

In the format provided by the authors and unedited.

Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning

Fabrice de Chaumont ^{1,6*}, Elodie Ey ^{2,6*}, Nicolas Torquet ³, Thibault Lagache¹,
Stéphane Dallongeville¹, Albane Imbert⁴, Thierry Legou⁵, Anne-Marie Le Sourd², Philippe Faure ^{3,7},
Thomas Bourgeron^{2,7} and Jean-Christophe Olivo-Marin ^{1,7*}

¹Institut Pasteur, BioImage Analysis Unit, CNRS UMR 3691, Paris, France. ²Human Genetics and Cognitive Functions, Institut Pasteur, UMR 3571 CNRS, University Paris-Diderot, Paris, France. ³Sorbonne Université, CNRS UMR 8246, INSERM, Neurosciences Paris Seine - Institut de Biologie Paris-Seine, Paris, France. ⁴Institut Pasteur, FabLab, Center for Innovation and Technological research, Paris, France. ⁵Aix-Marseille Université, CNRS, LPL, UMR 7309, Aix-en-Provence, France. ⁶These authors contributed equally: Fabrice de Chaumont, Elodie Ey. ⁷These authors jointly supervised this work: Philippe Faure, Thomas Bourgeron, Jean-Christophe Olivo-Marin. *e-mail: fabrice.de-chaumont@pasteur.fr; elodie.ey@pasteur.fr; jean-christophe.olivo-marin@pasteur.fr

Supplementary information – Table of content

Title	Page
Motivation and review of the existing tracking methods	2
Global tracking diagram	5
Setup calibration	6
Manual validation (extra information)	6
Automatic tracking quality control	9
Automatic tracking quality control based on RFID reading	9
Automatic tracking quality control based on detection	10
Social contact timelines	12
Collaborative website sharing data	13
Software Development Kit	13
Database description	14
Querying database information with Python	14
UDP live network information stream	14
MP4 video recording	14
Database player	15
Behavioural event extraction	15
Dynamics of subgroups of three mice within groups of four mice	18
Software and system general description	18
Supplementary results	20
Sex-related variations in the behaviour of C57BL/6J adult mice	28
Supplementary references	34

Supplementary films, source code and PDF

Films:	
1 - Live Mouse Tracker - Method overview	19
2 - Live Mouse Tracker - Hardware Blueprint and assembly instructions	19
3 - Live Mouse Tracker - Investing an abnormal nest building behaviour	19
4 - Live Mouse Tracker - Background height-map demo	19
5 - Live Mouse Tracker - Live 3D rendering demo	19
6 - Live Mouse Tracker - Validation	19
7 - Live Mouse Tracker - Tracking demo on mice having different coat colours	19
Source code:	
Live Mouse Tracker – Java source code v466	
Available at http://icy.bioimageanalysis.org/plugin/Live_Mouse_Tracker	
PDF:	
Assembly instructions	Separated file

Motivation and review of the existing tracking methods

We present here an overview of the current tracking methods used to follow mice and to extract behavioural information. We selected here the most representative ones of each type of methods used. We first present the current needs in phenotyping studies and then we detail to what extent the different methods answer these needs.

Phenotyping mouse models usually involves one individual screened through various behavioural tests to measure anxiety, learning, memory, or locomotion. Social aspects are determined in dyadic interactions, but these aspects are the least examined and remain to be complemented in existing phenotyping databases^{1,2}. With the aim of documenting in the most comprehensive way the phenotype of an individual, studying social aspects within a group of animals would allow to document both dyadic interactions and more complex ones involving more than one other individual. These two aspects of sociality can therefore be dissociated and will enrich phenotyping data³⁻⁷. Such approaches will increase the translational values of the phenotyping studies^{8,9}.

In clinical examination of patients with neuropsychiatric disorders affecting social relationships, both the initiation and the maintenance of these relationships are evaluated (ADI-ADOS for autism spectrum disorders). Therefore in mouse models for these disorders, we should also take into account these aspects by following the freely moving animals over longer durations⁹. Indeed, short experiments have been shown to be misleading in some cases¹⁰. New experiments should be built to gather precise data on individual mice over the long term to answer the more and more precise needs of phenotyping with ethologically-relevant measures^{1,2}.

Individual following is necessary to identify individual profiles. This is required to model personalized medicine. However, visual or olfactory marks could lead to modifications of the general behaviour of the animals (as in zebrafishes¹¹), biasing the results of the experiment. Therefore, individual following should be conducted without perceptible marks on the animals. To our knowledge, Radio Frequency IDentification (RFID) represents the most inconspicuous identity marker current available (particularly since they are getting smaller), once the subcutaneous implantation is done. Furthermore, RFID probes have the advantage of being a standard in some animal facilities where animals are systematically implanted.

To summarize these requirements, the current need in phenotyping is to measure automatically in a detailed way social and non-social behaviours of individuals within a group. The description of complex behaviours requires a high level of details. An attempt was already made to reach a high level of details in body postures with a depth sensor camera but this was used on an isolated mouse in an openfield¹². A high level of details in social contexts is currently reached when only two individuals are tested together over short durations (i.e., 10-15 min). The way such data is collected does not allow users to investigate more individuals simultaneously or over a longer duration. de Chaumont and colleagues (2012) published a semi-automatic software to describe precisely the types of interactions and the sequences of social events occurring between two visually indistinguishable mice, from a top-view video camera¹³, but this solution is limited as it needs to be supervised and corrected by an expert. Unger and colleagues (2017) built on MiceProfiler and used another segmentation method to improve detection and avoid manual corrections¹⁴ (manual interventions remained necessary at the beginning of the tracking). Hong and colleagues (2013) developed a system with two video cameras and one depth sensor to follow two mice¹⁵, but they needed to be of different coat colours to track the identities. The data extracted were detailed (attack, mounting, distances head to body and body to body), but they were calculated based on annotated video

frames and training of a classifier. Finally, the highest level of details was provided by Matsumoto and colleagues (2013). They developed a system using four depth sensors to follow in details the different steps of mating in rats in full 3D over ten minutes¹⁶. It only requires manual intervention to define the initial position of the models. Giancardo and colleagues (2013) have developed a system able to track unmarked mice over one hour¹⁷. The behavioural repertoire measured was large (types of contacts, above, following, non-social behaviours such as walking or standing alone), but the system needed an average of 1 manual intervention to correct identities every 34 s (in a 3-mice experiment). Overall, all these systems need manual interventions to provide the most relevant data. This is manageable over short experiments up to one hour, but not over longer experiments of several hours or days.

For such long-term experiments, the possibilities to track individually animals in groups are two folds. First, the animals are marked visually, using hair dye. These visual marks can be combined with radio frequency identification (RFID) tags or not. So et al. (2015) manually scan sampled over 21 days chasing, fighting, sniffing and allo-grooming³. This approach provided impressive data on hierarchy and social relationships within the group, but it is not manageable on long-term experiments in mass phenotyping. Similarly, in the Visible Burrow System¹⁸, Bove and colleagues (2018) manually annotated six 10-min periods per day over five days of a group of eight mice, either from the C57BL/6J strain or from the BTBR strain and highlighted a reduced level of social activity and an increased level of grooming, but no difference in environmental exploration, social avoidance or aggression (except on the first day)⁸. Howerton et al. (2012) scored the activity of two mice and their grouping within compartments automatically over 24 hours¹⁹. They used the reads of RFID tags when mice travelled through antennas and combined data with visual validation using hair dye. Similarly, Shemesh et al. (2013) proposed a system providing the activity as well as the spatial occupancy of the animals marked visually, whose identity was learned over labelled video frames⁴. Castelhana-Carlos et al. (2014) conducted the same kind of activity and position data on rats identified by RFID tags and pen marks on the tail²⁰. Interestingly, they also enriched their data with social behaviours that were annotated manually. The most advanced system in this category with animals visually marked is the system of Ohayon and colleagues (2013). They provided data over five days on the place preference of the mice, as well as their social grouping and following behaviours (based on pre-labelled videos)⁷. Altogether, these methods using visual marks to track animals (verified through RFID tag scanning or not) provided most of the time activity and place preference data. They are based on contrast and need specific fixed experimental conditions, obtained through background simplification or infrared lightning. Additional social behaviours needed to be manually annotated, except in the system of Ohayon et al. (2013)⁷.

For this second set of methods, the animals are not visually marked. In most of these systems, the animals are identified using RFID tags, and the only data extracted are on the activity and the spatial grouping over different compartments/zones of the cage^{6,21–23}. This method provides the longest experiment possible, but with no precise social data. Perez-Escudero and colleagues (2014) developed a very interesting and inspiring method to identify each individual based on a visual fingerprint²⁴, that requires neither visual nor RFID markers. This digital signature is coded by associating contrast information and distances within the animal. This system, idTracker, works very nicely with fishes and ants, as these animals are geometrically incompressible: indeed, even if they visually bend, they still always display the same surface. Mice in contrast are highly deformable, able to stretch or contract with a large spectrum of different conformations, eventually displaying different fur appearance, and, in

addition, they can stand up (known as rearing) and present yet another set of visual conformations. This makes idTracker unable to keep the identity of mice because the geometrical/digital signature used cannot deal with such a deformable model. Also, the level of details provided by this method is not sufficient to characterize globally the social behaviour of mouse models of neuropsychiatric disorders. With their system, Perez-Escudero and colleagues were able to extract spatial occupancy, velocity or relative position of the animals within a group of zebrafishes, but not robust data on mice. The most recent version of this system (idtracker.ai²⁵) outperforms the previous tracking of zebra fishes and claims to achieve 99.9% of detection accuracy with almost no mis-identification for mice. A video example is provided by the authors at <https://www.youtube.com/watch?v=ANsThSPgBFM>. In this video of a duration of 1 hour, 4 black animals are in a cage, and the system provide an (x,y) position (but not a full contour) for each animal. Even though the animals are not often in contact during this recording, there are enough of them to assess how the idTracker manages to resolve contact situations. It can be observed that the tracking performs very nicely to recover identities of animals before and after contacts, but that during contacts, the system loses the individual positions of the animals and uses linear interpolation to create estimated positions during the contacts (this can be verified at the following time points: 1:08, 2:23, 2:30, 2:36, 2:38, 3:06, 3:55). Moreover, it can be noticed that idTracker might swap identities (for example at 2:36, 3:07) or fall when animals change appearance in their fur or hair. In the later case, this requires a human operator (see at 25:54) to re-brush them and allow the system to detect animals again. Note that the previous idTracker version provided equivalent video of tracking, and did not interpolate results: <https://www.youtube.com/watch?v=yZxepx43h5Q>

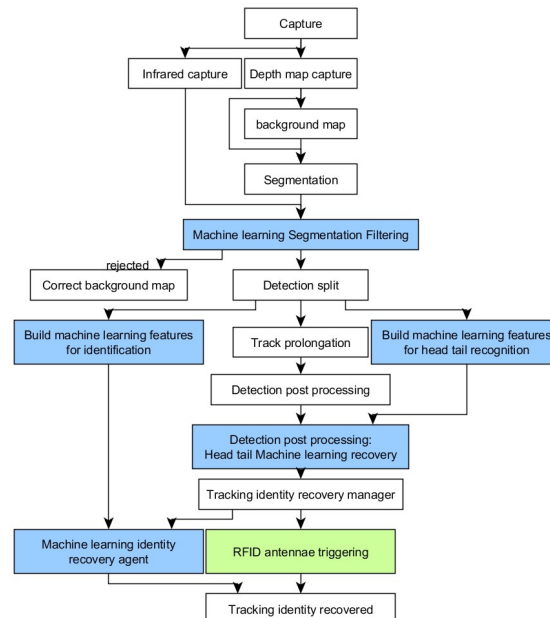
Mathis and colleagues proposed an open source software, DeepLabCut, that uses a convolutional deep learning to track several points selected by the user on the animals²⁶. For each desired point, the user manually annotates 1080 frames to train the system. This training can be performed on one or several animal(s), and then be applied to unknown animals. DeepLabCut deals with the recognition of specific parts of the animals. DeepLabCut community provides video where they do track mice: <https://twitter.com/GoldenNeuron/status/1097609848525484032>. In this one, 2 mice (one white, one black) are reliably tracked. We did not find example or project on long-term observation of 4 black mice with DeepLabCut yet. DeepLabCut is a very impressive and versatile system as it adapts to about any animal used in behavioural studies (fly, mice, zebra fishes, horses and also tracking of sub-region such as forepaw fingers). Another positive aspect is that it can potentially be real time to enable close loop, but this was not demonstrated. Second positive point is that it worked with existing footage, without specific camera or lighting. DeepLabCut provides the location of user-defined points, but an extra development is mandatory to interpret the behaviour related to the location of the points. Weissbrod and colleagues (2013) combined RFID identification with video recordings and provided social data reflecting relative movements, with a reduced behavioural repertoire but no robust data on the contacts between animals⁵. Alexandrov and colleagues (2015) proposed a system that can supposedly provide activity and contact data between non-visually marked animals but no information on the method used and on the validity of the tracking is available²⁷. Finally, Clever Sys Inc. proposed the RFID-Assisted SocialScan system that combines video tracking and RFID identification for tracking mice living in groups over several days, with a supposed large events of individual and social events automatically scored. The RFID allows to recover identities if they are lost, since animals are not visually marked. No validation of the tracking is provided and the sample video available (<https://www.youtube.com/watch?v=j5mBLG95Q1c>)

suggests that identities do swap very often, even in simple situations and are recovered only after a long delay, resulting in a high error probability in mouse identification (no validation method and no error rate were provided).

RFID tags are the most robust way to identify individuals but they have their own constraints in reading speed and in probe range. The RFID tag should be: 1) in the probe range for 1/10th s; 2) the only tag in this range; and 3) read by only one antenna, meaning only one antenna should be activated at a time. These constraints on RFID explain why their usage is not trivial, for instance, RFID floor, sold by the TSE company (TSE Systems GmbH European / Asian headquarters, Bad Homburg, Germany), cycles blindly over the antenna to read RFID, therefore the probability of reading an animal is reduced again by the number of antenna integrated in this device.

All these methods have their own advantages, but none of them provide the possibility of reliably track mice individually within a group over long-term experiments with the level of details that is requested for phenotyping mouse models of neuropsychiatric disorders. In addition, none of these systems is real time and therefore there is no possibility to directly interact with the experiment, despite the fact that several groups already asked for this feature^{15,16}. This would allow to increase phenotyping speed¹⁹, and reject all possibility of manually correcting of the data. Tracking errors should therefore be managed directly at the source, and not corrected afterwards. As fulfilling these needs with the current methods of animal tracking is not possible, we developed our own system working in real time and managing identification errors.

Global tracking diagram



Supplementary Figure S1 | Tracking diagram

Setup calibration

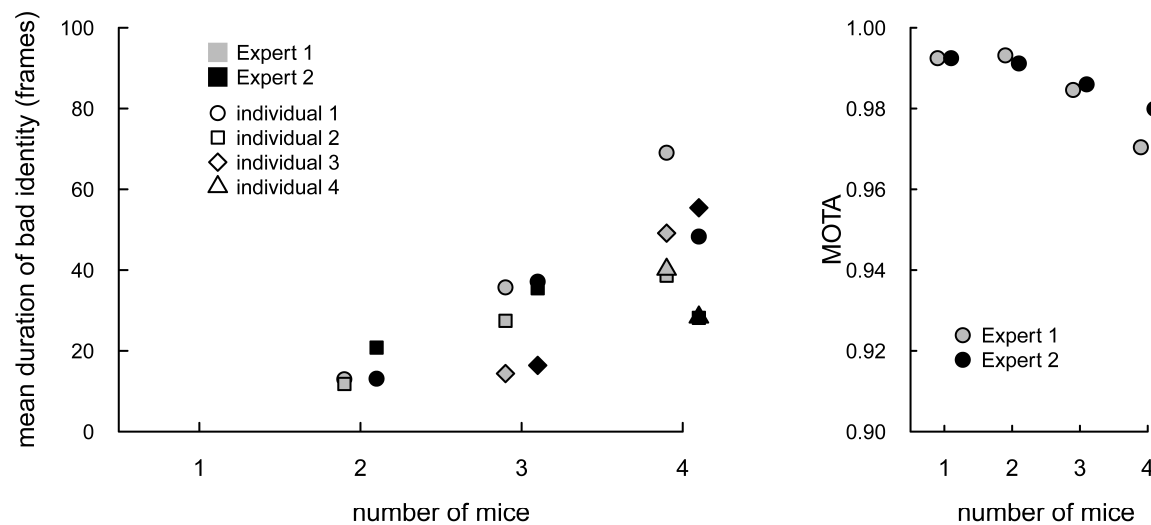


Supplementary figure S2 | Calibration of the setup. (a) The notice from the assembly instruction which indicates that Kinect should be perpendicular to the ground. (b) 3D image from blueprints. (c-d) View of the calibration software. All coloured dots display live the distance to the ground so that the user has just to move the Kinect by hand to watch if it is properly positioned. Red, orange, and green dots indicates respectively very bad, bad and correct positioning.

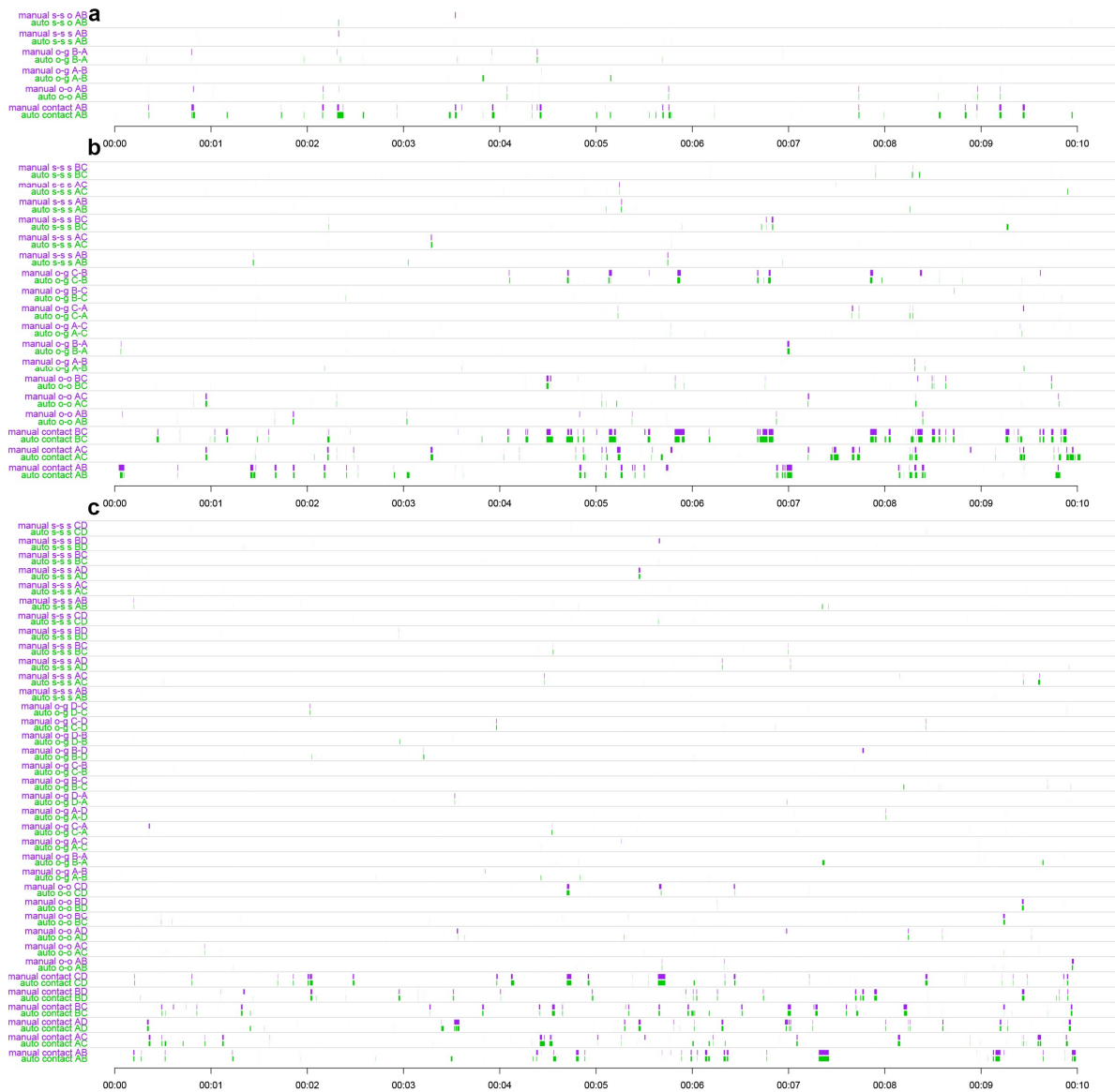
Note: The kinect v2 is now discontinued. Millions of items have been sold so this is not a rare product. Nevertheless, the method can be applied with other RGBD cameras if the driver to connect to the Java core program is available.

Manual validation (extra information)

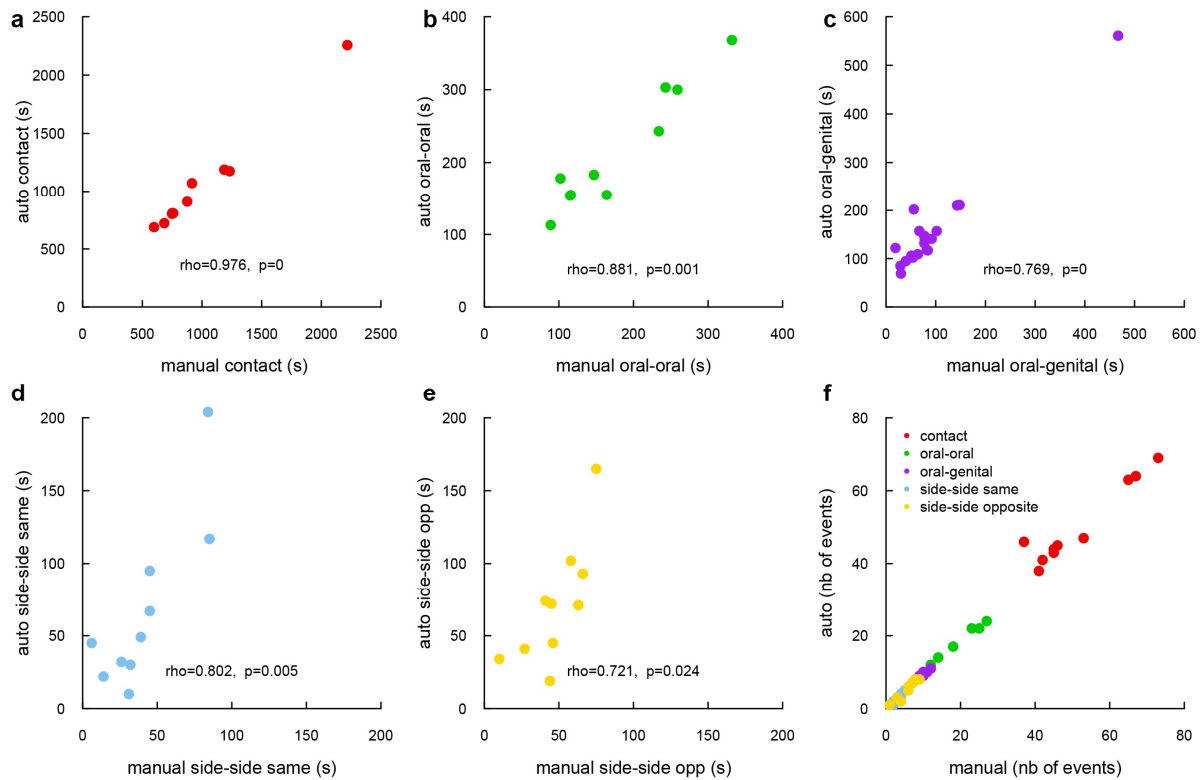
In the following figure (**Supplementary Fig. S3**), we detail the distribution of the mean duration of error and the representation of the MOTA index²⁸ depending on the expert. A video example of the tracking in operation is provided in the **Supplementary movie: LMT – Validation**.



Supplementary Fig. S3 | Manual validation of the tracking system. (a) Mean number of consecutive frames where the identity of a mouse is not correct. (b) Calculation of Multiple Object Tracking Accuracy (MOTA): $MOTA = 1 - \frac{\sum(t) (false\ negative(t) + false\ positive(t) + identity\ switch(t))}{\sum(t) (nb\ mice\ in\ the\ ground\ truth)}$, where t is a frame, a false negative is a frame where a mouse is not detected but it is there, a false positive is a frame where a mouse is detected but it is not there, an identity switch is a frame where identities are switched between individuals, and the number of mice in the ground truth is the number of mice that should be tracked in each frame.



Supplementary Fig. S4 | Timelines of manual validation of the tracking system (10 minutes). Each consecutive couple of row displays the manual validation (violet) against the same behaviour labelled automatically (green) over 10 minutes. (a) Experiment with 2 individuals. (b) 3 individuals. (c) 4 individuals.



Supplementary Fig. S5 | Manual validation of the tracking system. (a-e) Correlation between automatic and manual measurements for the different types of contact: (a) general contact, (b) oral-oral contact, (c) oral-genital contact, (d) side by side contact in the same direction and (e) side by side contact in the opposite direction (Spearman's rank correlation). (f) Number of events that were detected automatically as a function of the number of manual events detected by the experimenter.

The **Supplementary Fig. S5** displays the correlation (in total length of event found) between the manual and the automatic labelling on five different contact events. The **Supplementary Fig. S5a-e** displays linear correlation. For the time spent in contact, the time spent in oral-oral contact and the time spent in oral-genital contact, the manual and the automatic results display a significant positive correlation and therefore match. For the side-by-side contact events, the automatic method tends to find longer events than the manual method, but the relation between the automatic measure and manual measure is still linear. After further inspection, we realized that our expert was a little more restrictive on its criteria for side-by-side contact events. Indeed, the expert expected the overlap between the sides of the two animals to be larger than what the system evaluates, and discarded the frames in which the animals were at a whisker distance from one another, meaning that the system counted side-by-side contacts in slightly looser situations. This explains the reduced duration of manually measured events in comparison with automatic measures. We checked this assumption by comparing the two measures at the level of the event. For each manually labelled event, we searched for an event in the automatic labelling. The **Supplementary Fig. S5f** shows that very few events were missed by the automatic labelling, which confirms our explanation.

Automatic tracking quality control

Live Mouse Tracker records events in the database of the experiment to monitor the tracking. Five types of events are stored: 1. If the machine learning dedicated to the identification of tracks performs an association, an event “MACHINE LEARNING ASSOCIATION” is logged. 2. For the RFID, we log three events. a) If an anonymous track is assigned thanks to RFID, then “RFID ASSIGN ANONYMOUS TRACK” is triggered. b) If a RFID read corrects the identity of a track, then the “RFID MISMATCH” is triggered (and the track is corrected). c) Finally when all animals are identified, the RFID are used to control the identity of animals. In that case, at each reading confirming current track identity, a “RFID MATCH” is stored in database. 3. We can also monitor if an animal is not detected at a given frame. This happens when animals are hiding, or when they become indistinguishable, for instance in the nest condition. In this case, the event “DETECTED” is not triggered for the given animal. Those events (built automatically) allow building of several measures (online or in post-processing) to assess the quality of the tracking. Here we provide two different measurements:

Automatic tracking quality control based on RFID reading:

We extensively used the RFID technology to assess the reliability of the identity. In our experiments, the identity of each animal is checked through the RFID antennas 49 times per minute on average (from 19.7 to 100.9) (see **Supplementary Table 1**). At each RFID read, the identity of the animal is checked and corrected if necessary (the correction propagates back in time). In the table hereafter, the number of identity match over the total number of reads provides **98.47% ± 0.16 (mean ± SEM)** of success. The following table details the score for all animal of all 23h-experiments (**Supplementary Table 1**).

Supplementary table 1 - RFID check quality for 12H (all Shank2 mice)

Animal's RFID	Number of RFID reading	Number of identity match.	Number of identity mismatch.	Average number of RFID control per minute	Identity match rate
4064913	29387	28798	589	40,8	97,99
4064643	51857	51500	357	72,0	99,31
4064920	71911	71582	329	99,8	99,54
4065030	41690	41180	510	57,9	98,77
4064840	42066	41611	455	58,4	98,91
4064719	39505	39069	436	54,8	98,89
4064579	32602	32156	446	45,2	98,63
4064967	27319	26807	512	37,9	98,12
4064691	31278	30948	330	43,4	98,94
4064947	72662	72439	223	100,9	99,69
4065039	32261	31873	388	44,8	98,79
4064959	35075	34815	260	48,7	99,25
4064786	40895	40469	426	56,7	98,95
4064677	50371	50018	353	69,9	99,29
4064956	37645	37218	427	52,2	98,86
4064870	25437	24971	466	35,3	98,16
4064800	31164	30797	367	43,2	98,82
4064761	22723	22299	424	31,5	98,13

4064879	28768	28495	273	39,9	99,05
4064771	27424	27103	321	38,0	98,82
4064844	25818	25527	291	35,8	98,87
4064722	29560	29296	264	41,0	99,10
4064807	24824	24478	346	34,4	98,60
4064648	54120	53902	218	75,1	99,59
4065006	22632	22328	304	31,4	98,65
4064597	40729	40408	321	56,5	99,21
4395614	20725	20309	416	28,7	97,99
4395891	14185	13694	491	19,7	96,53
4064715	22679	22179	500	31,49	97,79
4396015	50673	50219	454	70,37	99,10
4395839	27279	26843	436	37,88	98,40
4395766	45341	44997	344	62,97	99,24
4395943	28466	28142	324	39,53	98,86
4395748	52810	52459	351	73,34	99,33
4396009	30079	29781	298	41,77	99,00
4064696	8673	8148	525	12,04	93,94
4395909	27077	26243	834	37,60	96,91
4395685	28969	28093	876	40,23	96,97
4395945	37733	37002	731	52,40	98,06
4395810	42675	42164	511	59,27	98,80
4395718	43721	43280	441	60,72	98,99
4395573	26722	26137	585	37,11	97,81
4395736	18510	17943	567	25,70	96,93
4395612	56164	55858	306	78,00	99,45
Average:				49,83	98,47

Automatic tracking quality control based on detection:

The second quality metric considers the number of frames the animal is detected over the total number of frames of the experiment (as used in validation) (**Supplementary Table 2**).

In our experiment, we observe animals in housing conditions. They cannot escape the observation. So we observe different group activity that can be condensed in 2 phases: activity or sleep. During sleep, the animals are all together in a nest for long periods, and may be covered by cotton. We assume that our system cannot track individuals during those phases. Still, we detect those events and we label such periods as “nesting”.

Therefore, the quality measurement during activity provides **92.91% ± 0.48 (mean ± SEM)** of success.

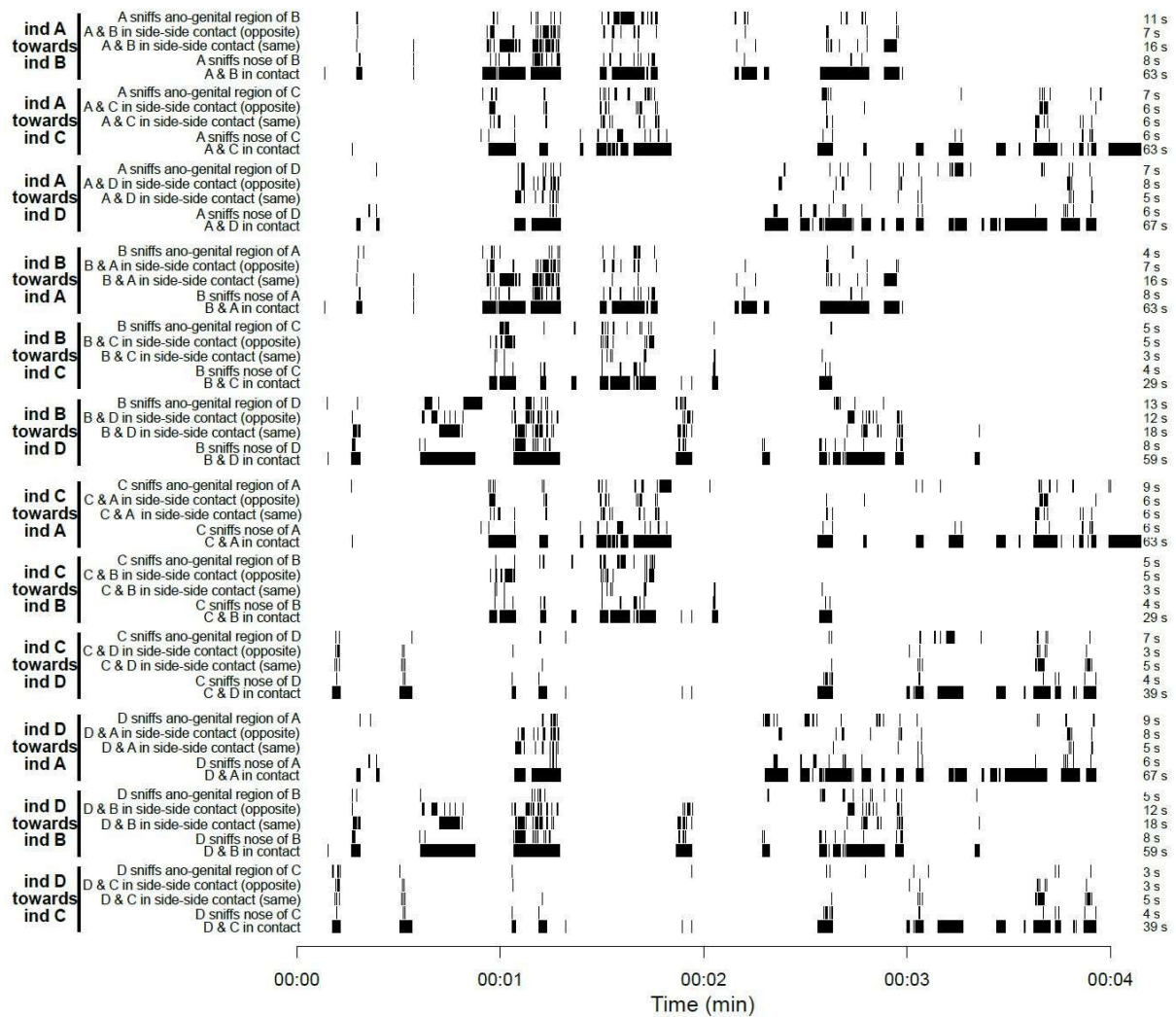
Supplementary Table 2 - Detection check quality for 12H (all Shank2 mice)

Animal's RFID	Total number of frame in activity	Total number of frame where animal is identified (in activity)	Detection ratio
4064913	1251999	1158732	92,55

4064643	1277002	1258198	98,52
4064920	1263583	1226301	97,04
4065030	1255989	1150209	91,57
4064840	1194591	1112645	93,14
4064719	1217785	1155538	94,88
4064579	1280143	1243262	97,11
4064967	1205802	1117508	92,67
4064691	1217524	1166273	95,79
4064947	1184428	1133754	95,72
4065039	1185552	1107423	93,40
4064959	1227086	1194256	97,32
4064786	1105668	1003393	90,74
4064677	1136796	1070961	94,20
4064956	1065620	969166	90,94
4064870	1065315	970214	91,07
4064800	1150021	1080028	93,91
4064761	1143709	1072717	93,79
4064879	1261700	1225298	97,11
4064771	1191558	1125411	94,44
4064715	1272050	1224720	96,27
4396015	1224397	1159482	94,69
4395839	1208929	1153056	95,37
4395766	1206880	1127104	93,38
4395943	1226894	1080982	88,10
4395748	1249159	1118764	89,56
4396009	1272463	1172354	92,13
4064696	1219752	1050078	86,08
4395909	1230852	1138858	92,52
4395685	1211635	1117577	92,23
4395945	1225521	1141087	93,11
4395810	1258067	1195796	95,05
4395718	1258095	1147781	91,23
4395573	1232080	1086625	88,19
4395736	1253415	1144180	91,28
4395612	1262161	1202839	95,29
4065006	1070942	952940	88,98
4064597	1177829	1051287	89,25
4395614	1153448	1056107	91,56
4395891	1090421	938421	86,06
Average:			92,91

Social contact timelines

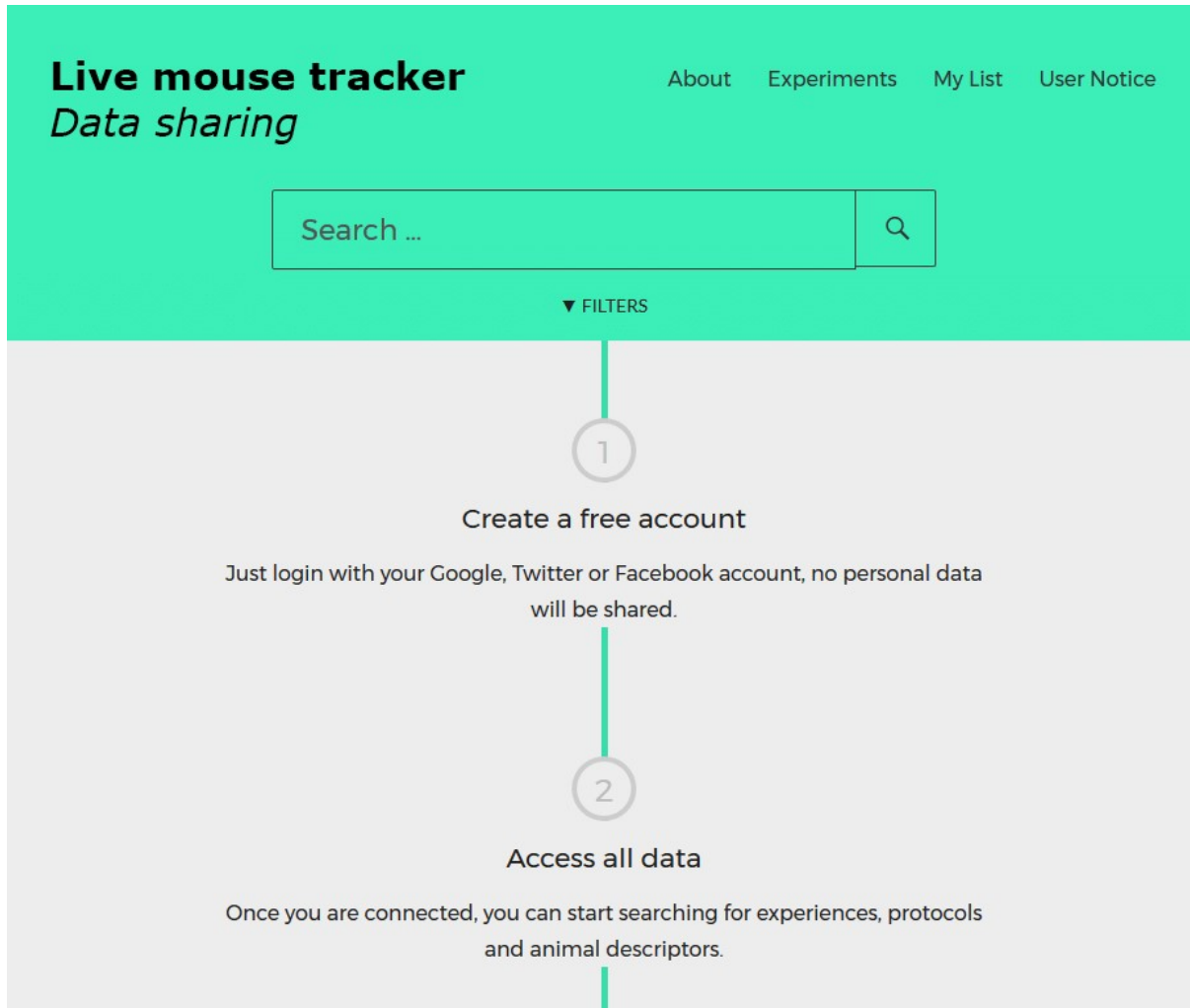
In the following figure (Supplementary Fig. S6), we provide additional timelines for one experiment of 4 individuals.



Supplementary Fig. S6 | Example of the distribution of the different types of contacts between each possible pair of individuals within a group of four mice.

Collaborative website sharing data

The website is opened to anybody (**Supplementary Fig. S7**). Users first store their data in the location they prefer, and then they reference their datasets within this website. People can then search for experiment and download databases, films, script to process data and any material users want to share.



Supplementary Figure S7 | Opening view of the website <http://livemousetracker.org> designed to share experiments.

Software Development Kit

This work represents two years of work dedicated to the development of this complete solution. A large part of the solution consists in connecting to driver, grabbing stream, ensuring that streams are synchronized, dealing with java real time consideration, creating code probe to help debugging, creating display overlay, creating streamed data structure, dealing with calibration, output of data as video stream, database or network stream for immediate control of third party devices. For developing purposes, a number of side tools have been created, that are only used by the developers, not by the users. They consist in small programs that aim at helping the developers to debug code and test new features. Those tools are also provided in open source to build on top of the existing tracking. Among those tools, developers

will find a raw data recorder and player, which mimics a live kinect device, but allows to play several times the exact same film.

A large part of those developments is not to be said “scientific”, but it is mandatory to build a final working product. We documented all this trunk of code so that newcomers can test their own techniques directly in this framework, in order to focus their energy on the problem of enhancing the phenotyping instead of working on all the side engineering problems.

Finally, to ease distribution of new versions, Live Mouse Tracker is developed as a plugin of Icy, which makes it easy to fork and further distributions.

Database description

All the data are stored locally on a simple standalone database, in a redundant manner to target different levels of query, which complies with the Big Data principles²⁹. This means that the database can be accessed by any third party software such as R, Matlab, Python or Java and that the user can query the database on simple or complex features. No server is needed to store or query databases as the data are shared between experimenters by simple file transfer (contrary to dedicated database servers that would lock-in the data into labs).

The software creates the database for the current experiment automatically as the tracking starts. We choose the SQLite database format as it is server free. The whole database is a single file, and each database represents an experiment, making it easy to backup, share, query and transform for users. Tables are Animal(Id, RFID, Genotype, Name), RfidEvent(Id, RFID, Time, X, Y), Frame(FrameNumber, TimeStamp, NumParticle, Paused), Event(Id, Name, Description, StartFrame, EndFrame, IdAnimalA, IdAnimalB), Detection(Id, FrameNumber, AnimalId, Mass_X, Mass_Y, Mass_Z, Front_X, Front_Y, Front_Z, Back_X, Back_Y, Back_Z, Rearing, Lool_Up, Look_Down, Data). The data field is an XML text containing the full mask of the animal in zip format. To perform fast SQL Queries, we advise to add indexes to the database (Provided in **Supplementary SQL script**). Database can be explored with the software DB Browser for SQLite (<http://sqlitebrowser.org>). We also provide scripts and examples to create all the figures and query shown in this paper from the software R (<https://www.r-project.org/>). Code: *experiment.Experiment.createDataBase*.

Querying database information with Python

Scripts are available on GitHub at <https://github.com/fdechaumont/lmt-analysis>.

UDP live network information stream

To perform closed-loop control over any device, we deliver live tracking information, streamed on network. To avoid network packet control, we choose the UDP protocol, which is efficient for continuous data streaming. Streaming message consists in location and direction of all animals present in the field.

MP4 video recording

To be able to go back to visual data for control, we record the infrared stream in MPEG. We add overlay information to display tracking information, RFID state, time of acquisition, memory load, CPU time computation for the last frame and real time/date. MPEG is recorded in full infrared resolution (512 x 424) at half the input rate by default (15 fps), customizable by user. MPEG is recorded in its own thread in low priority, meaning that in weak configuration, frames will be dropped from record. For storage and browsing convenience, each MPEG is 10 minutes long. Code: *package livemousetracker.MPEGRecorder*.

Database player

The database player allows to read database by replaying the data extracted from the software. Its GUI displays color-coded animal segmentation and the event corresponding to the current time points. Events can be displayed, no matter if they were computed from the original tracker, or subsequently in R, matlab, python or any software. If events are related to only one animal, the segmentation carries those descriptions. Otherwise, a column showing all interaction is colour coded to better understand which mice are involved. Events can be filtered to help studying specific events. Users can seek forward or backward in the database with different time-steps, or change the frame rate. Code: `package livemousetracker.dataplayer`.
Supplementary movie: LMT - Method overview.

Behavioural event extraction

We extracted individual, dyadic, configuration, dynamic as well as making / breaking group events (see examples in **Fig. 2**). The following list describes the events that were calculated. Events preceded by a star were used for the computation of the individual profiles.

<i>Individual events</i>	
*move alone	the mouse is moving at a speed higher than 5 pixels per frame and not in contact with any other mouse.
*stop alone	the mouse is stopped (speed lower than 5 pixels per frame) and not in contact with any other mouse.
*rear iso	the mouse is rearing (i.e. the height difference between the tail basis and the nose point is higher than 40 mm) and not in contact with any other mouse.
head up	the mouse is sniffing the air, nose up. (Supplementary methods - Head subpart processing)
*head down	the mouse is sniffing the ground. (Supplementary methods - Head subpart processing)
*huddled	the mouse is stopped (speed less than 5 pixels per frame) and huddling, with a ball shape from a top view (circularity>0.75).
*jump	the mouse is jumping against the walls of the cage (Oscillative position detection)
*SAP	the mouse is in stretched attend posture, i.e., the mouse is moving at a speed lower than 5 pixels per frame, its body length longer than the mean body length plus one standard deviation of body length and the height of the centre of mass is lower than the median height of the centre of mass.

Social dyadic events	
*contact	the mouse is in contact (i.e. the two masks have one common pixel) with any other mouse.
*move contact	the mouse is moving at a speed higher than 5 pixels per frame and in contact with another mouse.
stop in contact	the mouse is stopped (see above) and in contact with another mouse.
*rear contact:	the mouse is rearing (i.e. the height difference between the tail basis and the nose point is higher than 40 mm) and in contact with another mouse.
*contact side-side	the side of a mouse is within 30 pixels (52 mm) from the side of another mouse; the two animals are oriented in the same direction.
*contact side-side opp	the side of a mouse is within 30 pixels (52 mm) from the side of another mouse; the two animals are oriented in opposite directions.
*contact nose-nose	the mouse is sniffing the oral region of another mouse, i.e. the two nose points are less than 15 pixels (26 mm) from one another.
*contact nose-anogenital	the mouse is sniffing the ano-genital region of another mouse, i.e. the nose point of the first mouse is within 15 pixels (26 mm) from the tail point of the other mouse.

Configuration events	
*group2	the mouse is in contact with one and only one other mouse.
*group3	the mouse is in contact with two and only two other mice to form a group of three mice.
group4	the mouse is in contact with three other mice to form a group of four mice.
*out of nest	the mouse is isolated from the three other mice, which are stopped and form a nest (nest3).
nest4	the four mice are stopped and form a nest together.
*train2	the mouse is following another mouse (speed higher than 5 pixels per frame) while sniffing her ano-genital region.
*train3	the mouse is following two other mice in train2 (speed higher than 5 pixels per frame) while sniffing the ano-genital region of the last one.
*train4	the mouse is following three other mice in train3 (speed higher than 5 pixels per frame) while sniffing the ano-genital region of the last one.

<i>Dyadic dynamic events</i>	
*social approach	the mouse is approaching another one, i.e. the distance between the two animals shortens, the speed of the mouse is higher than the speed of the approached mouse, and the distance between the two animals is shorter than two mean body lengths (of the approached animal). This approach does not necessarily lead to a contact.
*approach reared mouse	the mouse is approaching another one, i.e. the distance between the two animals shortens, the speed of the mouse is higher than the speed of the approached mouse, and the distance between the two animals is shorter than two mean body lengths (of the approached animal). The approached mouse is rearing, i.e. the slope between the tail basis and the nose point. This approach does not necessarily lead to a contact.
*make contact	the mouse is approaching another one to make contact with it. The number and the duration of these approaches are computed.
*escape from social range	the mouse escapes from another mouse, i.e. its speed is higher than the speed of the escaped mouse, the distance between the two mice increases and the distance between the two animals is shorter than two mean body lengths (of the animal that the mouse is escaping from).
*break contact	the mouse breaks a contact with another mouse and moves away from this mouse.
*follow	the mouse is walking directly in the path of another mouse. Both mice move at a speed higher than 5 pixels per frame, with an angle between their direction vectors smaller than 45°, with the distance between their centres of mass shorter than two mean body lengths.

<i>Group making/breaking events</i>	
*make group3	the mouse is joining a group of two mice to form a group of three mice in contact.
*make group4	the mouse is joining a group of three mice to form a group of four mice in contact.
*break group3	the mouse is leaving a group of three mice.
*break group4	the mouse is leaving a group of four mice.
*seq o-o o-g:	the mouse is first doing an oral-oral contact with another mouse and then going to sniff the ano-genital region of this mouse.
*seq o-g o-o	the mouse is first sniffing the ano-genital region of another mouse and then establishes oral-oral contact with this mouse.

Dynamics of subgroups of three mice within groups of four mice

We aimed at studying the dynamic of groups. We detail here the formation and breaking of groups of three mice. We defined the joiner and the breaker of these groups by identifying the mouse approaching and getting into contact or breaking contact at the frame just preceding or just following the group3 event, respectively. In the following paragraph, we detail how we calculate the odds of these events.

We consider only the initiation of a group of three mice (but the reasoning is similar for the ending of a group of three mice). With two wild-type (WT_1, WT_2) and two mutant (KO_1, KO_2) mice initially in the cage, six pairs are possible: WT_1-WT_2 , KO_1-KO_2 , WT_1-KO_1 , WT_1-KO_2 , WT_2-KO_1 and WT_2-KO_2 .

When calculating the expected probability of who is joining two mice to form a group of three mice, the odds to have an initial pair including two wild-type mice are $1/6$. The probability of a mutant mouse to join this pair is $1/1$ since it is the only genotype available. Therefore, the chance to have a mutant mouse joining a pair of wild-type mice is $p(KO|WT-WT) = 1/6 \times 1/1 = 1/6$. With a similar reasoning:

- The odds to have an initial pair including two mutant mice are $1/6$. The probability of a wild-type mouse to join this pair is $1/1$ since it is the only genotype available. Therefore, the chance to have a wild-type mouse joining a pair of mutant mice is $p(WT|KO-KO) = 1/6 \times 1/1 = 1/6$.

- The odds to have an initial pair including one wild-type mouse and one mutant mouse are $4/6$. The probability of a wild-type mouse to join this pair is $1/2$ since a wild-type and a mutant mouse are available. Therefore, the chance to have a wild-type mouse joining a pair of one wild-type mouse and one mutant mouse is $p(KO|WT-KO) = 4/6 \times 1/2 = 1/3$.

- The odds to have an initial pair including one wild-type mouse and one mutant mouse are $4/6$. The probability of a mutant mouse to join this pair is $1/2$ since a wild-type and a mutant mouse are left. Therefore, the chance to have a mutant mouse joining a pair of one wild-type mouse and one mutant mouse is $p(WT|WT-KO) = 4/6 \times 1/2 = 1/3$.

These probabilities can also be found by counting the possible groups (initial pair - joiner or breaker):

$WT_1, WT_2 - KO_1$	$WT_1, WT_2 - KO_2$		
$KO_1, KO_2 - WT_1$	$KO_1, KO_2 - WT_2$		
$WT_1, KO_1 - WT_2$	$WT_1, KO_1 - KO_2$	$WT_1, KO_2 - WT_2$	$WT_1, KO_2 - KO_1$
$WT_2, KO_1 - WT_1$	$WT_2, KO_1 - KO_2$	$WT_2, KO_2 - WT_1$	$WT_2, KO_2 - KO_1$

Software and system general description

We developed the methods as a plugin of Icy. The code is developed in Java 8 (189 classes. 19934 lines of code. 950 Methods. 32 Packages.). We put specific effort to make the code readable and documented for further extension. Experiments were first conducted on intel core i5-6400/2.4GHz/16GB and then on Ryzen 1800x. We recommend to use the last Ryzen processors available. LMT is highly multithreaded, the more virtual cores available the better.

Supplementary films

Method overview: this film presents the different steps of the method, and a general overview of the system.

Hardware blueprint and assembly instructions: this film shows 3D animations of the assembly, and also demonstrates how to perform the calibration of the system.

Investigating an abnormal nest building behaviour: this film is a case study of a typical problem: a nest that is built by the animal at an unexpected location. It shows how the tracking of the identities helps to address this question.

Background height-map demo: this demo shows the film of the background height-map, which is basically a view of the cage without the animals. In this film, we can see cotton moved by animals to form the nest, and also the food usage. It also displays paths that are created by animals when they move and push the sawdust out of their way.

Live 3D rendering demo: the purpose of this video is to showcase what can be enabled with real-time tracking. To create this video, we connected Live Mouse Tracker to a 3D world representing the cage and the whole room that we created with Unreal engine <http://www.unrealengine.com>. This connection allows the 3D rendering to be performed on each view in real time. The link between LMT and the 3D rendering is ensured by a UDP network connection, and can be either rendered on the computer tracking animals or on any other computer connected to the same network.

Validation: this film shows the data used to validate the method. It addresses the 4 animals' scenario. It displays the real observation augmented with the tracking information, and also the synchronized database view showing labelled events.

Tracking demo on mice having different coat colours: this film shows how the system copes with different appearances and also with a mix of animals having different coat colours.

Supplementary results

Supplementary Table 3 - Phenotype of the *Shank2* and *Shank3* mutant mice in previous studies conducted with the classical short-term phenotyping methods (NS: non-significant differences between wild-type and mutant mice).

	<i>Shank2</i> exon 16	<i>Shank3</i> exon 11
<i>activity</i>	hyperactivity ^{7,19}	hypoactivity ¹⁶
<i>stereotyped behaviors</i>	subtle increase in self-grooming ⁷	increased self-grooming ^{7,16,26}
<i>social interest</i>	decreased/NS interest for social interactions ^{7,8}	decreased ²⁶ /NS ¹⁶ interest for social interactions
<i>social recognition</i>	decreased preference for social novelty ⁷	decreased preference for social novelty ²⁶
<i>ultrasonic vocalizations</i>	decreased call rate, atypical repertoire, NS temporal structure, shorter sequences, decreased peak frequency ^{7,8}	NS call rate, NS repertoire ¹⁶
<i>spatial learning and memory</i>	decreased acquisition and memory ²⁷	NS acquisition and memory ²⁶
<i>spatial reversal learning</i>	-	decreased reversal learning ²⁶
<i>object recognition</i>	-	NS object distinction ²⁶
<i>reaction to inanimate objects</i>	-	decreased marble burying, decreased nest building ²⁶
<i>motor functions</i>	NS motor coordination ⁷	decreased muscle strength, decreased, subtle deficits in motor coordination ²⁶
<i>anxiety</i>	increased/NS anxiety ²⁸	reduced ¹⁶ /NS ²⁶ anxiety
<i>sensitivity</i>	-	decreased pain sensitivity ²⁶

Supplementary Table 4: Statistical analysis of the individual profiles of Shank2 and Shank3 mice. The stars represent the level of significance after Bonferroni correction for 33 tests: *: p<0.05, **: p<0.01, ***: p<0.001

	Shank2-/- vs Shank2+/- Fig. 3a		Shank3-/- vs Shank3+/- Fig. 3b		Shank2 vs Shank3 Fig. 3		Shank3-/- vs Shank3+/- (old) Suppl. Fig. S11b		Shank2 vs Shank3 (old) Suppl. Fig. S11	
	unpaired two-sided t-test exact p-value		unpaired two-sided t-test exact p-value		unpaired two-sided Wilcoxon test exact p-value		unpaired two-sided t-test exact p-value		unpaired two-sided Wilcoxon test exact p-value	
	t	p-value	t	p-value	W	p-value	t	p-value	W	p-value
Headingdown	2,3257	0,0326665	0,5879	0,5684685	138,0	0,2153757	3,3722	0,0062274	93,0	0,5450733
Huddled	1,0161	0,3238402	-1,3598	0,2011269	48,0	0,0116808	1,5159	0,1577434	50,0	0,0148145
Jumping	1,9825	0,0638243	-1,2592	0,2340391	170,0	0,0077014	-1,0209	0,3292143	165,0	0,0150404
SAP	-9,9513	0,0000000 ***	1,9446	0,0778235	29,0	0,0004449 *	0,6395	0,5355623	14,0	0,0000117 ***
Movingalone	8,0386	0,0000003 ***	-4,9039	0,0004687 *	213,0	0,0000002 ***	-0,9519	0,3615925	205,0	0,0000045 ***
Stoppedalone	-5,9036	0,0000174 ***	3,1191	0,0097654	14,0	0,0000117 ***	0,9250	0,3748155	26,0	0,0002393 **
Rearingisolated	1,6430	0,1187405	-2,1296	0,0566218	154,0	0,0530969	-3,0152	0,0117561	162,0	0,0218050
Outofnest	2,0370	0,0575314	0,8448	0,4162370	119,0	0,6615220	1,0106	0,3339065	116,0	0,7546103
Movingincontact	7,8107	0,0000005 ***	-2,8883	0,0147510	214,0	0,0000001 ***	-0,3683	0,7196151	210,0	0,0000007 ***
Rearingincontact	-1,4503	0,1651848	-2,0012	0,0706578	111,0	0,9170568	0,0695	0,9458682	89,0	0,4394918
Contact	-1,5929	0,1296112	1,1202	0,2864733	69,0	0,1037996	0,1690	0,8688964	78,0	0,2153757
Sidebysidesw	-4,3923	0,0003977 *	1,1147	0,2887569	32,0	0,0007944 *	0,6584	0,5238509	36,0	0,0016269
Sidebysideow	-4,3383	0,0004466 *	1,7841	0,1019903	28,0	0,0003636 *	1,3327	0,2095685	28,0	0,0003636 *
Nose.nose	-1,8341	0,0842012	2,7607	0,0185323	51,0	0,0150404	2,3066	0,0415450	58,0	0,0346271
Nose.anogenital	-0,0258	0,9796852	1,4987	0,1620933	92,0	0,5175847	0,5776	0,5751734	104,0	0,8840755
Groupof2mice	-1,1273	0,2752622	1,8259	0,0951109	66,0	0,0787915	0,4080	0,6911054	75,0	0,1715033
Groupof3mice	0,2821	0,7812603	0,0097	0,9923999	94,0	0,5732471	-0,2343	0,8190617	111,0	0,9170568
Trainof2mice	5,2262	0,0000684 **	-1,8268	0,0949600	209,0	0,0000010 ***	-0,6831	0,5086485	199,0	0,0000274 ***
Trainof3mice	1,5740	0,1339244	-4,7333	0,0006160 *	164,5	0,0129894	-0,3484	0,7341378	147,5	0,0845795
Trainof4mice	-1,4577	0,1631389	-1,4832	0,1660868	114,0	0,6926743	NA	NA	204,0	0,0000029 ***
Followwithinthepath	4,5028	0,0003138 *	-0,2594	0,8001016	178,0	0,0022777	0,6287	0,5423573	171,0	0,0066823
Approachinsocialrange.number.	-0,7486	0,4643488	1,5807	0,1422484	80,0	0,2486420	0,1543	0,8801405	103,0	0,8512851
Approachamouseinrearinginsocialrange.number.	0,2819	0,7814233	0,7336	0,4785431	98,0	0,6920743	-0,1288	0,8998778	109,0	0,9833829
Makingcontact.duration.	5,4570	0,0000426 **	-0,0384	0,9700382	190,0	0,0002393 **	0,3012	0,7688496	189,0	0,0002956 **
Makingcontact.number.	-0,9324	0,3641944	0,7785	0,4527217	91,0	0,4908061	-0,1777	0,8621711	105,0	0,9170568
Groupmakingof3mice.number.	4,9883	0,0001123 **	-1,9903	0,0719872	200,0	0,0000208 ***	0,1565	0,8784738	182,0	0,0011455 *
Groupmakingof4mice.number.	4,1749	0,0006351 *	-2,9462	0,0133005	205,0	0,0000045 ***	-0,1026	0,9201441	181,0	0,0013675 *
Movingawayfromamouseinsocialrange.number.	-0,9996	0,3315442	1,6648	0,1241426	73,0	0,1460980	0,1332	0,8964318	99,0	0,7231173
Breakingcontact.number.	-0,9996	0,3315442	1,6648	0,1241426	73,0	0,1460980	-1,1869	0,2602664	112,0	0,8840755
Groupbreakingof3mice.number.	4,7086	0,0002025 **	-1,8312	0,0942667	196,0	0,0000599 *	0,3119	0,7609507	175,0	0,0036799
Groupbreakingof4mice.number.	3,5148	0,0026575	-3,0687	0,0106843	198,0	0,0000359 *	-0,0141	0,9889746	172,0	0,0057820
seqoraloral.oralgeni.number.	-2,6385	0,0172479	2,4129	0,0344389	49,0	0,0115977	2,5172	0,0286191	48,0	0,0101454
seqoralgeni.oraloral.number.	-2,6637	0,0163679	1,5428	0,1511491	47,0	0,0088510	2,9915	0,0122648	43,0	0,0049883

Supplementary Table 5 - Statistics for the comparison of the mean duration of the group3 and group4 events, according to the way they were constituted (see Supplementary fig. S13). Set1 and Set2 refer to groups that are compared. For each set, the first genotype is the one of the maker of the group and the second genotype is the one of the breaker of the group. The durations are compared using Mann-Whitney U tests (W: test statistics, N: sample size).

Group3

Shank2	ko_wt_wt				
set1	set2	N Group1	N Group2	W	p-value
wt-wt	ko-ko	9	9	22	0,9999990
wt-wt	wt-ko	9	9	1	0,0078125
ko-wt	wt-ko	9	9	32	0,3007813
wt-wt	ko-wt	9	9	0	0,0039063
ko-ko	wt-ko	9	9	0	0,0039063
ko-ko	ko-wt	9	9	0	0,0039063

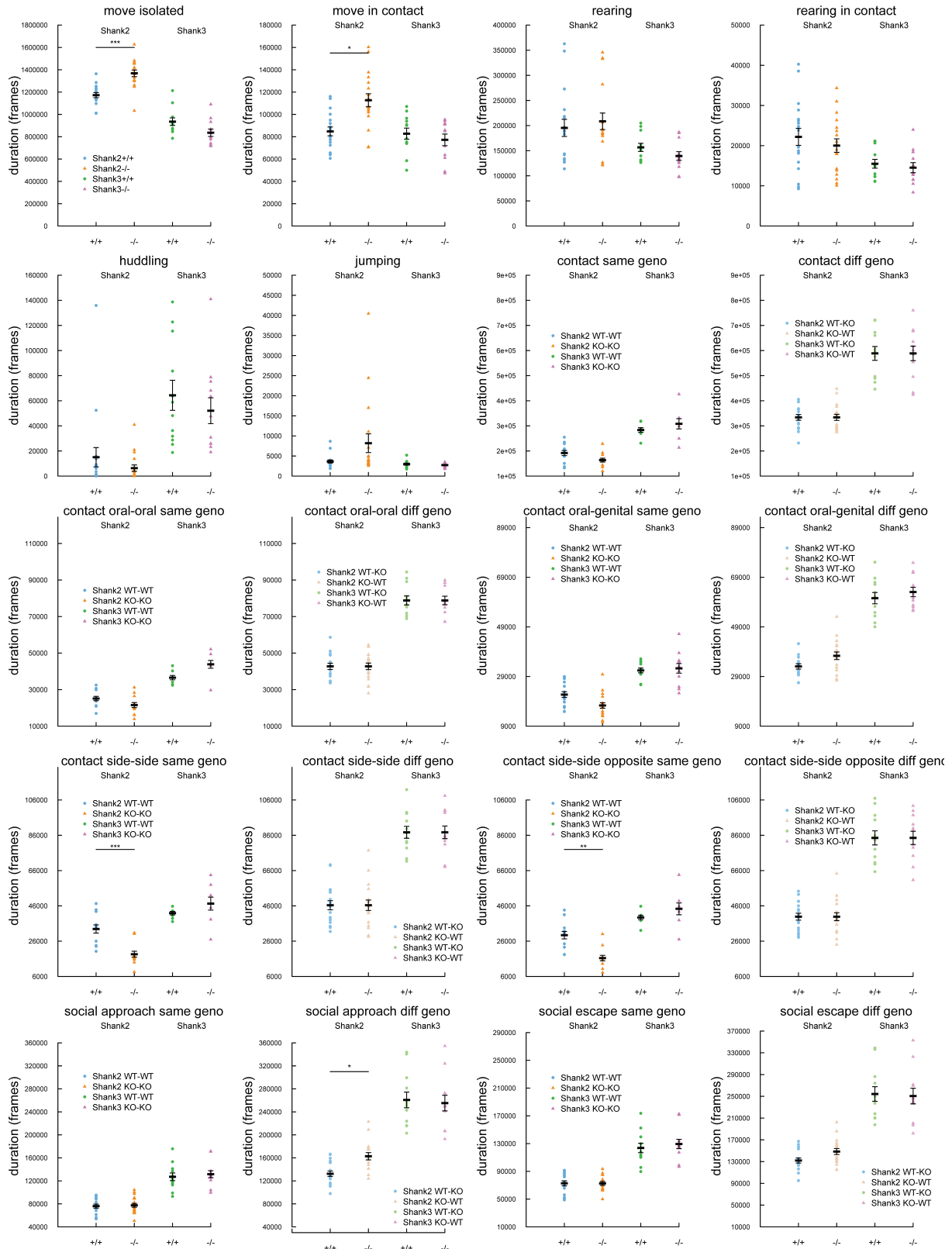
Shank2	wt_ko_ko				
set1	set2	N Group1	N Group2	W	p-value
wt-wt	ko-ko	9	9	30	0,4257813
wt-wt	wt-ko	9	9	3	0,0195313
ko-wt	wt-ko	9	9	35	0,1640625
wt-wt	ko-wt	9	9	0	0,0039063
ko-ko	wt-ko	9	9	0	0,0039063
ko-ko	ko-wt	9	9	0	0,0039063

Shank3	ko_wt_wt				
set1	set2	N Group1	N Group2	W	p-value
wt-wt	ko-ko	6	6	7	0,5625000
wt-wt	wt-ko	6	6	0	0,0312500
ko-wt	wt-ko	6	6	5	0,3125000
wt-wt	ko-wt	6	6	0	0,0312500
ko-ko	wt-ko	6	6	0	0,0312500
ko-ko	ko-wt	6	6	0	0,0312500

Shank3	wt_ko_ko				
set1	set2	N Group1	N Group2	W	p-value
wt-wt	ko-ko	6	6	17	0,2187500
wt-wt	wt-ko	6	6	0	0,0312500
ko-wt	wt-ko	6	6	0	0,0312500
wt-wt	ko-wt	6	6	0	0,0312500
ko-ko	wt-ko	6	6	0	0,0312500
ko-ko	ko-wt	6	6	0	0,0312500

Group4

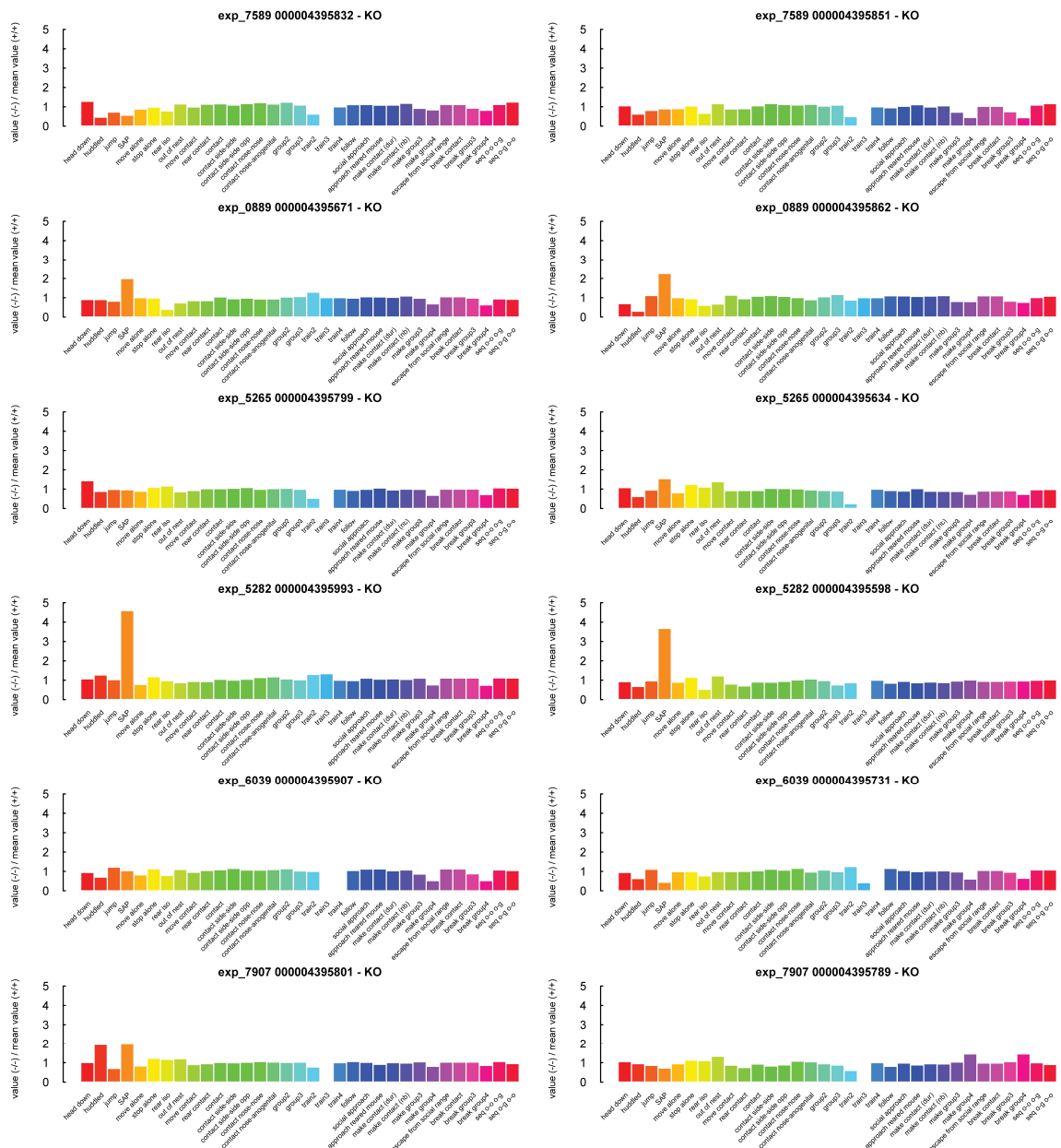
Strain	set1	set2	W	p-value
Shank2	wt-wt	ko-ko	19	0,7343750
Shank2	wt-wt	wt-ko	1	0,0078125
Shank2	wt-wt	ko-wt	1	0,0078125
Shank2	ko-ko	wt-ko	8	0,0976563
Shank2	ko-ko	ko-wt	0	0,0039063
Shank2	wt-ko	ko-wt	23	0,9999990
Shank3	wt-wt	ko-ko	3	0,1562500
Shank3	wt-wt	wt-ko	0	0,0312500
Shank3	wt-wt	ko-wt	0	0,0312500
Shank3	ko-ko	wt-ko	0	0,0312500
Shank3	ko-ko	ko-wt	0	0,0312500
Shank3	wt-ko	ko-wt	13	0,6875000



Supplementary Fig. S8 | Total duration of a subsample of the behavioural events extracted from the monitoring over 23h of *Shank2* and *Shank3* mixed-genotype groups of four mice (uncorrected p-values of two-sided Wilcoxon rank sum test, with a star when significance remains after Bonferroni correction for multiple testing; move isolated: *Shank2* mice: W=26,

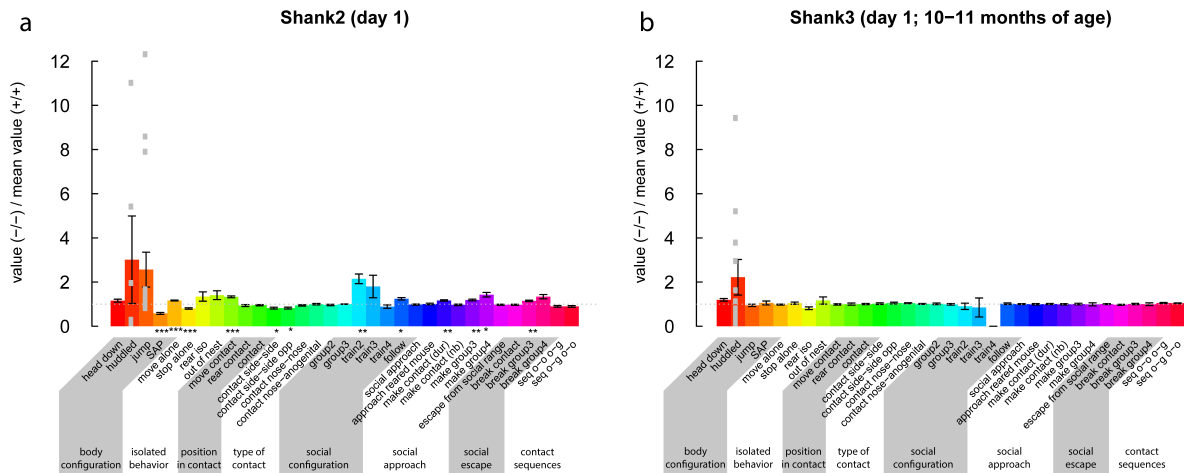
$p=2.533 \times 10^{-6}$ *, *Shank3* mice: $W=110$, $p=0.028$; move in contact: *Shank2* mice: $W=58$, $p=6.624 \times 10^{-4}$ *; jumping: *Shank2* mice: $W=94$, $p=0.0315$; contact oral-oral same genotype: *Shank3* mice: $W=28$, $p=0.0118$; contact oral-genital same genotype: *Shank2* mice: $W=242$, $p=0.011$; contact oral-genital different genotype: *Shank2* mice: $W=99$, $p=0.047$; contact side-side same genotype: *Shank2* mice: $W=292$, $p=4.097 \times 10^{-5}$ *; contact side-side opposite same genotype: *Shank2* mice: $W=296$, $p=2.351 \times 10^{-5}$ *; social approach different genotype: *Shank2* mice: $W=56$, $p=5.033 \times 10^{-4}$ *; social escape different genotype: *Shank2* mice: $W=95$, $p=0.034$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; *Shank2*: $N=9$ independent experiments; *Shank3*: $N=6$ independent experiments). Data are presented as mean $\pm$ s.e.m., as well as points for each experiment.

Supplementary Fig. S9 | Individual profiles of each *Shank2*^{-/-} mouse calculated over 23h of recording. One row represents the profiles of each of the two mutant mice from the same experiment (9 experiments in total). Traits that were not different from the mean value of the wild-type mice of the experiment were set at zero. Traits that are more expressed in mutant mice than in the mean of wild-type mice have positive values, while traits that are less expressed in mutant mice than in the mean of wild-type mice have negative values. Y-axis graduations represent the order of magnitude of the difference between mutant and wild-type mice, i.e., a value of 2 represents a trait that is two times more expressed in the mutant mice than in the wild-type mean mouse.

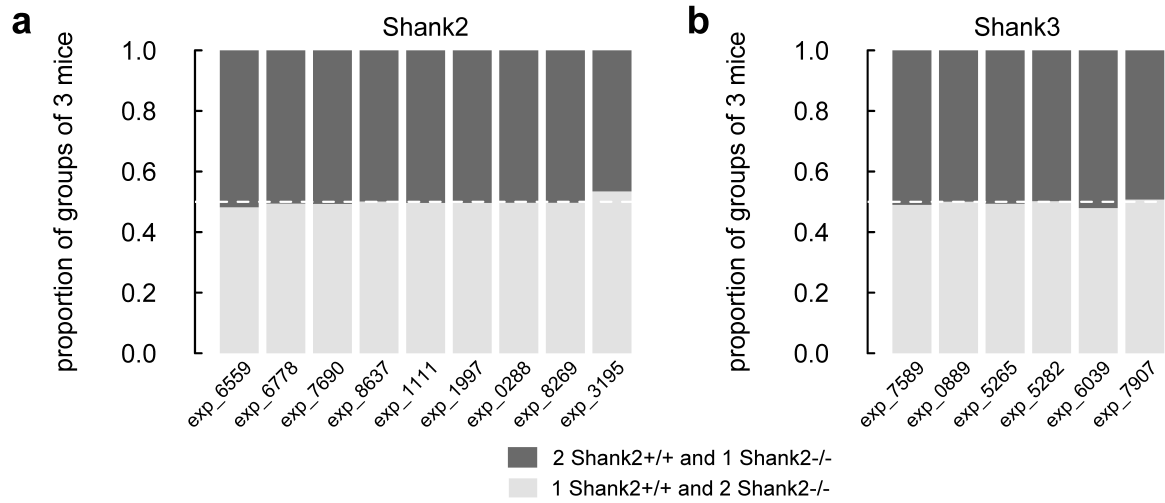


Supplementary Fig. S10 | Individual profiles of each *Shank3*^{-/-} mouse calculated over 23h of recording. One row represents the profiles of each of the two mutant mice from the same experiment (6 experiments in total). Traits that were not different from the mean value of the wild-type mice of the experiment were set at zero. Traits that are more expressed in mutant

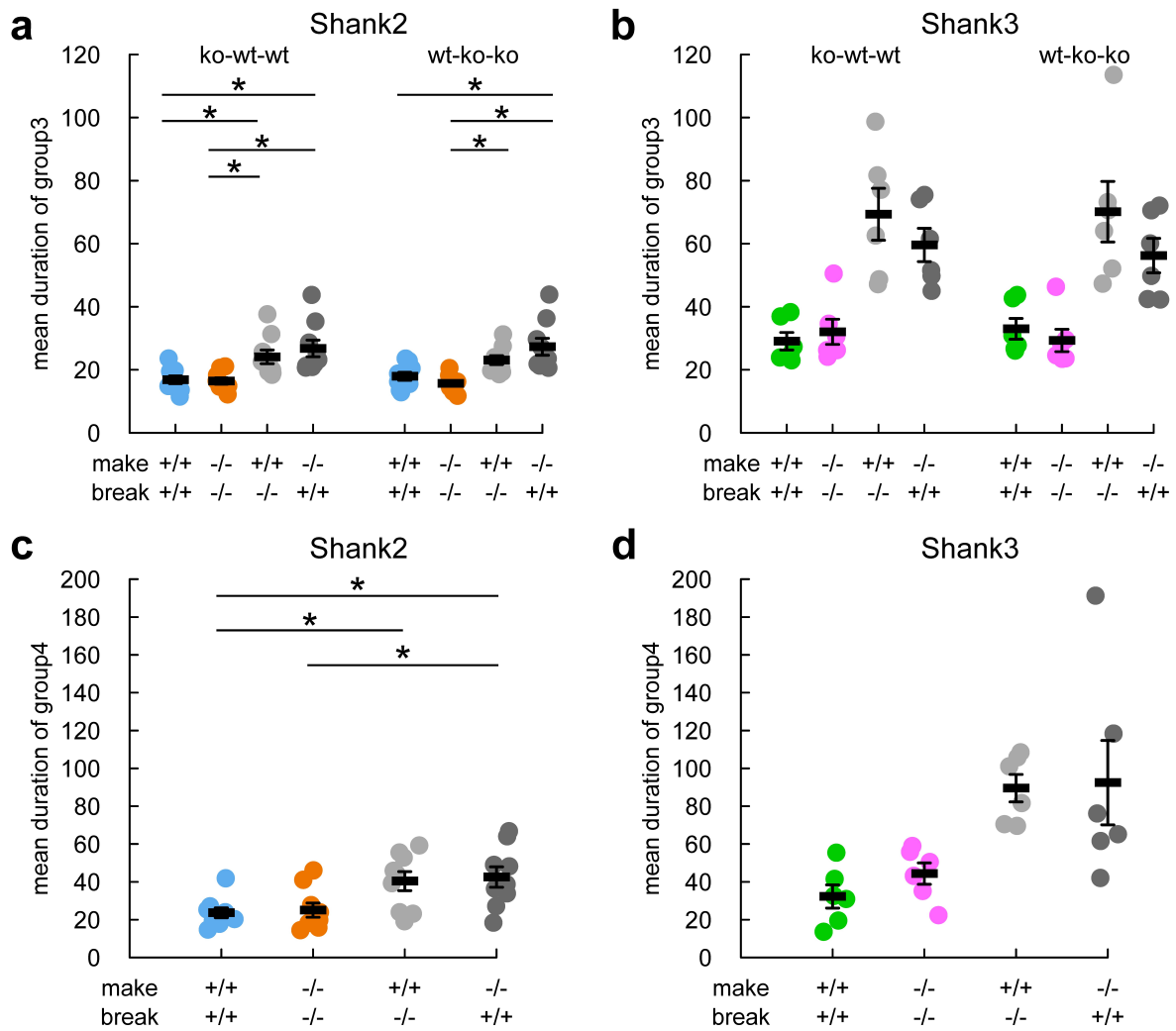
mice than in the mean of wild-type mice have positive values, while traits that are less expressed in mutant mice than in the mean of wild-type mice have negative values. Y-axis graduations represent the order of magnitude of the difference between mutant and wild-type mice, i.e., a value of 2 represents a trait that is two times more expressed in the mutant mice than in the wild-type mean mouse.



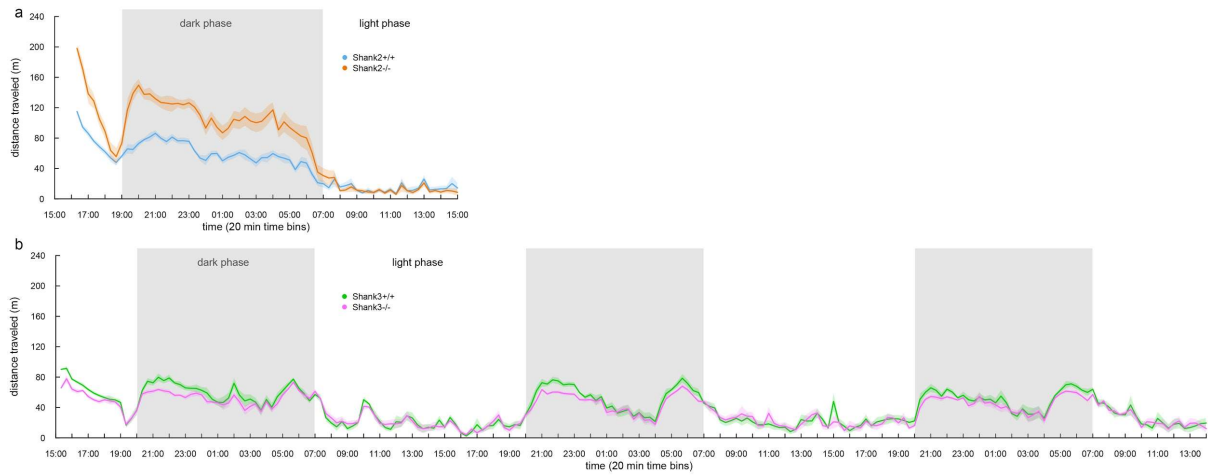
Supplementary Fig. S11 | Behavioural profiles of *Shank2*^{-/-} mice (a) and 10-11 months old *Shank3*^{-/-} mice (b). Over the first day of group monitoring, 33 behavioural traits were computed for each individual (see definitions in **Supplementary methods - Behavioural event extraction**). Each behavioural trait value for one mutant mouse divided by the mean of the two wild-type mice within each cage was compared to one using Wilcoxon tests (corrected for multiple testing, since 33 tests were conducted for each strain; after correction: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$; *Shank2*: $N = 18$ mice; *Shank3*: $N = 12$ mice). Traits that were not different from the mean value of the wild-type mice of the experiment were set at one. Traits that were expressed more in mutant mice than in the mean of wild-type mice had values larger than one, while traits that were expressed less in mutant mice than in the mean of wild-type mice had values smaller than one. Y-axis graduations represent the expression ratio between mutant and wild-type mice, i.e., a value of 2 represents a trait that was expressed two times more in the mutant mice than in the wild-type mean mouse. Significant differences emerged between the *Shank2* and the *Shank3* profiles in activity measures (Wilcoxon tests with Bonferroni corrections for multiple testing: time spent moving alone [$W = 205$, $p < 0.00003$] and in contact [$W = 210$, $p < 0.00003$], and time spent stopped alone [$W = 26$, $p = 0.00024$]), in exploratory measures (time spent in SAP [$W = 14$, $p = 0.00001$]), in time spent in side by side contact (opposite [$W = 28$, $p = 0.00036$] and same way [$W = 36$, $p = 0.00163$]), in the duration of social approaches (duration of making contact $W = 189$, $p = 0.00030$), in the time spent in train involving two mice ($W = 199$, $p = 0.00003$) and in the making of group3 ($W = 182$, $p = 0.00115$) and group4 ($W = 181$, $p = 0.00137$). A train involving the four mice together was not detected in the old *Shank3* mice. Data are presented as mean $\pm$ s.e.m. For complete statistical analyses, see **Supplementary Table 4**.



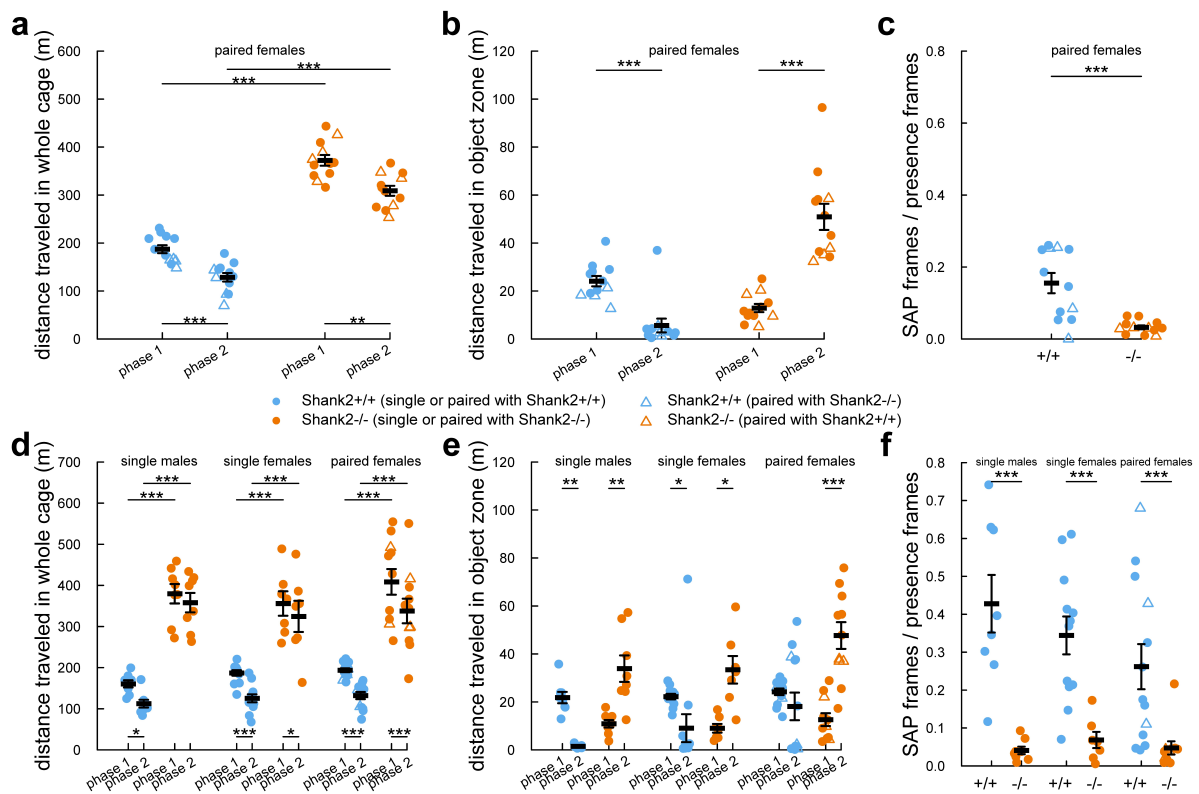
Supplementary Fig. S12 | Proportions of groups of three individuals for each one-day recording of four individuals for (a) *Shank2* mice and (b) *Shank3* mice. Two types of groups of three individuals were possible within the groups of four mice: two mutant mice with one wild-type mouse (light grey) and one mutant mouse with two wild-type mice (dark grey). The expected proportion (white line) was 0.5 given the composition of the cage (two mutant mice and two wild-type mice). We used a Chi square conformity test. No significant differences were found.



Supplementary Fig. S13 | Mean duration (frames) of groups of three and four individuals for each one-day recording of four individuals for (a) *Shank2* mice and (b) *Shank3* mice according to who is coming in and coming out. Wilcoxon tests with Bonferroni corrections for multiple testing. Data are presented as mean $\pm$ s.e.m. as well as points for each experiment (two-sided Wilcoxon rank sum test; p-values after correction for multiple testing: *: $p < 0.05$; for exact p-values, see **Supplementary Table 5**). *Shank2* strain: N=9 independent experimental groups; *Shank3* strain: N=6 independent experimental groups.



Supplementary Fig. S14 | Mean activity levels of (a) *Shank2* mice and (b) *Shank3* mice over one day and three days of recording, respectively. We computed the distance travelled by each mouse per 20 min time bins. We present the mean distance travelled per time bin for each genotype. *Shank2* strain: 18 *Shank2*^{-/-} mice & 18 *Shank2*^{+/+} mice; *Shank3* strain: 12 *Shank3*^{-/-} mice & 12 *Shank3*^{+/+} mice.

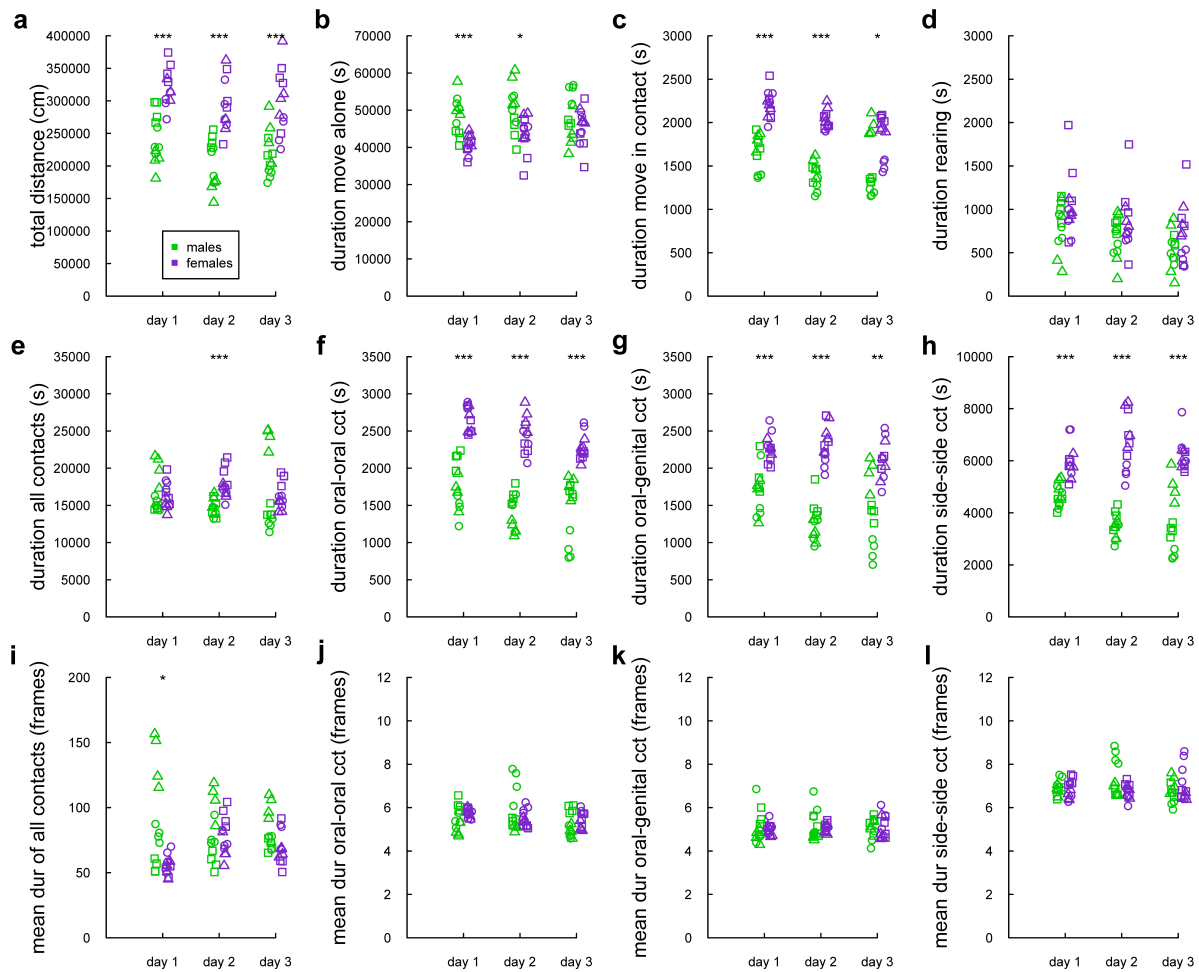


Supplementary Figure S15 | Hyperactivity and atypical exploration strategy in paired conditions and confirmed in a second cohort of *Shank2*^{-/-} mice. (a) Distance travelled in the entire cage in phase 1 and phase 2 for wild-type and *Shank2*^{-/-} female mice from cohort 1 in paired conditions (phase 1 vs phase 2 [paired Wilcoxon signed rank test]: *Shank2*^{+/+}: $W=78$, $p=4.883 \times 10^{-4}$; *Shank2*^{-/-}: $W=77$, $p=9.766 \times 10^{-4}$; wild-type vs mutant in phase 1 [unpaired

Wilcoxon rank sum test]: *Shank2*^{+/+} vs *Shank2*^{-/-}: $W=0$, $p=7.396 \times 10^{-7}$; wild-type vs mutant in phase 2 [unpaired Wilcoxon rank sum test]: *Shank2*^{+/+} vs *Shank2*^{-/-}: $W=0$, $p=7.396 \times 10^{-7}$). **(b)** Distance travelled in the object zone in phase 1 and phase 2 for wild-type and *Shank2*^{-/-} female mice from cohort 1 in paired conditions (phase 1 vs phase 2 [paired Wilcoxon signed rank test]: *Shank2*^{+/+}: $W=77$, $p=9.766 \times 10^{-4}$; *Shank2*^{-/-}: $W=0$, $p=4.883 \times 10^{-4}$). **(c)** Ratio of the number of frames in the object zone where the animal was detected in stretched attend posture over the total number of frames spent in the object zone (unpaired Wilcoxon rank sum test: *Shank2*^{+/+} vs *Shank2*^{-/-}: $W=128$, $p=6.560 \times 10^{-4}$). **(d)** Distance travelled in the entire cage in phase 1 and phase 2 for wild-type and *Shank2*^{-/-} mice from cohort 2 for males in the single condition, and for females in the single and paired conditions (phase 1 vs phase 2 [paired Wilcoxon signed rank test]: *Shank2*^{+/+} single males: $W=35$, $p=0.016$; *Shank2*^{-/-} single males: $W=25$, $p=0.382$; *Shank2*^{+/+} single females: $W=78$, $p=4.883 \times 10^{-4}$; *Shank2*^{-/-} single females: $W=26$, $p=0.046$; *Shank2*^{+/+} paired females: $W=91$, $p=2.441 \times 10^{-4}$; *Shank2*^{-/-} paired females: $W=66$, $p=9.766 \times 10^{-4}$; *Shank2*^{+/+} vs *Shank2*^{-/-} in phase 1 [unpaired Wilcoxon rank sum test]: single males: $W=0$, $p=1.554 \times 10^{-4}$; single females: $W=0$, $p=3.969 \times 10^{-5}$; paired females: $W=0$, $p=8.012 \times 10^{-7}$; *Shank2*^{+/+} vs *Shank2*^{-/-} in phase 2 [unpaired Wilcoxon rank sum test]: single males: $W=0$, $p=1.554 \times 10^{-4}$; single females: $W=2$, $p=1.588 \times 10^{-4}$; paired females: $W=0$, $p=8.012 \times 10^{-7}$). **(e)** Distance travelled in the object zone in wild-type mice and *Shank2*^{-/-} mice from cohort 2 in phase 1 and phase 2 for males in the single condition, and for females in the single and paired conditions (phase 1 vs phase 2 [paired Wilcoxon signed rank test]: *Shank2*^{+/+} single males: $W=36$, $p=0.008$; *Shank2*^{-/-} single males: $W=0$, $p=0.008$; *Shank2*^{+/+} single females: $W=66$, $p=0.034$; *Shank2*^{-/-} single females: $W=0$, $p=0.016$; *Shank2*^{+/+} paired females: $W=65$, $p=0.191$; *Shank2*^{-/-} paired females: $W=0$, $p=9.766 \times 10^{-4}$). **(f)** Proportion of frames in which mice are in SAP in the object zone over the total number of frames in which animals are detected in the object zone (*Shank2*^{+/+} vs *Shank2*^{-/-} in phase 2 [unpaired Wilcoxon rank sum test]: single males: $W=64$, $p=1.554 \times 10^{-4}$; single females: $W=64$, $p=1.554 \times 10^{-4}$; paired females: $W=132$, $p=1.562 \times 10^{-4}$). Data are presented as mean $\pm$ s.e.m. and individual points for: (a-c) cohort 1: 12 *Shank2*^{+/+} mice and 12 *Shank2*^{-/-} mice; (d-e) cohort 2: 12 *Shank2*^{+/+} female mice and 7 *Shank2*^{-/-} female mice and 8 *Shank2*^{+/+} male mice and 8 *Shank2*^{-/-} male mice in cohort 2. Exact p-values uncorrected for multiple testing are presented (two-sided Wilcoxon tests; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$).

Sex-related variations in the behaviour of C57BL/6J adult mice

The LMT system is a highly flexible system that allows a large diversity of experiments. In the present supplementary study, we aimed at pointing out the behavioural differences between males and females in adult C57BL/6J mice. For this purpose, we tracked over three days three groups of four male mice and three groups of four female mice aged of 3 months. We tracked individually the four mice constituting each group within the 50 x 50 x 25 cm test cage (Plexiglas, with fresh bedding, food and water *ad libitum*, red house and six cottons used for nesting). We tested each day separately using non-parametric Wilcoxon Mann-Whitney U-tests to control whether sex-related variations were more obvious in normal home cage behaviour (third day for instance) than in the exploration phase (first day).



Supplementary Figure S16 | Sex-related differences in activity and social behaviours in C57BL/6J mice. Behavioural measures automatically recorded during three-days tracking of four C57BL/6J mice in same-sex groups of males (green; three groups) and females (purple; three groups): (a) total distance travelled (unpaired Wilcoxon rank sum tests; day1: $W=5$, $p=1.405 \times 10^{-5}$; day2: $W=3$, $p=5.177 \times 10^{-6}$; day3: $W=12$, $p=2.012 \times 10^{-4}$), (b) total time spent moving alone (i.e., not in contact with any other individual) (unpaired Wilcoxon rank sum tests; day1: $W=133$, $p=1.442 \times 10^{-4}$; day2: $W=119$, $p=0.006$; day3: $W=86$, $p=0.443$), (c) total time spent moving in contact with at least one other individual (unpaired Wilcoxon rank sum tests; day1: $W=0$, $p=7.396 \times 10^{-7}$; day2: $W=0$, $p=7.396 \times 10^{-7}$; day3: $W=27$, $p=0.008$), (d) total time spent in rearing (i.e., standing up on the hind legs) (unpaired Wilcoxon rank sum tests; day1: $W=51$, $p=0.242$; day2: $W=52$, $p=0.266$; day3: $W=49$, $p=0.198$), (e) total time spent in any type of contacts (unpaired Wilcoxon rank sum tests; day1: $W=66$, $p=0.755$; day2: $W=7$, $p=3.328 \times 10^{-5}$; day3: $W=51$, $p=0.242$), (f) total time spent in oral-oral contacts (unpaired Wilcoxon rank sum tests; day1: $W=0$, $p=7.396 \times 10^{-7}$; day2: $W=0$, $p=7.396 \times 10^{-7}$; day3: $W=0$, $p=7.396 \times 10^{-7}$), (g) total time spent in oral-anogenital contacts (unpaired Wilcoxon rank sum tests; day1: $W=11$, $p=1.442 \times 10^{-4}$; day2: $W=0$, $p=7.396 \times 10^{-7}$; day3: $W=16$, $p=6.560 \times 10^{-4}$), (h) total time spent in side-by-side contacts (unpaired Wilcoxon rank sum tests; day1: $W=3$, $p=5.177 \times 10^{-6}$; day2: $W=0$, $p=7.396 \times 10^{-7}$; day3: $W=3$, $p=5.177 \times 10^{-6}$), (i) mean duration of one contact event, for any type of contact (unpaired Wilcoxon rank sum tests; day1: $W=120$, $p=0.005$; day2: $W=78$, $p=0.755$; day3: $W=110$, $p=0.028$), (j) mean duration of one event of oral-oral contact (unpaired Wilcoxon rank sum tests; day1: $W=50$, $p=0.219$; day2: $W=74$, $p=0.932$; day3: $W=53$, $p=0.291$), (k) mean duration of one event of oral-genital contact (unpaired Wilcoxon rank sum tests; day1:

W=65, $p=0.713$; day2: W=58, $p=0.442$; day3: W=73, $p=0.977$) and (I) mean duration of one event of side-by-side contact (unpaired Wilcoxon rank sum tests; day1: W=69, $p=0.887$; day2: W=100, $p=0.114$; day3: W=62, $p=0.590$). On the plots, mice belonging to the same group are depicted by the same shape (triangles, squares, and circles). Data are presented as individual points (12 individuals of each sex) and tested between sexes with non-parametric two-sided Wilcoxon rank sum tests (uncorrected p-values: *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$). No significant differences emerged over the three days of recording within each sex.

Overall, using the LMT system, we highlighted a different level of activity, as well as increased specific types of social contacts in females in comparison with males in C57BL/6J mice. C57BL/6J female mice were significantly more active in comparison with C57BL/6J male mice over each of the three days of tracking (**Supplementary Fig. S16a**). Interestingly, in comparison with males, females spent a reduced amount of time moving alone (**Supplementary Fig. S16b**), while they spent more time moving in contact with other mice (**Supplementary Fig. S16c**). The duration of vertical exploration (namely rearing duration) did not differ significantly between male and female mice (**Supplementary Fig. S16d**). Interestingly, in the social domain, sex-related differences were not obvious in the total time spent in any types of contacts (**Supplementary Fig. S16e**), while they became significant when specific types of contacts were considered. Indeed, females spent significantly longer time in nose-to-nose contacts (**Supplementary Fig. S16f**), in nose-to-anogenital contacts (**Supplementary Fig. S16g**), and in side-by-side contacts (**Supplementary Fig. S16h**) in comparison with males. These increases in the total duration of these different types of contacts were not due to an increase in the mean duration of contact events (**Supplementary Fig. S16i,j,k,l**), but therefore to the increase in the number of events.

Supplementary references

1. Anonymous. Building a better mouse test. *Nat Meth* **8**, 697–697 (2011).
2. Wright, J. Why studying autism in mice may be doomed to fail. *Spectrum / Autism Research News* (2018).
3. So, N., Franks, B., Lim, S. & Curley, J. P. A Social Network Approach Reveals Associations between Mouse Social Dominance and Brain Gene Expression. *PLoS One* **10**, e0134509 (2015).
4. Shemesh, Y. *et al.* High-order social interactions in groups of mice. *Elife* **2**, (2013).
5. Weissbrod, A. *et al.* Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nat. Commun.* **4**, (2013).
6. Freund, J. *et al.* Emergence of Individuality in Genetically Identical Mice. *Science* (80-.). **340**, 756–759 (2013).
7. Ohayon, S., Avni, O., Taylor, A. L., Perona, P. & Roian Egnor, S. E. Automated multi-day tracking of marked mice for the analysis of social behaviour. *J. Neurosci. Methods* **219**, 10–19 (2013).
8. Bove, M. *et al.* The Visible Burrow System: A behavioral paradigm to assess sociability and social withdrawal in BTBR and C57BL/6J mice strains. *Behav. Brain Res.* **344**, 9–19 (2018).
9. Esler, A. & Grissom, N. Why we need a mouse version of a diagnostic test for autism.

Spectrum / Autism Research News (2018). Available at: <https://spectrumnews.org/opinion/viewpoint/need-mouse-version-diagnostic-test-autism/>.

10. Fonio, E., Golani, I. & Benjamini, Y. Measuring behavior of animal models: faults and remedies. *Nat. Publ. Gr.* **9**, 1167–1170 (2012).
11. Dahlbom, S. J., Lagman, D., Lundstedt-Enkel, K., Sundström, L. F. & Winberg, S. Boldness Predicts Social Status in Zebrafish (*Danio rerio*). *PLoS One* **6**, e23565 (2011).
12. Wiltshko, A. B. *et al.* Mapping Sub-Second Structure in Mouse Behavior. *Neuron* **88**, 1121–1135 (2015).
13. de Chaumont, F. *et al.* Computerized video analysis of social interactions in mice. *Nat. Methods* **9**, 410-U134 (2012).
14. Unger, J. *et al.* An unsupervised learning approach for tracking mice in an enclosed area. *BMC Bioinformatics* **18**, 272 (2017).
15. Hong, W. *et al.* Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proc. Natl. Acad. Sci.* **112**, E5351–E5360 (2015).
16. Matsumoto, J. *et al.* A 3D-Video-Based Computerized Analysis of Social and Sexual Interactions in Rats. *PLoS One* **8**, e78460 (2013).
17. Giancardo, L. *et al.* Automatic Visual Tracking and Social Behaviour Analysis with Multiple Mice. *PLoS One* **8**,
18. Blanchard, D. C. *et al.* Visible burrow system as a model of chronic social stress: Behavioral and neuroendocrine correlates. *Psychoneuroendocrinology* **20**, 117–134 (1995).
19. Howerton, C. L., Garner, J. P. & Mench, J. A. A system utilizing radio frequency identification (RFID) technology to monitor individual rodent behavior in complex social settings. *J. Neurosci. Methods* **209**, 74–78 (2012).
20. Castelhana-Carlos, M., Costa, P. S., Russig, H. & Sousa, N. PhenoWorld: a new paradigm to screen rodent behavior. *Transl. Psychiatry* **4**, e399 (2014).
21. König, B. *et al.* A system for automatic recording of social behavior in a free-living wild house mouse population. *Anim. Biotelemetry* **3**, 39 (2015).
22. Bains, R. S. *et al.* Analysis of Individual Mouse Activity in Group Housed Animals of Different Inbred Strains using a Novel Automated Home Cage Analysis System. *Front. Behav. Neurosci.* **10**, 106 (2016).
23. Torquet, N. *et al.* Social interactions impact on the dopaminergic system and drive individuality. *Nat. Commun.* **9**, 3081 (2018).
24. Pérez-Escudero, A., Vicente-Page, J., Hinz, R. C., Arganda, S. & de Polavieja, G. G. idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat. Methods* **11**, 743–748 (2014).
25. Romero-Ferrero, F., Bergomi, M. G., Hinz, R. C., Heras, F. J. H. & de Polavieja, G. G. idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. *Nat. Methods* **16**, 179–182 (2019).
26. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

27. Alexandrov, V., Brunner, D., Hanania, T. & Leahy, E. Highthroughput analysis of behavior for drug discovery. *Eur. J. Pharmacol.* **750**, 82–89 (2015).
28. Bernardin, K., Elbs, A. & Stiefelbogen, R. *Multiple object tracking performance metrics and evaluation in a smart Room environment. Proceedings of IEEE International Workshop on Visual Surveillance* (2006).
29. Lemberger, P., Batty, M., Morel, M. & Raffaelli, J.-L. *Big Data et Machine Learning - Manuel du data scientist.* (2016).