the mice.

A major improvement in measuring locomotor activity is to automate recording of mouse movements, typically using video tracking, photocell beams or electromagnetic detector technologies[29,32–34] (see Supplementary information S2 (movie)). These can be mounted in the home cage as an alternative to using separate banks of test cages, avoiding the anxiety that is induced by a new environment. The SEE (Software for the Exploration of Exploration) tool[34] is an advanced example of video tracking technology that breaks explorative motor behaviours down into discrete aspects and provides a highly detailed assessment of open-field activity. The advantage of automated systems such as the SEE tool is that they provide standardized, objective measurements of activity that can be relatively complete and can be collected in different bin sizes over short or extended periods of time.

Table 1 | **Criteria for selecting optimal motor tests**

Criterion	Sub-criteria	Description
Validity	Construct validity	Does test performance rely on the same neurobiological processes that are thought to underlie the disease?
	Face validity	Does the animal's behaviour seem similar to a relevant class of human movement?
	Predictive validity	Do the effects of treatments (e.g. drugs) in animals predict what those same treatments will do in humans?
Reliability	Inter-rater reliability	Do two experimenters produce the same result from the same behavioural observation?
	Test–retest reliability	Does a repeat test on the same animal produce the same result?
Sensitivity	Sensitivity	Is the measure sufficiently fine-grained to detect small differences in behavioural performance?
	Power	Is the variability of random differences in performance small with respect to the differences associated with target experimental manipulations?
Utility	Cost efficiency	Is the test cheap and easy to implement?
	Testing efficiency	Is the test quick and easy to run, without requiring extensive training?
	Objectivity	Is the behavioural measure automated or subject to objective criteria, or does scoring require subjective experimenter ratings and judgements?
	Training	Can the test be rapidly implemented in a standard form by new technical staff, or does it require extensive training, technical skill to conduct or a high level of experience to judge performance?
	Bilaterality	Is the test suitable for determining bilateral phenotypes, or is it more appropriate for comparing performance on the two sides of the body within a single animal?

Wheel running. An alternative measurement of voluntary locomotor activity is provided by the tendency of mice to engage in extended periods of running when they are provided with a running wheel in the home cage[22,35–37] (FIG. 1b). In wheel running studies, the total number of wheel rotations (distance travelled), the speed of running (rotations per 3 minutes), the time to initiate running and the number of breaks from running can all be used as measurements of activity. Wheels vary slightly in size but are generally ~10–15 cm in diameter. In such a wheel, a normal C57BL6/J mouse will run ~5 km per day[38].

The advantages of running wheel tasks are that they produce high-quality data, are fully automated and can be carried out in the home cage. One performance issue is that the oestrous cycle in female animals affects their running wheel behaviour[39], so mixed-sex mouse cohorts should not be used in these tasks. A second potential confound is that wheel running might be an anxiolytic behaviour in mice that reduces the anxiety associated with several measurements of motor behaviour, including open-field activity and acoustic startle[40]. Consequently, running wheel behaviour in the mouse might at least partially reflect the underlying level of stress, which might be an unexpected side effect of a genetic modification.

Circadian rhythms. The identification of clock genes and their analysis in genetically modified mice have attracted a lot of interest over the past two decades[41]. Mice are nocturnal animals, and their activity peaks within the first hour of the dark phase. Their motor activity changes as a consequence of changes in the light–dark circadian cycle, and these changes are clearly manifested in a number of transgenic models of disease, such as the R6/2 mouse strain that carries the HD mutation[42] and mice infected with bovine spongiform encephalitis[43]. These changes are reflected in the flattening and progressive disruption of the 24 h activity cycle.

Rotation. For four decades, rotational behaviour has been a widely used measurement of motor asymmetry in the unilateral 6-OHDA lesion rat model of PD, but an increasing number of studies have extended this model to mice[14,44–46]. Following unilateral forebrain dopamine depletion, rats or mice exhibit a postural bias to the side of the lesion; this can be amplified and transformed by pharmacological activation into a head-to-tail turning (rotation) response. Stimulation of dopamine release from intact terminals by indirect agonists (for example, amphetamines) enhances the spontaneous bias, causing ipsilateral rotation, whereas activation of 'supersensitive' receptors on the denervated side by low doses of a direct agonist (for example, apomorphine) drives vigorous contralateral rotation. The rotation response itself is recorded either by direct observation or, more typically, in automated rotometer test bowls[47] (FIG. 1c). The rotometer apparatus is particularly suitable for evaluating the temporal profiles of pharmacological[48,49], cell[45,50] or gene[51] therapies with dopaminergic targets of action.

The strength of the rotation tests is that the 6-OHDA lesion is a well-validated model of PD with highly

Box 2 | **SHIRPA neurological test battery**

The 'SHIRPA' (SmithKline Beecham, Harwell, Imperial College and Royal London Hospital phenotype assessment) battery was developed as a generalized neurological screen to assess phenotypes in a large-scale multicentre random mutagenesis programme[9]. The primary SHIRPA screen covers simple measurements of physical and neurological health, weight, body condition and sensory and motor function, and serves to identify global disturbances in gait, posture and muscle tone deficits, as well as motor control and coordination abnormalities. The secondary screen incorporates several of the tests of motor function described in more detail in the main text, including the accelerating rotarod, the spontaneous locomotor activity test and open-field rearing measurements. The tertiary assessment focuses on psychiatric and cognitive aspects of behaviour and includes prepulse inhibition. Of the three stages of assessment it is the primary assessment that offers the greatest general overview of any motor abnormalities in mice. During the primary screen, 40 different observational ratings are made of tremor, body position, spontaneous activity and posture, followed by assessments of grip strength, limb clasping and trunk curl during suspension by the tail, as well as several automatic responses including pinna, toe pinch and righting reflexes (see the MRC Harwell web page on SHIRPA). The SHIRPA core has subsequently undergone modification with additional morphological profiles of coat colour, hair length and morphology of the head, tail, whiskers, teeth, limbs and ears[124]. Although the primary SHIRPA screen has been found to differentiate inbred mouse strains[11] and to detect a wide range of genetically modified mice and mouse models of disease[125–129], it is not particularly sensitive compared with other methods of assessment. For example, transgenic synphilin 1 mice, which might be a useful model of Parkinson's disease, exhibit marked deficits in rotarod and footprint testing at a stage of disease in which the primary SHIRPA screen fails to identify any difference from wild-type mice[23]. The SHIRPA screen also failed to detect differences between 129/SvJ and C57BL/6J mice that were developmentally deficient in vitamin D and control mice, but differences in the level of exploration were identified using the hole board test[130]. The disadvantage of observational tests is that there is no real way to increase their sensitivity other than by using video analysis techniques, which reduces the rapid assessment capability that the screen was designed to have. However, the SHIRPA is still an excellent initial screen for large numbers of mice with unknown phenotype.

specific neuropathology, and rotational measurements are sensitive, reliable and reproducible. However, rotation tests are relevant only to unilateral lesion models, and are not directly applicable to the wide range of non-lateralized phenotypes associated with genetic manipulation *per se*.

Motor coordination and balance

Rotarod. The rotarod was specifically designed for making automated measurements of neurological deficits in rodents[52], and is one of the most commonly used tests of motor function in mice (FIG. 1d; see Supplementary information S3 (movie)). Early designs use a rotating rod of ~3 cm diameter, on which the mouse is placed and has to maintain its balance; a trip switch on the floor below is set to record the latency until the mouse falls from the rotating rod. Mice are tested on separate trials at a series of fixed speeds (for example, see FIG. 2b), or speed increases can be incorporated into a single trial by using an accelerating version of the test[53]. In accelerating versions, the range of rod rotation speeds can differ markedly between studies, but typically revolutions of the rod accelerate smoothly from 0 to 40 rpm over a 5 minute period.

The accelerating test is quicker and more efficient, but it confounds motor coordination at different speeds with fatigue, whereas the fixed-speeds test provides separate data on each range of rotation speeds and is probably more sensitive[54]. The fixed-speeds test has been used to demonstrate that the age of onset of transgenic phenotypes is dependent on task difficulty and the sensitivity of the test to motor symptoms[6] (FIG. 2a).

There are several common confounds of the rotarod test. The first is that some animals may cling to the beam, and rotate with it, rather than fall when they lose balance. This is due to some commercial models having a rod that is grooved to aid grip, but to which the mice can cling by their claws; a simple solution is to cover the rod with a layer of coarse rubber. The second confound relates to individual animals that refuse the test and simply fall as soon as they are placed on the rod. This is especially relevant in longitudinal assessments, during which the animals can learn over repeated tests that the consequences of falling are innocuous. Fortunately these animals are relatively rare and their performance is conspicuous relative to that of the other mice in the cohort, thus they can be (and need to be) excluded as 'outliers' in any statistical analysis. A third confound relates to mouse weight: heavy mice perform worse than light mice. Thus, genetic or lesion-induced weight loss can offset motor disability and potentially skew results. Finally, and particularly with the accelerating version of the rotarod, speed is confounded by fatigue at progressively longer latencies. However, demonstration of differential deficits at higher rotation rates in a series of fixed-speed tests[6] can be used to ensure that a more rapid fall is indeed attributable to problems with motor coordination rather than to greater susceptibility to fatigue. Despite these confounds, the rotarod remains one of the main tests of motor function in the mouse owing to its ease of use and sensitivity.

Balance beams. The balance beam assesses a mouse's ability to maintain balance while traversing a narrow beam to reach a safe platform (FIG. 1e). It was originally designed to assess motor deficits in aged rats[55] and has proved equally useful in assessing motor coordination and balance in young, lesioned and genetically altered mice[6,56,57]. Measurements recorded include the time taken to cross the beam and the number of paw faults or slips. Some versions use a range of cross sections and diameters to vary the task difficulty[6,57]; others use a beam that becomes progressively narrower as it approaches the safety platform[58]. In early versions of the test, which use simple square and round cross sections, the mice may occasionally fall, and the frequency of falls becomes an additional dependent variable. Two useful modifications of beam design can promote a mouse's willingness to progress rapidly across to the escape platform rather than simply cling on to prevent falling: an additional ledge can be placed either side of the platform so that even when the paw slips grip is maintained; and an inclined beam can be used instead of a horizontal beam, as mice seem to have a natural tendency to run upwards to escape (see Supplementary information S4 (movie)). These modifications help to increase the sensitivity of the test, so it can be used for the early detection of motor deficits in mouse models of HD (S.P.B., unpublished observations).

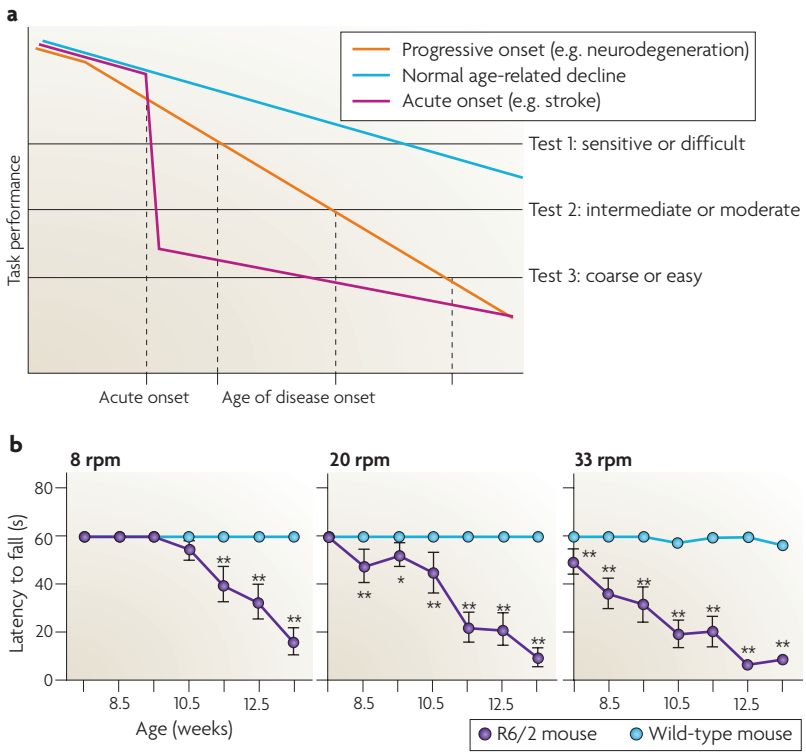

Figure 2 | **Rates of progression of neurological impairment and onset of phenotype. a** | Normal age-related decline plotted against the development of sudden or acute and chronic or progressive neurological impairment. Depending on the sensitivity of the test, the age of onset can vary markedly. **b** | Rotarod performance in R6/2 mice carrying the Huntington's disease mutation and in wild-type mice at three levels of task difficulty. The harder the task, the earlier and greater the deficit. Figure is modified, with permission, from REF. 6 © (1999) Society for Neuroscience.

Gait analysis. A more detailed analysis of motor coordination and synchrony is provided by examining gait during normal walking. One approach is to use a high-speed video camera to measure stride length while the mouse is running on a treadmill[59]. However, such equipment is expensive. The more commonly used method for assessing gait is the 'footprint' test. The fore and hind paws are painted with dyes of different colours and the mouse is encouraged to walk in a straight line (typically in a narrow corridor) over absorbent paper[6,60,61] (FIG. 1f). The footprint patterns are then analysed for a range of measurements, including stride length, base width, overlap between fore and hind paws, and paw and finger splay.

Gait analysis is not only simple, and the scoring straightforward, but also it is one of the few tests that translates directly from animal to human studies. Although automated video analysis is possible, it is too early to determine whether this matches the sensitivity of the time-consuming manual analysis; for example, one study failed to detect quite clear overt impairments in superoxide dismutase 1-mutant and MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine)-treated mice when using automated analysis[62]. Despite this, advances in video analysis technology provide the opportunity to assess aspects of movement in detail that the manual footprint test cannot achieve.

Swimming. Coordinated limb use during voluntary locomotion can also be assessed when mice are swimming. The mouse is video recorded while swimming in water from one end of a clear perspex tank to a visible escape platform at the other end (FIG. 1g; see Supplementary information S5 (movie)). The video is then analysed for the number of left and right forepaw and hindpaw strokes, swimming speed and latency to traverse the tank, and note is taken of any incoordination[57]. Swimming has been found to be a sensitive test of motor coordination in normal[7] and genetically modified mice[6,57,61]. The video analysis component of the test provides great sensitivity, but also increases the time to process the data. There may also be a risk of drowning in some mouse lines, especially at advanced stages of disease, so appropriate caution must be exercised.

Climbing. Climbing tests are used to monitor motor impairments in aged animals[55] and sensorimotor impairments associated with hypothalamic and nigrostriatal motivational systems in rats[63] and mice. The mouse is placed on a tilting or vertical grid or frame that induces a particular pattern of escape response (for example, climbing) in normal animals and then its selection from alternative behaviours, the speed of its completed response and/or the accuracy and precision with which it grips and coordinates its escape behaviour are recorded.

Various climbing frames have been used. In the vertical-pole task, the mouse is placed on a vertical wooden or wire-mesh pole with its head facing upwards[64]. A normal mouse will grip the pole before turning through 180° and slowly climbing down to the base of the pole. Latencies to turn through the 180°, to descend once turned and to complete the task are primary performance measurements, and abnormalities in response can include climbing, falling or slipping down backwards[22,65,66]. Alternatively, if the mouse is placed on a horizontal pole or grid which is then raised to a 45° angle it will climb[67]. Rope or string climbing tests have also been used to differentiate mutant mice from wild-type mice. In these tests the mouse is simply placed at the bottom of a suspended piece of rope with regularly placed knots, and its ascent is timed[68]. A further variant is the 'coat hanger' test, which times a mouse climbing from a hanging position on the base of the coat hanger to the uppermost point of the central vertical handle at the top[32,69–71].

These are relatively simple but nonspecific (climbing involves several aspects of motor function) tests of motor dysfunction, and have value as rapid screening methods for disability. However, their quantitation is relatively coarse, they are difficult to standardize and they provide little information about the nature of the underlying disorder.

Grid stepping. There are several variants of this test, which involves video recording mice from below while they walk across a horizontal grid and counting the number of left and right paw slips through the grid[26,72,73]. Perhaps the most useful version is the ladder rung task[74]. In this test the rat or mouse has to cross a horizontal

ladder, grasping each rung with its fore and hind limbs. The spacing of the rungs can be made regular or variable to vary the difficulty of coordinating step and grasp movements. Making the spacing variable enhances the sensitivity of the test to deficits caused by small focal cortical lesions[73,74]. Although they are sensitive to specific ipsilateral and contralateral lesions, such stepping tasks require time-consuming video analysis and subjective observer ratings, and have not been widely used to evaluate non-lateralized impairments in genetically altered animals.

Grip strength. In contrast to the nonspecificity of the climbing tasks, grip strength tests are highly specific, in that they attempt to measure a single, well-defined aspect of behaviour. Grip strength has traditionally been measured in one of three ways: by testing the ability of the mouse to remain clinging to an inverted or tilted surface such as a wire grid or a cage lid[25,75] for a period of time, usually up to 1 minute (see Supplementary information S6 (movie)); by testing the ability of the mouse to hang on a wire with its forepaws for a preset length of time or until its grip fails[76]; or by making specific measurements of the force required to pull the mouse off a narrow bar that it is gripping[61,75,77] (see Supplementary information S7 (movie)). The test has been used in mice to measure tolerance to drugs, such as ethanol (which induces myorelaxation)[78], and to assess disease progression in complexin 1-knockout[61] or cerebellar lurcher mice[79].

Each of these methods has limitations. The inverted-grid test lacks sensitivity, and accurate assessment of grip strength is dependent on homogeneity of mouse weight across the cohorts. In addition, the mice tend to move around when the grid has been inverted, and consequently different muscle groups are being used or rested while the animal is moving. This contrasts with testing mice on the mechanical systems, where only forelimb muscles are being strained. Mouse movement can be difficult to overcome and tends to be strain dependent; agitating the grid before inverting it partially resolves the problem, as the animals increase their grip as the grid is shaken. The same weight problem applies to the wire hang test. The mechanical grip strength meters suffer from a single major problem: the unwillingness of the mice to hang on to the grip bars. This can be overcome to some extent by providing a metal grid for the mice to hang on to rather than a bar. Invariably several mice have to be excluded from the study on the grounds of non-participation, and multiple trials per animal are usually required to generate reliable data. One attempt to increase the sensitivity of the grip strength test has been to add different masses to a grid that the mouse grips while suspended by its tail[25].

Skilled limb use

Although in many instances it is desirable to measure gross aspects of motor function, such as stamina or balance, in other cases it is necessary to analyse finer aspects of voluntary motor control in normal, lesioned and/or genetically modified animals. Mice have good manual dexterity and a range of tests is available to describe and measure aspects of precision movements involved in food manipulation, reaching for food, and so on.

Descriptive paw use. Different rodent species exhibit marked commonality in the ways that they manipulate food. Whishaw and colleagues[80] used frame-by-frame video analysis combined with a dance movement notation system[81] to define several key features of food manipulation. This analysis provides a highly detailed characterization of fine motor control in a defined laboratory environment and, although it is not practical for routine experimental studies, provides an analytic framework in which many operational reaching tasks have been subsequently developed and validated[82–86].

Reaching tests. The first and simplest reaching tasks require the animal to reach through the bars of a wire mesh cage to retrieve food pellets, and the reaching success is rated[87]. Video recording and analyses of individual reaching movements in a single such task[84] have been widely used to characterize the effects of lesions, grafts and drugs in rats and mice[86]. However, to date, analyses in mice have focused on normal performance and the effects of unilateral lesions[83,88], and these tasks have had limited application to genetically altered mouse models.

Other tests, such as the Collins test[89], have recorded the paw preferences of mice collecting food pellets from feeding tubes. Handedness and therefore hemispheric laterality are determined by the number of reaches made by each forepaw out of a total of 50 reaches for each of several sessions, with scores of ≥29 considered to represent a significant degree of lateralized bias. Changes have been demonstrated following lateralized lesions[90], but such spontaneous preferences have not been widely used to characterize genetic mutations. These tests, like the cylinder and rotation tests, are obviously not appropriate for characterizing non-lateralized impairments.

Staircase test. The staircase test provides an entirely objective test of skilled reaching. Animals are allowed to access food pellets by reaching down on either side of a raised plinth located along the centre of a corridor in order to grasp and retrieve food pellets located on the steps of a baited descending staircase (FIG. 1h; see Supplementary information S8 (movie)). The narrow width of the corridor restricts the animal from turning around, and so the number of pellets that remain on each side provide a simple measurement of the reach and grasp coordination of the respective forelimbs. In mice this test is sensitive to cortical[91], basal ganglia[92], spinal[93] and ischaemic lesions[94]. Notably, although it was originally designed to measure lateralized reaching, the staircase test is equally sensitive to bilateral impairment[94].

Acoustic startle response

The startle reflex is "an unconditional reflexive response to a sudden environmental stimulus"[95]. It is usually assessed in a microprocessor-controlled startle chamber, in which the animal container is mounted on a force transducer to measure the startle reaction to

unpredictable, brief, intense acoustic stimuli (typically 50 ms bursts of 105–120 dB white noise) (FIG. 1i; see Supplementary information S9 (movie)). The startle response is inhibited if the stimuli are preceded by a brief warning stimulus (known as the 'prepulse'). Typically this is a 20 ms low-intensity burst of white noise at 4–8 dB above the background noise level that occurs 20–500 ms earlier than the stimulus[96]. Prepulse-mediated startle inhibition is affected in a range of neurological and psychiatric conditions, including HD and schizophrenia, in which the primary auditory response itself is unaffected, implying that there is a disturbance of a more central process that has been proposed to involve sensory gating of the motor reflex[97–99]. Similarly, psychoactive drugs[100,101] also disrupt prepulse-mediated startle inhibition, and in transgenic models of HD this disruption can appear before other aspects of the phenotype emerge[6,102].

However, the test is sensitive to a number of confounds: there are marked strain and sex differences[7,103]; body weight has a direct influence on the primary startle measurement (involving force)[23]; some transgenic strains and the normal ageing process are associated with overt hearing loss, affecting both the primary and the prepulse startle responses[104]; there is very wide variation and little standardization of test parameters between studies and laboratories; and, perhaps most crucially, the functional significance of a deficit in prepulse inhibition to primary motor coordination and control remains ambiguous. Nevertheless, because of their objectivity, automation, sensitivity and ease of application, prepulse inhibition and acoustic startle tests remain a popular tool both for screening new phenotypes and as an assay of the onset and alleviation of impairment[121].

Operant tasks

The most detailed analysis of motor (and cognitive) performance in rats and mice is provided by operant learning paradigms. Operant tasks in the classic automated lever pressing chambers ('Skinner boxes') have been the mainstay for animal learning studies in experimental psychology for almost a century[105], but they have not proved to be practical for testing mice because of the delicacy of the equipment required and the tendency of mice to satiate rapidly on even small food pellets. Operant testing in mice has become more popular since the introduction of related operant apparatus, the 'nine-hole box'[106]. In this apparatus (FIG. 3a) an array of holes in the chamber wall provides response locations. It is much easier to train mice to poke their noses into holes, owing to the similarity to their natural behavioural repertoire, as the primary response choice than it is to train them to press levers. Each hole contains a light to provide discriminative stimuli and a photocell to detect nose poke responses, so there are no moving parts to break down. A further adaptation is to introduce reinforcement using sweet fluids that can be delivered in small quantities using a peristaltic pump.

There is extensive evidence that the frequency and timing of reinforcement delivery influence animals' performance to maximize access to reward, be it food, water, sex, exploration or avoidance of pain. Several mouse studies have explored how different schedules of reinforcement influence the rates, timing and persistence of responding[107–109]. To date, these tests have mainly been used to assess specific aspects of motor learning, habit formation and attention.

Five-choice serial reaction time (5-CSRT) test. The 5-CSRT task was first developed for rats[110] and subsequently adapted to probe vigilance and attention in mice[106]. The task requires mice to respond rapidly to stimuli appearing at random in one of five holes in the array. In addition to allowing automated control and providing accurate estimates of reaction time and accuracy of responding, the task can be adapted to varying levels of difficulty by shortening the stimulus durations within or between sessions. Underlying response strategies are revealed by presenting two light stimuli, eliminating all stimuli, adding distracting stimuli (for example, a burst of white noise) on infrequent probe trials, and exploring patterns of responding in extinction. The 5-CSRT test has been widely used to probe the effects of psychoactive drugs, cortical and subcortical lesions, strain differences, chromosomal aberrations and the consequences of genetic knockout and disease-causing gene mutations in mice[110]. The reaction time measurement provides a good assay of basic motor capacity, but the 5-CSRT test is essentially a test of attentional function and is not ideal for specific motor analysis and habit learning.

Sequence learning. Sequence learning tasks in mice are essentially extensions of the choice reaction time task, with several stimuli linked in predictable sequences[111,112] or presented as embedded sequences among randomly presented comparators[113,114]. In the serial implicit learning task (SILT), mice are required to respond rapidly to two stimuli presented in sequence; most sequences are random, but when the first stimulus occurs in one specific location (for example, the second hole from the left), the location of the second stimulus is fully predictable (the hole two positions to the right of the first stimulus)[113,114] (see Supplementary information S10 (movie)). Mice learn to perform this motor habit learning task with a high degree of reliability and exhibit a distinctive pattern of responding: the speed and accuracy with which they move to the first hole are highest if the hole is in the centre and decline towards more lateral holes in the array, and the speed and accuracy of the second response diminishes the greater the size of the required movement (the number of holes between the first and second stimuli) (FIG. 3b). Both mice with striatal lesions and transgenic mice carrying the HD mutation are impaired on this task, demonstrating reduced accuracy and increased response latencies in the learned habit[113,114] (FIG. 3b). The fact that speed and accuracy of response in the SILT are higher for second stimuli that are predictable than for unpredictable stimuli (FIG. 3b) indicates that it is a good probe of implicit learning. However, both mice with striatal lesions and HD-transgenic mice were able to recognise predictable stimuli, and demonstrated higher

Operant tasks
Behavioural tests based on the principles of 'operant' (or instrumental) learning theory, in which the experimental subject makes voluntary responses of a specified type or timing to obtain a reward or to avoid punishment. The testing is typically conducted in automated apparatus, known as operant chambers.

Implicit learning
Learning that occurs without the explicit awareness of the learner. The classic example of this is a child's learning of the syntax and rules of language before schooling.

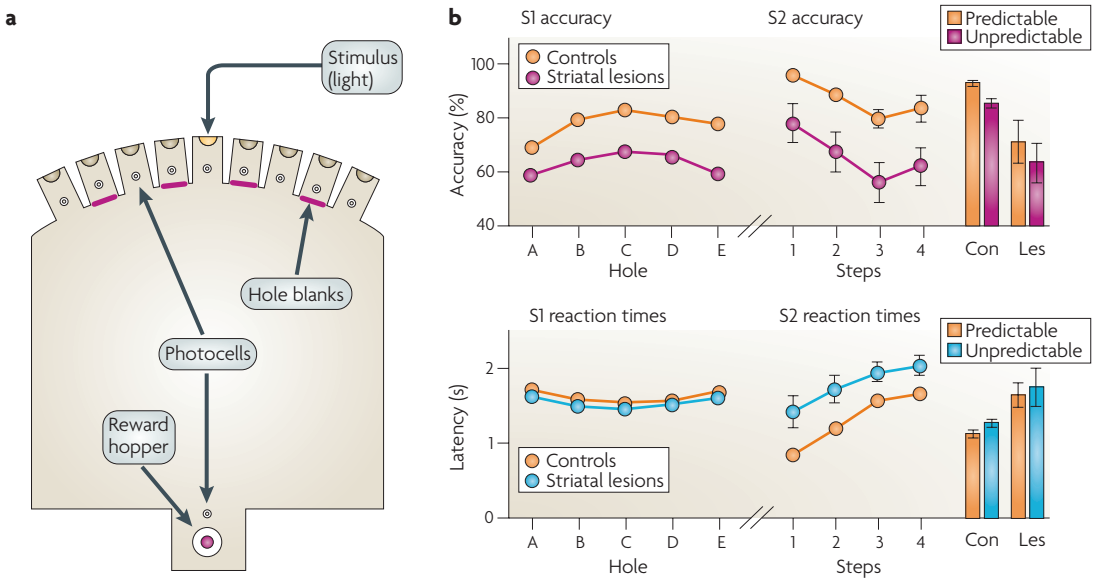

Figure 3 | **Serial choice responding in the SILT operant task. a** | A schematic illustration of the 'nine-hole box' apparatus used in the serial implicit learning task (SILT). **b** | Both the accuracy and the speed of the motor performance of the mice are disrupted by striatal lesions on successive choice responses, whereas the lesioned mice continue to show the benefit of implicit knowledge on predictable trials. Con, controls; Les, lesioned mice. Data are from REF. 113.

accuracy and decreased reaction times on the predictable trials than on the unpredictable trials [113,114]. This suggests that the fundamental deficit in these mice, as in people with HD, is in the performance, not the learning, of the task. Nevertheless, the sensitivity of the task to early impairments in striatal function is revealed by the fact that the deficit appears with the first signs of diffuse striatal huntingtin pathology and is apparent many months before either overt inclusions form or an overt phenotype becomes detectable using simple motor tests like the rotarod[113].

As with all tests of motor learning, operant procedural motor learning tasks have disadvantages. Weeks or even months of daily training can be needed to achieve a stable baseline of performance, which makes it difficult to assess mouse lines with an aggressive and rapidly progressing phenotype, such as the HD R6/2 mouse[24]. The mice all have to have their food intake restricted to similar levels, and this can be compromised by genetic, sex or strain differences in body weight, motivation or metabolism. Nevertheless, it is much easier in operant tasks than in other motor learning tasks to probe the precise functional locus of deficits, owing to the tasks' high sensitivity. It is also easier to determine the underlying motivation of the animals — by varying task parameters such as stimulus light intensity[106], stimulus duration, inter-trial influences on impulsivity, or reward magnitude[115,116]. Operant tasks are extremely powerful, in particular because they generate data from many standardized, automated, replicable trials daily, resulting in low within-subject variability.

Conclusion

Of all the domains of mouse behavioural science, motor testing is the most accessible. It does not necessarily require the use of expensive equipment, and the aspects of behaviour that are being assessed may be directly observable. However, as with all behavioural testing, careful planning of the experimental approach to select the most efficient, reliable and valid tests for the specific experimental purpose is the key to producing relevant data of a high quality. The demands differ depending on whether that purpose is to provide a behavioural assay with which to monitor a treatment in a mouse model of disease or a tool for functional analysis of a phenotype that is associated with a particular genetic mutation. It is often not sufficient to select just a single 'best' test, as different tests assess different aspects of motor performance and so a range of tests will be required to capture the mouse's condition in full and to maximize the chance of achieving a successful experimental outcome. Finally, the health and well-being of the animal are paramount to achieving a valid and interpretable outcome of any behavioural testing regime.

1. Nolan, P. M. *et al.* A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nature Genet.* **25**, 440–443 (2000).
2. Lalonde, R. & Strazielle, C. Spontaneous and induced mouse mutations with cerebellar dysfunctions: behavior and neurochemistry. *Brain Res.* **1140**, 51–74 (2007).
3. Lalonde, R. & Strazielle, C. Motor performance and regional brain metabolism of spontaneous murine mutations with cerebellar atrophy. *Behav. Brain Res.* **125**, 103–108 (2001).
4. Triarhou, L. C. *Neural Transplantation in Cerebellar Ataxia* (Landes, 1997).
5. McIlwain, K. L., Merriweather, M. Y., Yuva-Paylor, L. A. & Paylor, R. The use of behavioral test batteries: effects of training history. *Physiol. Behav.* **73**, 705–717 (2001).
6. Carter, R. J. *et al.* Characterisation of progressive motor deficits in mice transgenic for the human Huntington's disease mutation. *J. Neurosci.* **19**, 3248–3257 (1999).
7. Brooks, S. P., Pask, T., Jones, A. L. & Dunnett, S. B. Behavioural profiles of inbred mouse strains used as transgenic backgrounds. I: motor tests. *Genes Brain Behav.* **3**, 206–215 (2004).
8. Crawley, J. N. & Paylor, R. A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice. *Horm. Behav.* **31**, 197–211 (1997).

9. Rogers, D. C. *et al.* Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. *Mamm. Genome* **8**, 711–713 (1997).
10. Irwin, S. Comprehensive observational assessment: Ia. A systematic, quantitative procedure for assessing the behavioral and physiologic state of the mouse. *Psychopharmacologia* **13**, 222–257 (1968).
11. Rogers, D. C. *et al.* Use of SHIRPA and discriminant analysis to characterise marked differences in the behavioural phenotype of six inbred mouse strains. *Behav. Brain Res.* **105**, 207–217 (1999).
This study demonstrated the sensitivity of the SHIRPA screen for characterizing different mouse lines.
12. Schallert, T. & Tillerson, J. L. in *Central Nervous System Diseases: Innovative Animal Models from Lab to Clinic* (eds Emerich, D. F., Dean, R. L. & Sanberg, P. R.) 131–151 (Humana, Totowa, New Jersey, 2000).
13. Cenci, M. A. & Lundblad, M. in *Animal Models of Movement Disorders* (ed. LeDoux, M.) 193–209 (Elsevier, Amsterdam, 2005).
14. Iancu, R., Mohapel, P., Brundin, P. & Paul, G. Behavioral characterization of a unilateral 6-OHDA-lesion model of Parkinson's disease in mice. *Behav. Brain Res.* **162**, 1–10 (2005).
This paper describes the unilateral 6-OHDA model of PD and various behavioural tests that have been used to characterize the lesion effects.
15. Cenci, M. A., Lee, C. S. & Björklund, A. L-DOPA-induced dyskinesia in the rat is associated with striatal overexpression of prodynorphin- and glutamic acid decarboxylase mRNA. *Eur. J. Neurosci.* **10**, 2694–2706 (1998).
16. Steece-Collier, K. *et al.* Embryonic mesencephalic grafts increase levodopa-induced forelimb hyperkinesia in Parkinsonian rats. *Mov. Disord.* **18**, 1442–1454 (2003).
17. Lane, E. L., Winkler, C., Brundin, P. & Cenci, M. A. The impact of graft size on the development of dyskinesia following intrastriatal grafting of embryonic dopamine neurons in the rat. *Neurobiol. Dis.* **22**, 334–346 (2006).
18. Lundblad, M. *et al.* Pharmacological validation of a mouse model of L-DOPA-induced dyskinesia. *Exp. Neurol.* **194**, 66–75 (2005).
19. Santini, E. *et al.* Critical involvement of cAMP/DARPP-32 and extracellular signal-regulated protein kinase signaling in L-DOPA-induced dyskinesia. *J. Neurosci.* **27**, 6995–7005 (2007).
20. Gold, S. J. *et al.* RGS9–2 negatively modulates L-3,4-dihydroxyphenylalanine-induced dyskinesia in experimental Parkinson's disease. *J. Neurosci.* **27**, 14338–14348 (2007).
21. Glynn, D., Sizemore, R. J. & Morton, A. J. Early motor development is abnormal in complexin 1 knockout mice. *Neurobiol. Dis.* **25**, 483–495 (2007).
22. Hickey, M. A. *et al.* Extensive early motor and non-motor behavioral deficits are followed by striatal neuronal loss in knock-in Huntington's disease mice. *Neuroscience* **157**, 280–295 (2008).
23. Lin, C. H. *et al.* Neurological abnormalities in a knock-in mouse model of Huntington's disease. *Hum. Mol. Genet.* **10**, 137–144 (2001).
24. Mangiarini, L. *et al.* Exon 1 of the *HD* gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**, 493–506 (1996).
25. Barnéoud, P., Lolivier, J., Sanger, D. J., Scatton, B. & Moser, P. Quantitative motor assessment in FALS mice: a longitudinal study. *Neuroreport* **8**, 2861–2865 (1997).
26. Ma, M., Basso, D. M., Walters, P., Stokes, B. T. & Jakeman, L. B. Behavioral and histological outcomes following graded spinal cord contusion injury in the C57Bl/6 mouse. *Exp. Neurol.* **169**, 239–254 (2001).
27. de Visser, L., van den Boss, R., Kuurman, W. W., Kas, M. J. & Spruijt, B. M. Novel approach to the behavioural characterization of inbred mice: automated home cage observations. *Genes Brain Behav.* **5**, 458–466 (2006).
This paper describes how home cage monitoring systems can be used to produce a detailed analysis of mouse home cage activity.
28. Menalled, L. B. *et al.* Early motor dysfunction and striosomal distribution of huntingtin microaggregates in Huntington's disease knock-in mice. *J. Neurosci.* **22**, 8266–8276 (2002).
29. Grusser, C. & Grusser-Cornehls, U. Improvement in motor performance of Weaver mutant mice following lesions of the cerebellum. *Behav. Brain Res.* **97**, 189–194 (1998).
30. De Leonibus, E. *et al.* Spatial deficits in a mouse model of Parkinson disease. *Psychopharmacology (Berl.)* **194**, 517–525 (2007).
31. Kas, M. J., de Mooij-van Malsen, A. J., Olivier, B., Spruijt, B. M. & Van Ree, J. M. Differential genetic regulation of motor activity and anxiety-related behaviors in mice using an automated home cage task. *Behav. Neurosci.* **122**, 769–776 (2008).
32. Thifault, S., Lalonde, R., Sanon, N. & Hamet, P. Longitudinal analysis of motor activity and coordination, anxiety, and spatial learning in mice with altered blood pressure. *Brain Res.* **910**, 99–105 (2001).
33. Brooks, S., Kaur, S., Starr, B. S. & Starr, M. S. Motor actions of eliprodil in the normal and monoamine-depleted mouse: a role in the treatment of Parkinson's disease? *J. Neural Transm.* **103**, 737–748 (1996).
34. Drai, D., Kafkafi, N., Benjamini, Y., Elmer, G. & Golani, I. Rats and mice share common ethologically relevant parameters of exploratory behavior. *Behav. Brain Res.* **125**, 133–140 (2001).
35. Hickey, M. A., Gallant, K., Gross, G. G., Levine, M. S. & Chesselet, M. F. Early behavioral deficits in R6/2 mice suitable for use in preclinical drug testing. *Neurobiol. Dis.* **20**, 1–11 (2005).
36. Valentinuzzi, V. S. *et al.* Locomotor response to an open field during C57BL/56J active and inactive phases: differences dependent on conditions of illumination. *Physiol. Behav.* **69**, 269–275 (2000).
37. Van Raamsdonk, J. M. *et al.* Phenotypic abnormalities in the YAC128 mouse model of Huntington disease are penetrant on multiple genetic backgrounds and modulated by strain. *Neurobiol. Dis.* **26**, 189–200 (2007).
38. Harri, M. *et al.* Effect of access to a running wheel on behavior of C57BL/56J mice. *Lab. Anim. Sci.* **49**, 401–405 (1999).
39. Kent, S., Hurd, M. & Satinoff, E. Interactions between body temperature and wheel running over the estrous cycle in rats. *Physiol. Behav.* **49**, 1079–1084 (1991).
40. Salam, J. N. *et al.* Voluntary exercise in C57 mice is anxiolytic across several measures of anxiety. *Behav. Brain Res.* **197**, 31–40 (2008).
41. Hastings, M. H., Maywood, E. S. & Reddy, A. B. Two decades of circadian time. *J. Neuroendocrinol.* **20**, 812–819 (2008).
42. Morton, A. J. *et al.* Disintegration of the sleep-wake cycle and circadian timing in Huntington's disease. *J. Neurosci.* **25**, 157–163 (2005).
43. Dell'Omo, G. *et al.* Early behavioural changes in mice infected with BSE and scrapie: automated home cage monitoring reveals prion strain differences. *Eur. J. Neurosci.* **16**, 735–742 (2002).
44. Nielsen, D. M. *et al.* Paw preference, rotation, and dopamine function in Collins HI and LO mouse strains. *Physiol. Behav.* **61**, 525–535 (1997).
45. Nishimura, F. *et al.* Potential use of embryonic stem cells for the treatment of mouse parkinsonian models: improved behavior by transplantation of *in vitro* differentiated dopaminergic neurons from embryonic stem cells. *Stem Cells* **21**, 171–180 (2003).
46. Zhang, X., Andren, P. E., Greengard, P. & Svenningsson, P. Evidence for a role of the 5-HT_{1B} receptor and its adaptor protein, p11, in L-DOPA treatment of an animal model of Parkinsonism. *Proc. Natl Acad. Sci. USA* **105**, 2163–2168 (2008).
47. Ungerstedt, U. & Arbuthnott, G. W. Quantitative recording of rotational behaviour in rats after 6-hydroxydopamine lesions of the nigrostriatal dopamine system. *Brain Res.* **24**, 485–493 (1970).
48. Matsuya, T. *et al.* Synergistic effects of adenosine A2A antagonist and L-DOPA on rotational behaviors in 6-hydroxydopamine-induced hemi-Parkinsonian mouse model. *J. Pharmacol. Sci.* **103**, 329–332 (2007).
49. Yu, L., Schwarzschild, M. A. & Chen, J. F. Cross-sensitization between caffeine- and L-dopa-induced behaviors in hemiparkinsonian mice. *Neurosci. Lett.* **393**, 31–35 (2006).
50. Björklund, L. *et al.* Embryonic stem cells develop into functional dopaminergic neurons after transplantation in a Parkinson rat model. *Proc. Natl Acad. Sci. USA* **99**, 2344–2349 (2002).
51. Bensadoun, J. C., Mirochnitchenko, O., Inouye, M., Aebischer, P. & Zurn, A. D. Attenuation of 6-OHDA-induced neurotoxicity in glutathione peroxidase transgenic mice. *Eur. J. Neurosci.* **10**, 3231–3236 (1998).
52. Dunham, N. W. & Miya, T. S. A note on a simple apparatus for detecting neurological deficit in rats and mice. *J. Am. Pharm. Assoc.* **46**, 208–209 (1957).
53. Jones, B. J. & Roberts, D. J. A rotarod suitable for quantitative measurements of motor inco-ordination in naive mice. *Naunyn-Schmiedebergs Arch. Pharmacol.* **259**, 211 (1968).
54. Monville, C., Torres, E. M. & Dunnett, S. B. Comparison of two different protocols of the rotarod test to assess motor deficit and lesion size in the 6-OHDA model. *J. Neurosci. Methods* **158**, 219–223 (2006).
Although this study used rats, the findings are applicable to mouse models of disease. They demonstrate that different lesion types are differentially sensitive to different rotarod test protocols.
55. Wallace, J. E., Krauter, E. E. & Campbell, B. A. Motor and reflexive behavior in the aging rat. *J. Gerontol.* **35**, 364–270 (1980).
56. Kiernan, B. W. *et al.* Myelination and behaviour of tenascin-C null transgenic mice. *Eur. J. Neurosci.* **11**, 3082–3092 (1999).
57. Perry, T. A. *et al.* Cognitive and motor function in transgenic mice carrying excess copies of the 695 and 751 amino acid isoforms of the amyloid precursor protein gene. *Alz. Res.* **1**, 5–14 (1995).
58. Schallert, T., Woodlee, M. T. & Fleming, S. M. in *Pharmacology of Cerebral Ischemia* (eds Krieglstein, J. & Klumpp, S.) 201–216 (Medpharm Scientific Publishers, Stuttgart, 2002).
59. Heglund, N. C., Taylor, C. R. & McMahon, T. A. Scaling stride frequency and gait to animal size: mice to horses. *Science* **186**, 1112–1113 (1974).
60. Clarke, K. A. & Still, J. Development and consistency of gait in the mouse. *Physiol. Behav.* **73**, 159–164 (2001).
Demonstrates the use of video technology to measure several aspects of mouse gait in detail.
61. Glynn, D., Drew, C. J., Reim, K., Brose, N. & Morton, A. J. Profound ataxia in complexin I knockout mice masks a complex phenotype that includes exploratory and habituation deficits. *Hum. Mol. Genet.* **14**, 2369–2385 (2005).
62. Guillot, T. S., Asress, S. A., Richardson, J. R., Glass, J. D. & Miller, G. W. Treadmill gait analysis does not detect motor deficits in animal models of Parkinson's disease or amyotrophic lateral sclerosis. *J. Mot. Behav.* **40**, 568–577 (2008).
63. Marshall, J. F., Richardson, J. S. & Teitelbaum, P. Nigrostriatal bundle damage and the lateral hypothalamic syndrome. *J. Comp. Physiol. Psychol.* **87**, 808–830 (1974).
64. Ogawa, N., Hirose, Y., Ohara, S., Ono, T. & Watanabe, Y. A simple quantitative bradykinesia test in MPTP-treated mice. *Res. Commun. Chem. Pathol. Pharmacol.* **50**, 435–441 (1985).
65. Ohno, Y. *et al.* Evaluation of the antibradykinetic actions of 5-HT_{1A} agonists using the mouse pole test. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **32**, 1302–1307 (2008).
66. Thullier, F., Lalonde, R., Cousin, X. & Lestienne, F. Neurobehavioral evaluation of lurcher mutant mice during ontogeny. *Brain Res. Dev. Brain Res.* **100**, 22–28 (1997).
67. Crawley, J. N. *What's Wrong With My Mouse?: Behavioral Phenotyping of Transgenic and Knockout Mice* (Wiley, New York, 2000).
Probably the most informative text on behavioural phenotyping of the mouse.
68. Kashiwabuchi, N. *et al.* Impairment of motor coordination, Purkinje cell synapse formation, and cerebellar long-term depression in GluRδ2 mutant mice. *Cell* **81**, 245–252 (1995).
69. Lalonde, R., Qian, S. & Strazielle, C. Transgenic mice expressing the PS1-A246E mutation: effects on spatial learning, exploration, anxiety, and motor coordination. *Behav. Brain Res.* **138**, 271–279 (2003).
70. Lalonde, R. & Strazielle, C. Motor coordination, exploration, and spatial learning in a natural mouse mutation (*nervous*) with Purkinje cell degeneration. *Behav. Genet.* **33**, 59–66 (2003).
71. Le Cudennec, C., Faure, A., Ly, M. & Delatour, B. One-year longitudinal evaluation of sensorimotor functions in APP751SL transgenic mice. *Genes Brain Behav.* **7** (Suppl. 1), 783–791 (2008).
72. Baskin, Y. K., Dietrich, W. D. & Green, E. J. Two effective behavioral tasks for evaluating sensorimotor dysfunction following traumatic brain injury in mice. *J. Neurosci. Methods* **129**, 87–93 (2003).
73. Farr, T. D., Liu, L., Colwell, K. L., Whishaw, I. Q. & Metz, G. A. Bilateral alteration in stepping pattern after unilateral motor cortex injury: a new test strategy for analysis of skilled limb movements in neurological mouse models. *J. Neurosci. Methods* **153**, 104–113 (2006).

74. Metz, G. A. & Whishaw, I. Q. Cortical and subcortical lesions impair skilled walking in the ladder rung walking test: a new task to evaluate fore- and hindlimb stepping, placing, and co-ordination. *J. Neurosci. Methods* **115**, 169–179 (2002).
75. Cabe, P. A., Tilson, H. A., Mitchell, C. L. & Dennis, R. A simple recording grip strength device. *Pharmacol. Biochem. Behav.* **8**, 101–102 (1978).
76. Joseph, J. A. & Lippa, A. S. Reduction of motor behavioral deficits in senescent animals via chronic prolactin administration. II. Non-stereotypic behaviors. *Neurobiol. Aging* **7**, 37–40 (1986).
77. van Riezen, H. & Boersma, L. A new method for quantitative grip strength evaluation. *Eur. J. Pharmacol.* **6**, 353–356 (1969).
78. Jodogne, C., Tirelli, E., Klingbiel, P. & Legros, J. J. Oxytocin attenuates tolerance not only to the hypothermic but also to the myorelaxant and akinesic effects of ethanol in mice. *Pharmacol. Biochem. Behav.* **40**, 261–265 (1991).
79. Lorivel, T. & Hilber, P. Motor effects of delta 9 THC in cerebellar Lurcher mutant mice. *Behav. Brain Res.* **181**, 248–253 (2007).
80. Whishaw, I. Q., Sarna, J. R. & Pellis, S. M. Evidence for rodent-common and species-typical limb and digit use in eating, derived from a comparative analysis of ten rodent species. *Behav. Brain Res.* **96**, 79–91 (1998).
81. Eshkol, N. & Wachmann, A. *Movement Notation* (Weidenfeld and Nicholson, London, 1958).
82. Whishaw, I. Q., Woodward, N. C., Miklyaeva, E. & Pellis, S. M. Analysis of limb use by control rats and unilateral DA-depleted rats in the Montoya staircase test: movements, impairments and compensatory strategies. *Behav. Brain Res.* **89**, 167–177 (1997).
83. Farr, T. D. & Whishaw, I. Q. Quantitative and qualitative impairments in skilled reaching in the mouse (*Mus musculus*) after a focal motor cortex stroke. *Stroke* **33**, 1869–1875 (2002).
84. Whishaw, I. Q. & Pellis, S. M. The structure of skilled forelimb reaching in the rat: a proximally driven movement with a single distal rotatory component. *Behav. Brain Res.* **41**, 49–59 (1990).
85. Whishaw, I. Q., Pellis, S. M., Gorny, B. P. & Pellis, V. C. The impairments in reaching and the movements of compensation in rats with motor cortex lesions: an endpoint, videorecording, and movement notation analysis. *Behav. Brain Res.* **42**, 77–91 (1991).
86. Whishaw, I. Q. An endpoint, descriptive, and kinematic comparison of skilled reaching in mice (*Mus musculus*) with rats (*Rattus norvegicus*). *Behav. Brain Res.* **78**, 101–111 (1996).
87. Whishaw, I. Q., O'Connor, W. T. & Dunnett, S. B. The contributions of motor cortex, nigrostriatal dopamine and caudate-putamen to skilled forelimb use in the rat. *Brain* **109**, 805–843 (1986).
88. Wells, J. E. *et al.* Matrix metalloproteinase (MMP)-12 expression has a negative impact on sensorimotor function following intracerebral haemorrhage in mice. *Eur. J. Neurosci.* **21**, 187–196 (2005).
89. Collins, R. L. On the inheritance of handedness. I. Laterality in inbred mice. *J. Hered.* **59**, 9–12 (1968).
90. Cabib, S. *et al.* Paw preference and brain dopamine asymmetries. *Neuroscience* **64**, 427–432 (1995).
91. Baird, A. L., Meldrum, A. & Dunnett, S. B. The staircase test of skilled reaching in mice. *Brain Res. Bull.* **54**, 243–250 (2001).
Demonstration of the staircase test in mice as a method to assess fine motor control.
92. Puskovic, V. *et al.* Prolonged biologically active transgene expression driven by HSV LAP2 in brain *in vivo*. *Mol. Ther.* **10**, 67–75 (2004).
93. Starkey, M. L. *et al.* Assessing behavioural function following a pyramidotomy lesion of the corticospinal tract in adult mice. *Exp. Neurol.* **195**, 524–539 (2005).
94. Bouet, V. *et al.* Sensorimotor and cognitive deficits after transient middle cerebral artery occlusion in the mouse. *Exp. Neurol.* **203**, 555–567 (2007).
95. Paylor, R. & Crawley, J. N. Inbred strain differences in prepulse inhibition of the mouse startle response. *Psychopharmacology (Berl.)* **132**, 169–180 (1997).
This paper shows the sensitivity of prepulse inhibition across a range of inbred mouse strains.
96. Hoffman, H. S. & Searle, J. L. Acoustic variables in the modification of startle reaction in the rat. *J. Comp. Physiol. Psychol.* **60**, 53–58 (1965).
97. Grillon, C., Ameli, R., Charney, D. S., Krystal, J. & Braff, D. Startle gating deficits occur across prepulse intensities in schizophrenic patients. *Biol. Psychiatry* **32**, 939–943 (1992).
98. Perriol, M. P. *et al.* Disturbance of sensory filtering in dementia with Lewy bodies: comparison with Parkinson's disease dementia and Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **76**, 106–108 (2005).
99. Swerdlow, N. R. *et al.* Impaired prepulse inhibition of acoustic and tactile startle response in patients with Huntington's disease. *J. Neurol. Neurosurg. Psychiatry* **58**, 192–200 (1995).
100. Semenova, S., Geyer, M. A., Sutcliffe, J. G., Markou, A. & Hedlund, P. B. Inactivation of the 5-HT_7 receptor partially blocks phencyclidine-induced disruption of prepulse inhibition. *Biol. Psychiatry* **63**, 98–105 (2008).
101. Yee, B. K., Chang, D. L. & Feldon, J. The effects of dizocilpine and phencyclidine on prepulse inhibition of the acoustic startle reflex and on prepulse-elicited reactivity in C57BL6 mice. *Neuropsychopharmacology* **29**, 1865–1877 (2004).
102. Van Raamsdonk, J. M. *et al.* Cognitive dysfunction precedes neuropathology and motor abnormalities in the YAC128 mouse model of Huntington's disease. *J. Neurosci.* **25**, 4169–4180 (2005).
103. Plappert, C. F., Rodenbucher, A. M. & Pilz, P. K. Effects of sex and estrous cycle on modulation of the acoustic startle response in mice. *Physiol. Behav.* **84**, 585–594 (2005).
104. Willott, J. F. *et al.* The BALB/c mouse as an animal model for progressive sensorineural hearing loss. *Hear. Res.* **115**, 162–174 (1998).
105. Mackintosh, N. J. *The Psychology of Animal Learning* (Academic, 1974).
106. Humby, T., Laird, F. M., Davis, W. & Wilkinson, L. S. Visuospatial attentional functioning in mice: interactions between cholinergic manipulations and genotype. *Eur. J. Neurosci.* **11**, 2813–2823 (1999).
Demonstration of the five-choice reaction time task for operant motor learning in the mouse in the nine-hole box.
107. Baron, S. P. & Meltzer, L. T. Mouse strains differ under a simple schedule of operant learning. *Behav. Brain Res.* **118**, 143–152 (2001).
108. Bensadoun, J. C., Brooks, S. P. & Dunnett, S. B. Free operant and discrete trial performance of mice in the 9-hole box apparatus: validation using amphetamine and scopolamine. *Psychopharmacology (Berl.)* **174**, 396–405 (2004).
109. Derenne, A., Arsenault, M. L., Austin, D. P. & Weatherly, J. N. Weaver mutant mice exhibit long-term learning deficits under several measures of instrumental behavior. *Physiol. Behav.* **92**, 1002–1009 (2007).
110. Carli, M., Robbins, T. W., Evenden, J. L. & Everitt, B. J. Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats; implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. *Behav. Brain Res.* **9**, 361–380 (1983).
111. Cho, Y. H., Delcasso, S., Israel, A. & Jeantet, Y. A long list visuo-spatial sequential learning in mice. *Behav. Brain Res.* **179**, 152–158 (2007).
112. Christie, M. A. & Hersch, S. M. Demonstration of nondeclarative sequence learning in mice: development of an animal analog of the human serial reaction time task. *Learn. Mem.* **11**, 720–723 (2004).
113. Trueman, R. C., Brooks, S. P. & Dunnett, S. B. Implicit learning in a serial choice visual discrimination task in the operant 9-hole box by intact and striatal lesioned mice. *Behav. Brain Res.* **159**, 313–322 (2005).
114. Brooks, S. P., Trueman, R. C. & Dunnett, S. B. Striatal lesions in the mouse disrupt acquisition and retention, but not implicit learning, in the SILT procedural motor learning task. *Brain Res.* **1185**, 179–188 (2007).
115. Trueman, R. C., Brooks, S. P., Jones, L. & Dunnett, S. B. Time course of choice reaction time deficits in the $Hdh^{Q92/Q92}$ knock-in mouse model of Huntington's disease in the operant Serial Implicit Learning Task (SILT). *Behav. Brain Res.* **189**, 317–324 (2008).
This paper demonstrates the sensitivity of the SILT to motor learning in a knock-in mouse model of disease.
116. Wrenn, C. C. *et al.* Performance of galanin transgenic mice in the 5-choice serial reaction time attentional task. *Pharmacol. Biochem. Behav.* **83**, 428–440 (2006).
117. Marsden, C. D. & Parkes, J. D. Success and problems of long-term levodopa therapy in Parkinson's disease. *Lancet* **1**, 345–349 (1977).
118. Ohl, F. Animal models of anxiety. *Handb. Exp. Pharmacol.* **169**, 35–69 (2005).
119. Caligiuri, M. P. & Buitenhuys, C. Do preclinical findings of methamphetamine-induced motor abnormalities translate to an observable clinical phenotype? *Neuropsychopharmacology* **30**, 2125–2134 (2005).
120. Willner, P. The validity of animal models of depression. *Psychopharmacology (Berl.)* **83**, 1–16 (1984).
121. Lalonde, R. & Strazielle, C. Brain regions and genes affecting postural control. *Prog. Neurobiol.* **81**, 45–60 (2007).
122. Winer, B. J., Brown, D. R. & Michels, K. M. *Statistical Principles in Experimental Design* (McGraw-Hill, New York, 1991).
123. Robinson, V. Finding alternatives: an overview of the 3Rs and the use of animals in research. *Sch. Sci. Rev.* **87**, 111–114 (2005).
124. Masuya, H. *et al.* Implementation of the modified-SHIRPA protocol for screening of dominant phenotypes in a large-scale ENU mutagenesis program. *Mamm. Genome* **16**, 829–837 (2005).
125. Inoue, M. *et al.* Sensory stimulation accelerates dopamine release in the basal ganglia. *Brain Res.* **1026**, 179–184 (2004).
126. Ferdinandusse, S. *et al.* Ataxia with loss of Purkinje cells in a mouse model for Refsum disease. *Proc. Natl Acad. Sci. USA* **105**, 17712–17717 (2008).
127. Lackner, P. *et al.* Behavioural and histopathological alterations in mice with cerebral malaria. *Neuropathol. Appl. Neurobiol.* **32**, 177–188 (2006).
128. Lalonde, R., Dumont, M., Paly, E., London, J. & Strazielle, C. Characterization of hemizygous SOD1/wild-type transgenic mice with the SHIRPA primary screen and tests of sensorimotor function and anxiety. *Brain Res. Bull.* **64**, 251–258 (2004).
129. Rafael, J. A., Nitta, Y., Peters, J. & Davies, K. E. Testing of SHIRPA, a mouse phenotypic assessment protocol, on Dmd(mdx) and Dmd(mdx3cv) dystrophin-deficient mice. *Mamm. Genome* **11**, 725–728 (2000).
130. Harms, L. R., Eyles, D. W., McGrath, J. J., Mackay-Sim, A. & Burne, T. H. Developmental vitamin D deficiency alters adult behaviour in 129/SvJ and C57BL/56J mice. *Behav. Brain Res.* **187**, 343–350 (2008).

Acknowledgements
Work from the authors' laboratory has been funded by the Medical Research Council, the Biotechnology and Biological Sciences Research Council, the Cure Huntington's Disease Initiative, the Hereditary Disease Foundation and the European Union Seventh Framework programme.

Competing interests statement
The authors declare competing financial interests: see web version for details.

DATABASES
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Huntington's disease | Parkinson's disease

FURTHER INFORMATION
Stephen B. Dunnett's homepage: http://www.cf.ac.uk/biosi/staffinfo/dunnett/
MRC Harwell web page on SHIRPA: http://www.har.mrc.ac.uk/services/phenotyping/neurology/shirpa.html

SUPPLEMENTARY INFORMATION
See online article: S1 (movie) | S2 (movie) | S3 (movie) | S4 (movie) | S5 (movie) | S6 (movie) | S7 (movie) | S8 (movie) | S9 (movie) | S10 (movie)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF