

# CELLMISSY MANUAL

## INTRODUCTION

This document is intended as a supporting material for researchers that wish to use the **CellMissy** tool.

**CellMissy** is a cross-platform, generic and easily extensible data management and analysis system for cell migration experiments, being focused in its current version on wound healing-like experiments. It is entirely written in Java and is freely available under the Apache2 open source license at <https://github.com/compomics/cellmissy/>.

**CellMissy** is described in “*CellMissy: a tool for management, storage and analysis of cell migration data produced in wound healing-like assays.*” (P. Masuzzo, N. Hulstaert, L. Huyck, C. Ampe, M. Van Troys and L. Martens).

## 1. HOW TO RUN CELLMISSY FOR THE FIRST TIME

In its simple, single-user setup, **CellMissy** can run on minimal hardware, so any modern laptop or desktop PC is more than sufficient. Furthermore, since **CellMissy** is written in Java, it can run on any platform that supports a Java Virtual Machine version 1.8.0 or above (Windows, Linux, and Mac OS-X). However, if **CellMissy** is to be used as a shared system between many different users, it will be more practical to set up a central database (DB) that all users can access simultaneously. This DB server again needs not be a high-end machine, and the task can easily be handled by any modern desktop machine with sufficient storage space. **CellMissy** handles both scenarios (single-user or multi-user) with equal ease, and has been designed to support a full-blown production environment in a large lab without hiccups.

Before you can correctly use **CellMissy** for the first time, you need to follow some configuration steps, in order to set up a DB connection and create the MySQL schema for the application:

1. connect to a MySQL server (e.g. with MySQL Workbench, see <http://www.mysql.com/products/workbench/>)
2. create a new schema in the connected server and set it as the default schema (i. e. make the schema the active one in the current session)

3. run the SQL script "***cellmissy\_schema\_1.4.0.sql***" (you can get it from <https://github.com/compomics/cellmissy/blob/master/supportFiles/>, do NOT use the old 1.0.4 version as this is incompatible with the current version of CellMissy): this will create the tables for the DB, set the indexes for them, and insert basic records into the DB (e.g. some cell lines, migration/invasion assays, extracellular matrix compositions...)

Please note that the provided SQL script works for MySQL relational databases; if you want to use different DB types, let us know and we'll try to provide you with another script.

After you have configured the DB connection for **CellMissy**, you can run the application by double clicking the executable *.jar* file present in the "**CellMissy**" folder (note that you need to unzip the compressed CellMissy folder before you can execute the *.jar* file). At this stage, you can edit **CellMissy** properties that establish the connection to the DB, by clicking the "*edit properties*" button in the login dialog, as shown below.

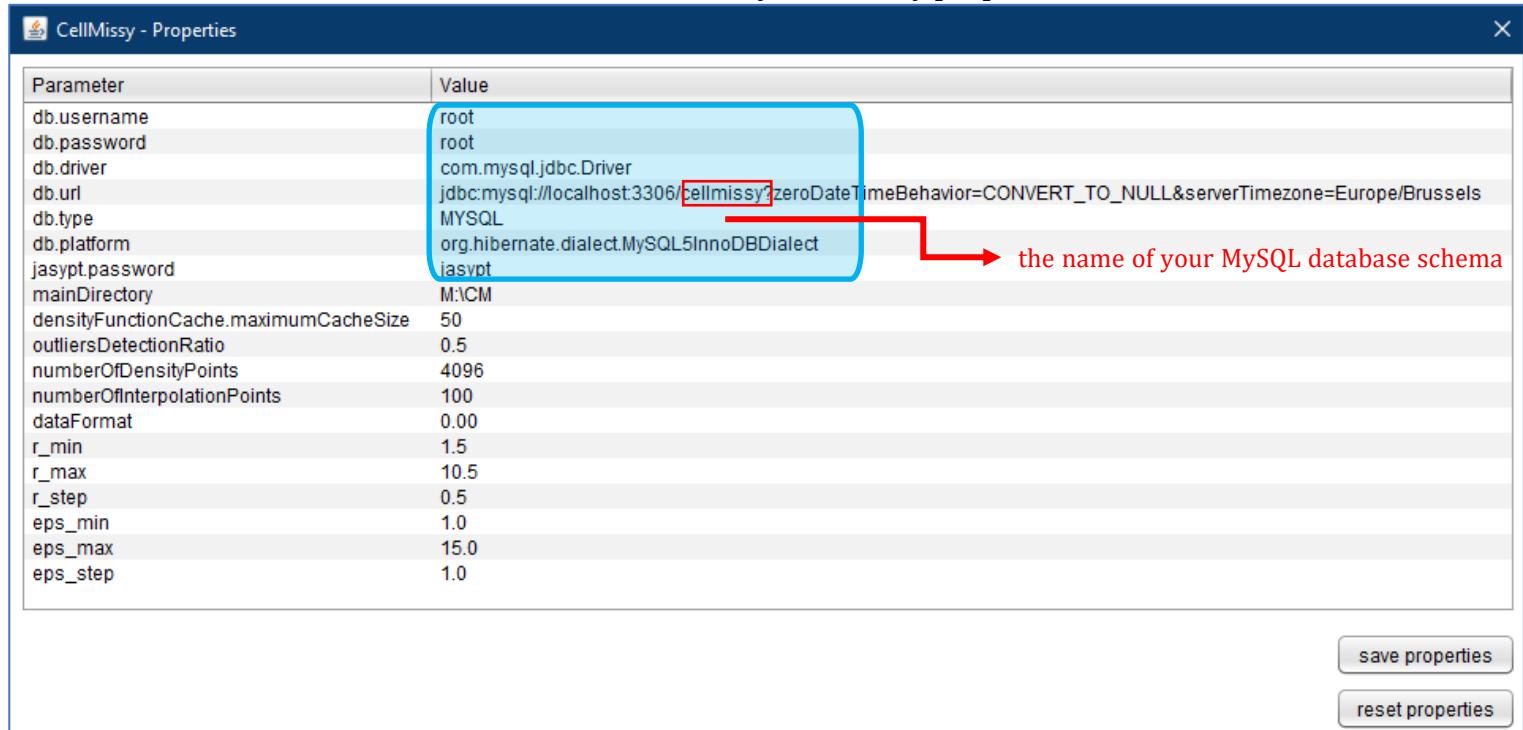


Another dialog will pop up with a table containing the properties of the software and their values, as shown in the screenshot below. Here, you can set the connection parameters (*db.username*, *db.password*, *db.driver*, *db.url*, *db.type*) according to the configuration properties you have chosen in the three steps above. Once the new properties have been saved, the application will automatically shut down: you can then restart **CellMissy** (again double clicking the executable *.jar* file) and use it with the current DB parameters.

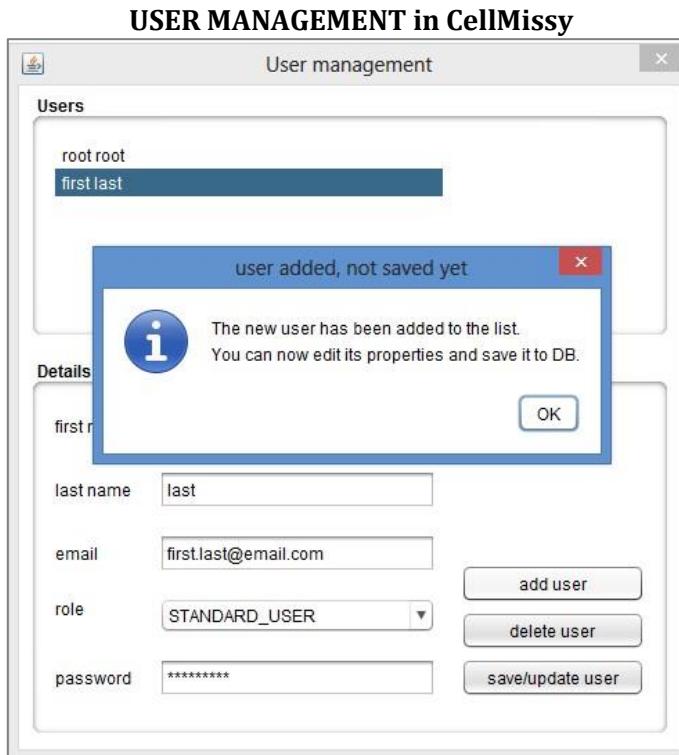
To login into the application, you can use the credentials from the *root application user* (this *root application user* is inserted into the DB when you run the "***cellmissy\_schema.sql***" script):

username	password
<i>root</i>	<i>cellmissy</i>

## CONFIGURING CellMissy – CellMissy properties



Having logged in with this root user, you will have *ADMIN* rights: this means that you will also have access to the ***User Management*** module (through the *Edit* menu); here, you can delete the root user or change its credentials, and you can add other application users as well. Each time you add a user, this is automatically inserted in the list present in the GUI (see following screenshot), you can then edit the data of the new added user and finally save the user to the DB. Note that if you enter the application as an *ADMIN* user, you can also select a user from the list and delete it from the DB.

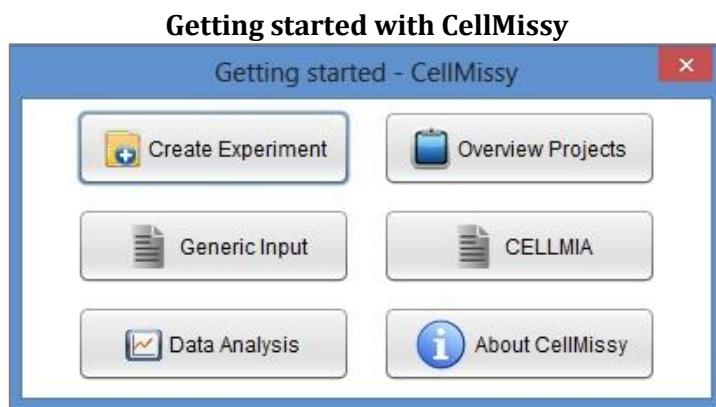


## 2. HOW TO USE CELLMISSY

Developed to follow the steps typically encountered in a cell migration experiment, **CellMissy** is mainly composed of three modules, as described below.

- **Experiment Manager** → to set up a new cell migration/invasion experiment.
- **Data Loader** → to import and store cell migration data; for a typical wound healing-like experiment these are values of measured area in time.
- **Data Analyzer** → to explore and analyze cell migration data; perform statistics, and finally create analysis reports.

Each of these main modules can be accessed in **CellMissy** through the main *File* menu, as well as through the start-up dialog that will automatically appear when you run **CellMissy**, as shown in the next screenshot.



On top of these modules, **CellMissy** also provides means for import/export of an entire experiment, along with the import/export of templates containing all the experimental set-up metadata. These functionalities can be accessed through the *File* menu.

The next sections of this document will provide more details on each of the module of **CellMissy**.

## 2.1 EXPERIMENT MANAGER - CREATE A NEW CELL MIGRATION EXPERIMENT



This module guides you through the set-up of a new cell migration/invasion experiment.

### EXPERIMENT MANAGER - experiment metadata

**Overview**

Projects	Experiments (status)
P000	E005, IN_PROGRESS
P001	E006, IN_PROGRESS
P002	E007, IN_PROGRESS
P004	E008, IN_PROGRESS E009, IN_PROGRESS

**Experiment Data**

Number\* 15

Date\* 31-Jul-2019

Insert Experiment Purpose:  
experiment purpose goes here...

**Project Description**

demo\_project

**Microscope Data**

Select Instrument generic microscope

Select Magnification 10x

**Image Analysis Data**

You will analyze your images with:

CELLMIA

another image software

**New Project...**

You first choose a project to which the experiment is going to belong, and provide a number and a short description for the experiment (see the screenshot above, experiment metadata). If the experiment you want to add needs to be part of a new project (i.e. a project that is not present in the DB yet), you can add a new project from this interface.

Then, you can define the experimental set-up on a multi-well plate view (see the screenshot below, plate view and biological conditions). Common multi-well plate formats are rendered on the view, namely 96 (8 columns x 12 rows), 48 (6 x 8), 24 (4 x 6), 12 (3 x 4), 6 (2 x 3).

### EXPERIMENT MANAGER - plate view and biological conditions

**Project Experiment Miscellaneous**

**Experiment metadata**

**Project:** P000  
**Number:** E003  
**Purpose:** experiment purpose

**Conditions**

- Add condition
- Remove condition

**Condition 1** **Condition 2** **Condition 3** **Condition 4** **Condition 5**

**Plate**

Select a plate format: 96 (8x12)

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2												
3												
4												
5												
6												
7												
8												

**Randomize wells**   **Clear last selection**   **Clear all**

**Info**

Add conditions and select wells for each condition. Conditions details can be chosen in the right panel.

Different informative metadata variables can be added or chosen from drop down lists for each biological condition in the right panel of the GUI. Here, three tabs show different views according to the variables to be supplied (see the next three screenshots):

1. **Cell Line:** here the cell line used in the biological condition can be chosen; you can define and customize parameters such as seeding density (expressed in number of cells per well), growth medium and serum type and concentration. Cell lines can be added to a drop down list (and thus to the used **CellMissy** DB).
2. **Assay\_ECM:** here you can characterize the extracellular matrix (ECM) condition, the dimensionality (are you setting up a migration (2D) or an invasion (3D) experiment?), the

coating type (collagen, fibronectin), coating/matrix polymerization conditions (temperature, time...) and so on.

3. **Treatments:** the last tab gives you the possibility to specify the treatment or compound to which the cells were subjected in a given biological condition, for example the type and concentration of a drug, the presence of only drug solvent (e.g. in control condition), a protein overexpression or a siRNA treatment, etc. Treatments can be added to a drop down list (and thus to the used CellMissy DB).

**EXPERIMENT MANAGER – cell line properties**

Conditions Setup

**Cell Line** **Assay-Ecm** **Treatments**

**Choose a Cell Line**

Select a cell line	MDA-MB-231 ▾
Seeding Density	50000 cells/well
Seeding Time	day -1
Growth Medium	DMEM ▾
Serum	FBS hi ▾
Serum Concentration	10.0 %

**Add a new Cell Line**

If the cell line you want to use is not present, add it:

Cell line name  Add Cell Line

## EXPERIMENT MANAGER - assay-ECM properties

Conditions Setup

**Cell Line** **Assay-Ecm** **Treatments**

Select ECM dimension **2D**

Select a migration assay **scratch**

Extra Cellular Matrix

Composition **Collagen I (bovine)**  
(monomeric coating)

Add new composition

Concentration **0.04** mg/ml

Volume **100.0** µl

Coating time (min) **60**

Coating temperature **RT**

## EXPERIMENT MANAGER - treatments properties

Conditions Setup

**Cell Line** **Assay-Ecm** **Treatments**

Drugs

IPA3  
IPA5

Add >>

Treatments

Control + Drug Solvent  
WT

Remove <<

Control

Time of Addition **0 h**

Concentration **0** µM SFC\* **0** %

Drug Solvent

Assay Medium **DMEM**  
Serum **FBS hi** SC\* **1.0** %

Medium Volume **10.0** µl

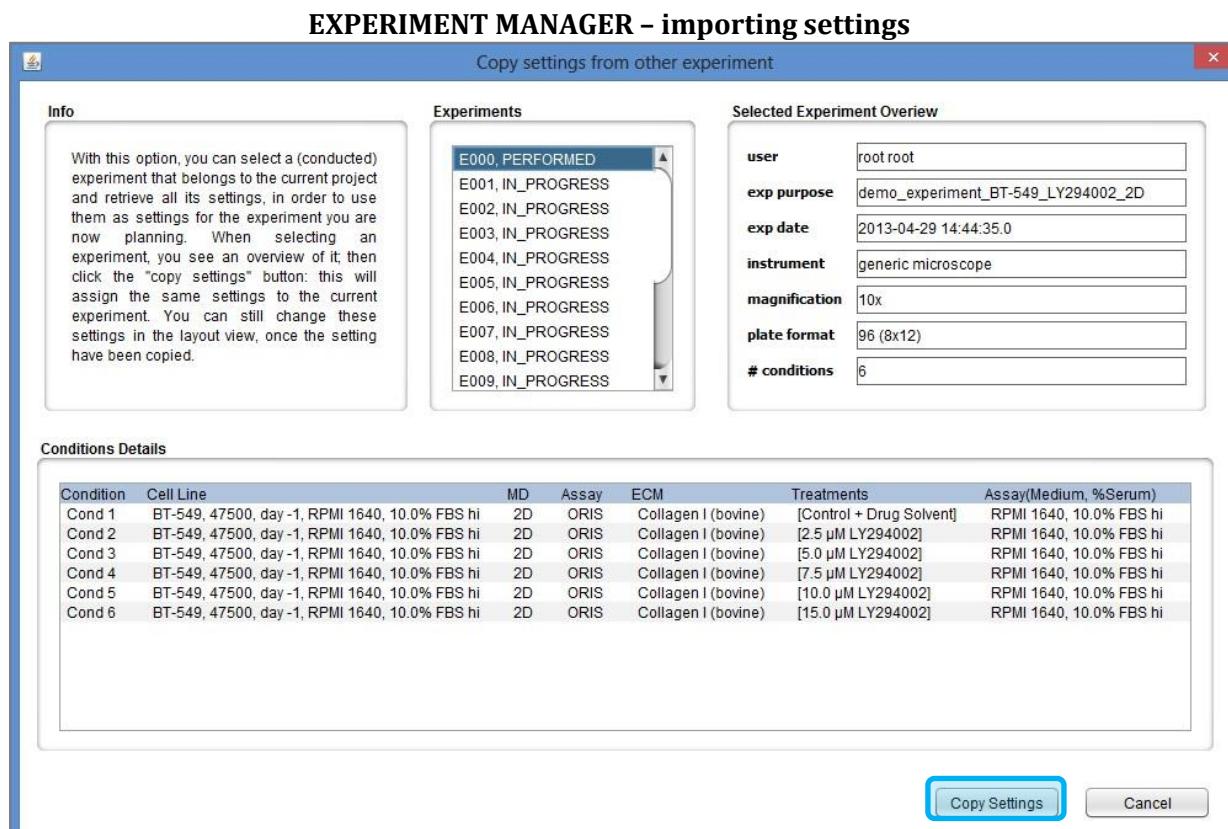
SFC\* = Solvent Final Concentration  
SC\* = Serum Concentration

Add new drugs/treatments...

Once this experimental set-up is finalized, you can export the design as a PDF document that can be used as a reference in the lab while conducting the experiment. Then, the plate lay-out, well assignments and related metadata are all stored in the database.

## 2.1.1 EXPERIMENT MANAGER – IMPORT SET-UP SETTINGS FROM ANOTHER EXPERIMENT

**CellMissy** provides the possibility, while setting-up a new experiment, to retrieve the settings from an experiment that was already planned (thus present in the DB) for the current project. You can easily use this functionality in the Experiment Manager module, through the “*Import Settings...*” button. Clicking this button will make a dialog appear, where you can select the experiment from which you want to retrieve the settings, as shown in the following screenshot. Clicking an experiment will render its details in the right panel, while a table with all the biological conditions details will be shown at the bottom of the dialog. Clicking the “*Copy Settings*” button will assign the selected experiment’s set-up to the currently planned experiment, and this will automatically update the plate lay-out, as well as the conditions list.



### 2.1.2 EXPERIMENT MANAGER – EXPORT A TEMPLATE TO AN XML FILE

Once the experimental set-up is finalized, just after you have created a PDF report, you can export the set-up of an experiment to an XML file, creating thus a template that can be exchanged and easily re-imported into **CellMissy** to reproduce someone else's workflow/setup. This can be done in the Experiment Manager module clicking the “*Export Template...*” button: all you have to do is choose a directory to save the XML file and click the “*save*” button. A name for the template file will be automatically created by **CellMissy**, of the type: *setup\_template\_ExXX\_PYY*, with *X* and *Y* to be replaced with the number of the experiment and the number of its project, respectively.

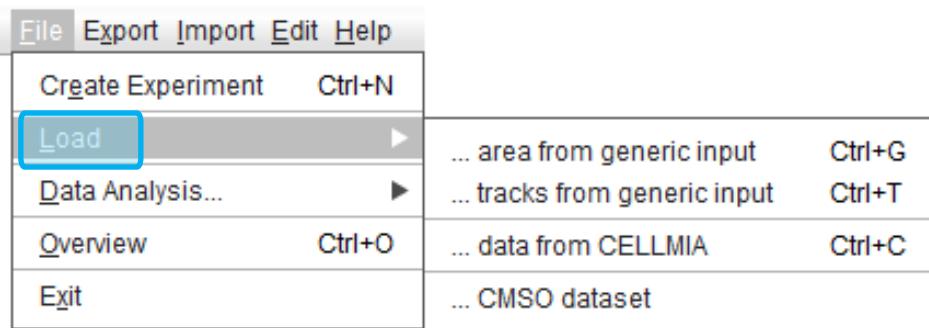
At <https://github.com/compomics/cellmissy/tree/master/supportFiles> you can find an example of this set-up template (the *setup\_template\_E000\_P000.xml* file).

The same function can be reached in **CellMissy** also through the *File* menu: in this case, you first need to choose an experiment from which you want to export the template, and then save the generated file to a directory.

### 2.1.3 EXPERIMENT MANAGER – IMPORT A TEMPLATE FROM AN XML FILE

What if you want to use the settings of an experiment that was not saved in your **CellMissy** database? If this is the case, during the set-up of a new experiment, you can import an external XML file containing the experiment template (a file generated in **CellMissy** as described in section 2.1.2 above). In the Experiment Manager, clicking the “*Import Template...*” button will let you choose an XML file to import; **CellMissy** will then retrieve the settings from this file and assign them to the new experiment. The plate lay-out and the conditions list will be then automatically updated.

## 2.2 DATA LOADER - LOAD MIGRATION DATA...



Once the experiment is performed and the acquired images are analyzed by the image processing software of your choice, you return to **CellMissy** to import the relevant cell migration data. Data import and storage in **CellMissy** are either based on using the generic migration input format (requiring you to connect text files with a replicate (i.e. a well in the experimental set-up) of a specific biological condition) or can be fully automated once tailored to a customized system. A third option is to (re-)analyze a CMSO-conform dataset. See the four following sections for more details.

### 2.2.1 ... from generic input - area

A wound healing-like experiment followed by image processing will generally result in a list of area values in time for each technical replicate of a biological condition. This can be for a limited number of time points or for a large number (e.g. based on a time-lapse experiment). The image processing will either have given you as output the area values of the wound/gap/cell-free zone (that decreases in time as cells in the sheet migrate) or the area of the cell-covered zone (that increases in time as cells in the sheet migrate). **CellMissy** is designed such that both types of area vs. time lists can be used as import data when present in a generic input file (see "*example\_dataset\_scratch*", see also below, Section 3).

For generic input, you first need to provide some experiment metadata: experiment duration, time interval, etc. (see next screenshot).

**DATA LOADER – experiment metadata**

**Experiment metadata**

Time Frames*	<input type="text"/>
Interval*	<input type="text"/> MINUTES <input type="button" value="▼"/>
Duration*	<input type="text"/> hours

i Please fill in experiment metadata.

---

**Info**

Select project/experiment in progress to load motility data; provide experiment metadata to start with the import.

Then, **CellMissy** expects for each well at least one tab-separated text file containing two columns, as shown in the following figure.

**INPUT DATA FILE – example**

1	time	area
2	0	0
3	30	4.155276505
4	60	9.919203631
5	90	19.28003667
6	120	23.29390134
7	150	29.54343138
8	180	37.01400293
9	210	46.8975504
10	240	48.85572839
11	270	54.60291895
12	300	61.74627854
13	330	64.44038024
14	360	72.94498036
15	390	77.04713647
16	420	79.31239403
17	450	85.06734648
18	480	87.10290076
19	510	89.78463195
20	540	92.40887669
21	570	94.6932964
22	600	96.33930109
23	630	96.4307458
24	660	98.60261818
25	690	99.00041478
26	720	99.10034904

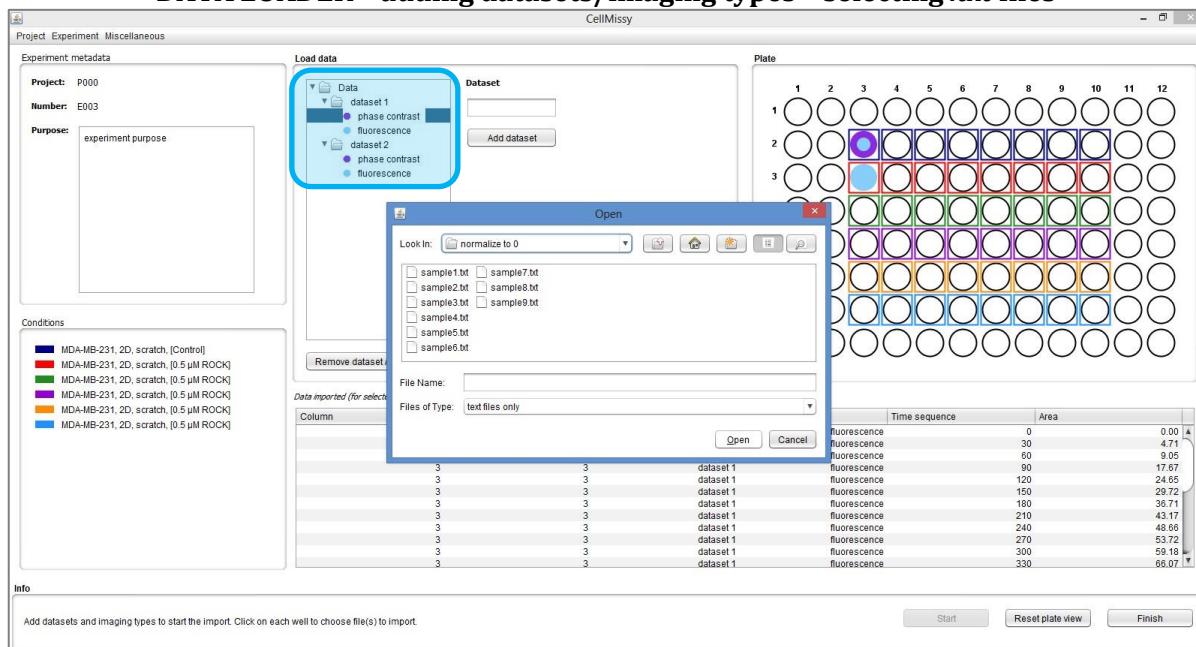
The first column holds the time information (this can be minutes or simply time sequence steps) and the second column the area values (these can be expressed in  $\mu\text{m}^2$ , pixels or area percentage %).

You start the import by adding names for datasets and imaging types of interest (e.g. phase contrast imaging, fluorescent imaging) (see screenshot below). One dataset could e.g. stand for a collection of data generated through a specific setting of the algorithm during image processing. Once an imaging type is added to the session, it is automatically added to each dataset. However,

a well (or a biological condition) may have been imaged with only a specific imaging technique, or a certain dataset may have been generated only for a specific group of conditions. That's why **CellMissy** does not require every combination of dataset/imaging type to be supplied for each well. Furthermore, you can load multiple files per well, even for a certain combination of dataset/imaging type, allowing the use of multiple imaging locations within the well.

A table underneath the plate view keeps track of the imported area values, together with current dataset, imaging type, and well's column, row. Once the import is done, and experiment metadata are supplied, you can store the data clicking the “*Finish*” button.

### DATA LOADER - adding datasets/imaging types – selecting .txt files



#### 2.2.2 ... from generic input - single cell

The module expects text files containing four distinct columns: (i) the trajectory unique ID, (ii) the time point, (iii) the x coordinate (either in pixel or  $\mu\text{m}$ ), and (iv) the y coordinate (either in pixel or  $\mu\text{m}$ ). Formats accepted: tab-separated-values, comma-separated-values, XLS. Once the experiment has been setup (using the dedicated **Experiment Manager** module), single-cell tracking data can be imported and stored into the relational database of **CellMissy**.

### INPUT DATA FILE - example

Track	time	x	y
1	0	333.6	707.8
1	1	331.909091	709.363636
1	2	333	710.5
1	3	332	710.5
1	4	331.846154	711.461538
1	5	331.615385	712.384615
1	6	332	712
1	7	331.925926	712.407407
1	8	331.875	713.125
1	9	332.888889	713.333333
1	10	333	713.2
1	11	333.166667	712.541667
1	12	335.628571	714.657143
1	13	347.6	718.933333
1	14	349.8	714.5
1	15	350.178571	706.892857
1	16	349.5	693.5
1	18	345.1	670.966667
1	19	355.419753	666.679012
1	20	371	666.5

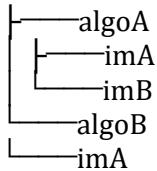
Header of the text file

Just follow these steps:

1. Click on the **Load Directory** button.

This expects you to point to a directory that looks like the following:

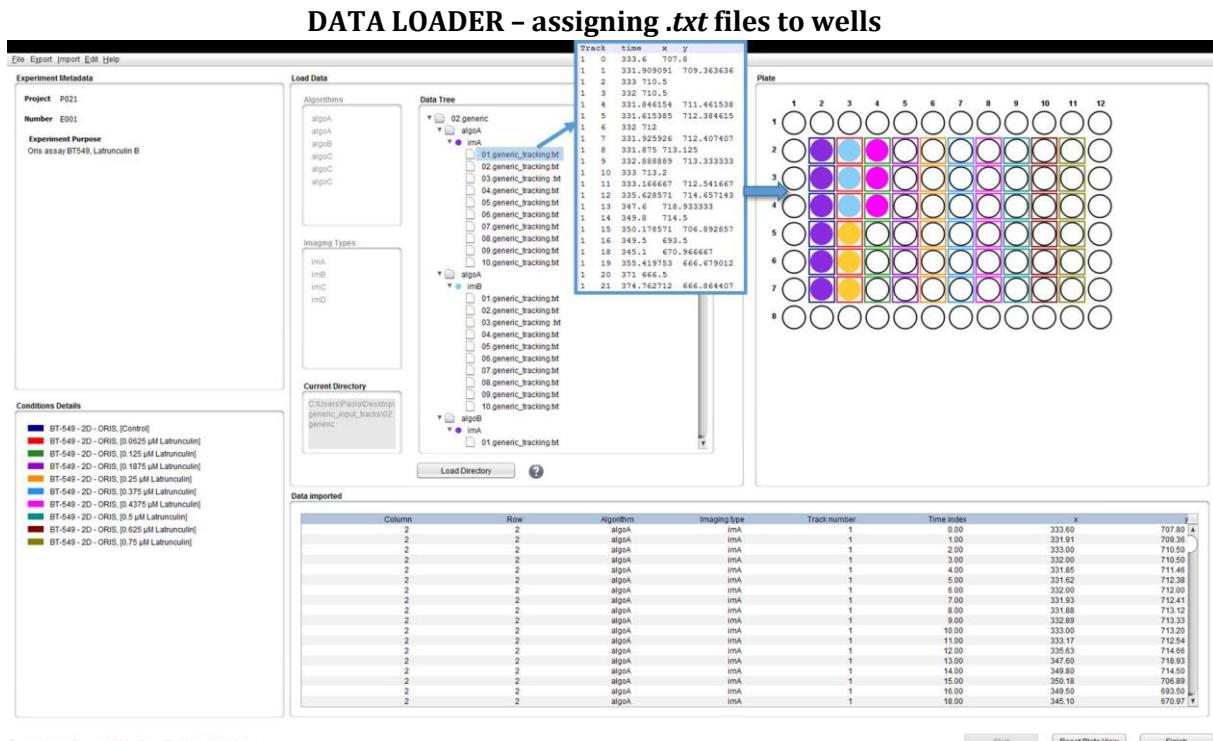
your\_folder:



In this case, you'll be loading two datasets (one processed with algoA, one processed with algoB), and, in particular, algoA will contain data imaged with imA and imB imaging techniques, while algoB will contain data imaged only with imA.

2. If the operation above was successful (this can only fail if your directory structure is wrong), the *Algorithms* and the *Imaging Types* will be automatically populated, and the *Data Tree* will be filled.

3. Then, you just need to drag and drop each text file to the well it belongs to. For each file successfully imported, the data table gets automatically updated. Finally, just click on the Finish button to store the data to the database.

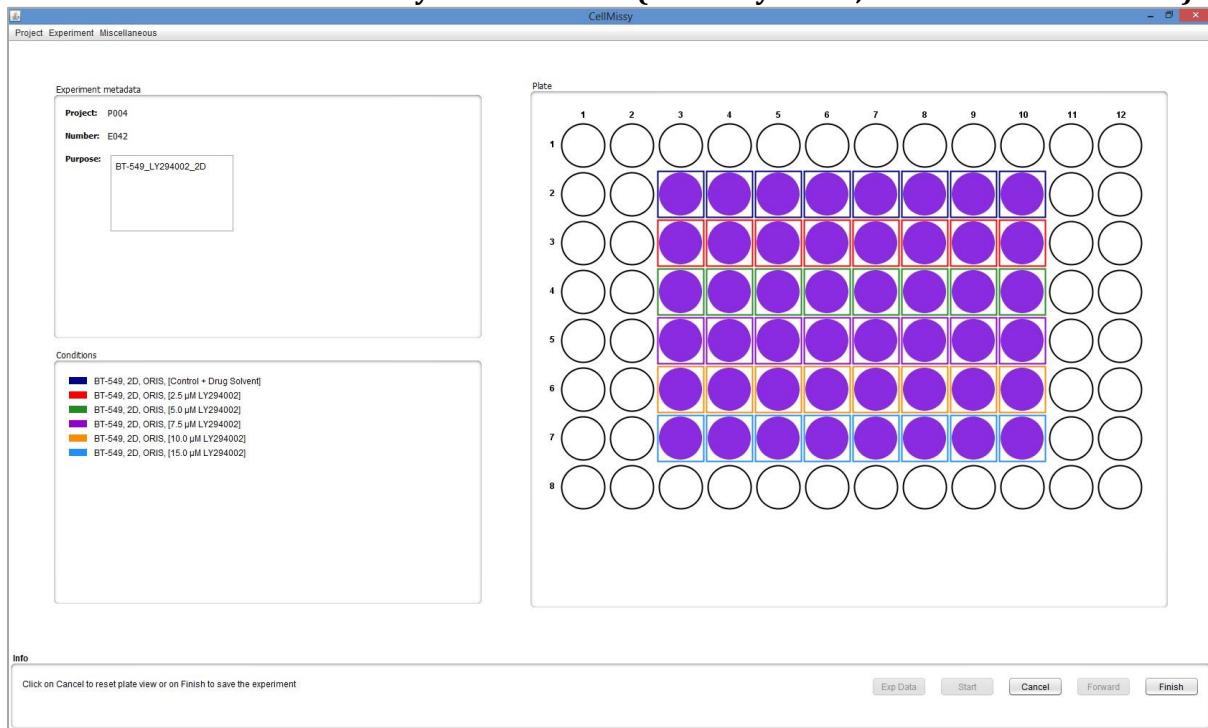


### 2.2.3 ... automatically from a customized combination of imaging system and image processing software

**CellMissy** has capabilities to automatically load data (i.e. the text files generated by the image processing software) by reading metadata coupled to the images, to the processed images and, when provided, with information on the imaging order of the wells. This makes it amenable to high-throughput data processing. In the current version, this is established for the customized set-up used in our group, i.e. an *Olympus xcellence CellM* system and custom imaging software we generated in collaboration with DciLabs Belgium (Van Troys et al., Analysis of Invasion Dynamics of Matrix-Embedded Cells in a Multisample Format. *Methods Mol Biol.* 2018).

Once you select the experiment for which you want to import and store the data, **CellMissy** looks for a file generated at microscope imaging time; this file has an *.obsep* extension (see <http://loci.wisc.edu/software/bio-formats> for further information), and contains the experiment metadata information, as well as the names and the metadata of imaging types used during the image recording and the names of the positions list(s) defined and generated. If the file is not in the right folder or more than one file is present, you can select a file to associate to the current session. Once these metadata are retrieved, **CellMissy** processes the data text files: all you have to do is to click on the well/sample that was first imaged; all the others wells will be automatically

## DATA LOADER – automatically from CELLMIA (Van Troys et al., *Methods Mol Biol. 2018.*)



highlighted (if imaged), according to the position list associated to the current imaging type (see screenshot above). The plate view can be reset at any time and the import can be cancelled for a certain batch of data. Once the import is finished, you can save the cell migration data to the database clicking the “*Finish*” button.

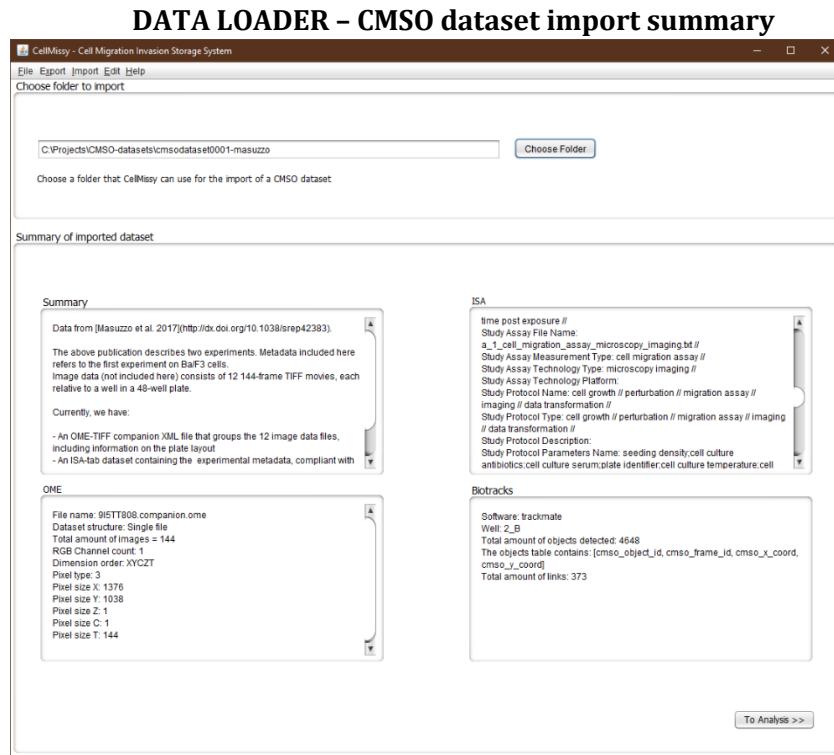
Of course, automated import from other existing image analysis tools will require the writing of a small piece of adapter code, and we foresee requests from the community to provide such interfaces, since we are fully committed to supporting such work, or undertaking it ourselves.

### 2.2.4 ... from a CMO-conform dataset

The purpose of standardization is to allow communication of data, results and metadata between researchers. In practice, however, data gathering, processing and formatting is performed by software, and standards are therefore typically handled by software applications rather than by researchers directly. As a result, it is important that data management software tools like **CellMissy** are conversant in the standards developed so that the data held by these tools can be assembled and formatted for direct submission to a standards-compliant repository. The opposite, where data in such a repository can be imported into a local data analysis tool for visualization and analysis, is also vital to enable re-use of the data.

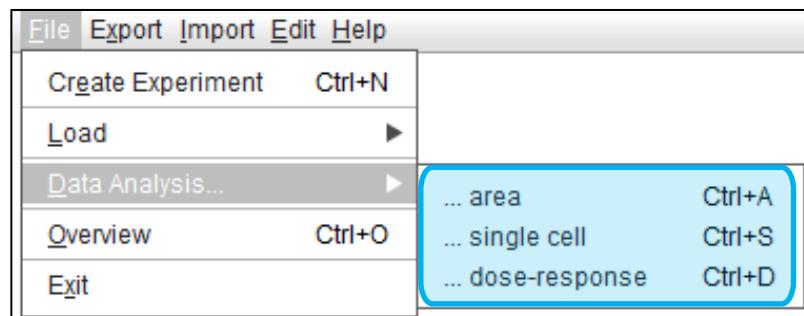
The Cell Migration WIS Repository (<https://repo.cellmigration.org>) is a community accessible cell migration data repository and knowledge-base, developed as part of the H2020 MULTIMOT

project, to facilitate data exchange for cell migration data. This repository is based on the development of extensible community standards by CMSO, the Cell Migration Standardisation Organisation (<https://cmso.science/>). As such, **CellMissy** has the capability to load and analyze datasets that comply with these standards. CMSO datasets are not saved to the database when loading. You can find the example dataset used in the screenshot below here: <https://github.com/CellMigStandOrg/CMSO-datasets/tree/master/cmsodataset0001-masuzzo>.



To load a CMSO dataset, select the folder containing the entire dataset. CellMissy will read all individual parts (experimental and imaging metadata, cell tracking data and a readme file) and will show an overview of these parts. You can then continue to the analysis of the single-cell tracks.

## 2.3 DATA ANALYZER – ANALYZE, EXPLORE AND REPORT DATA



Once cell migration data are loaded in the system, interpretation and reporting of these data can be performed by **CellMissy** in the Data Analyzer module.

### DATA ANALYZER -analysis preferences/experiment metadata

**Overview**

**Projects**  
P000  
P001

**Performed Experiments**  
E012

**Project Description**  
tesssssst

**Experiment Details**

User	root
Instrument	generic microscope
Number of time frames	108
Purpose	HT1080 2D single cell dose-response, Y27632. In a 48-well plate. 14 conditions, concentration ranged from 10 $\mu$ M to 100 $\mu$ M.

**Analysis Preferences**

Outliers Detection Algorithm: iQR\_R\_algorithm

Kernel Density Estimation: normal\_Kernel

**Metadata**

Dataset: MIA\_algo-1

Imaging type: BF

Coordinates unit of measurement: pixels

Select a dataset and an imaging type to analyze.

Specify also the unit of measurement for the tracks coordinates.

Please select a project and an experiment to visualize and analyse single cell data.

<<Previous    Next>>    Cancel    Start

Here, you select an experiment in the main view (left side above screenshot), and a small summary of it is provided. For the chosen experiment, you need to select a dataset and an imaging type of

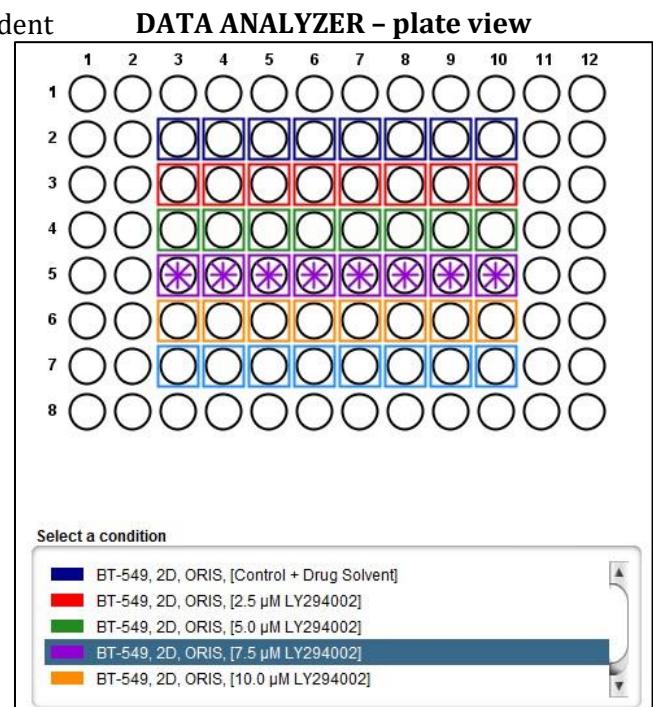
interest: cell migration data associated with these two will be loaded from the DB, allowing data inspection and further analysis.

Before starting an analysis session, you need to provide the unit of measurement for area data this can be  $\mu\text{m}^2$ , pixels or %; for single-cell it is pixels or  $\mu\text{m}$ . For area analysis, you also need to specify whether the measured area in the wound healing-like assay is the cell-covered area (i.e. related to the wound closure) or the open area (i.e. the wound area). Furthermore, you can here select some analysis preferences: the algorithm to use for the outliers detection, the distance metric (for area analysis quality control on technical replicates level), and finally the kernel function for the probability density estimation. Note that every new available implementation for these analysis features will be automatically picked up by **CellMissy** and presented to you in its interface. For more details on the software extension, see section 2.5.

The general workflow for the area data analysis is the same for both types of readouts to analyze, cell-covered area and open area, but open area values are first transformed to cell-covered area values and always expressed in area %.

The first two steps in the data analysis are performed on the level of one biological condition: the left side of **CellMissy** interface is here composed of a plate lay-out showing the set-up of the experiment, along with a list reporting the annotated biological conditions (colors, followed by details). As shown in the following screenshot, when clicking on a biological condition, the correspondent replicates (wells) are marked with a star.

In the following sections we go into the analysis possibilities of **CellMissy** and explain all steps for the area, single-cell and dose-response analysis.



### 2.3.1 Area analysis

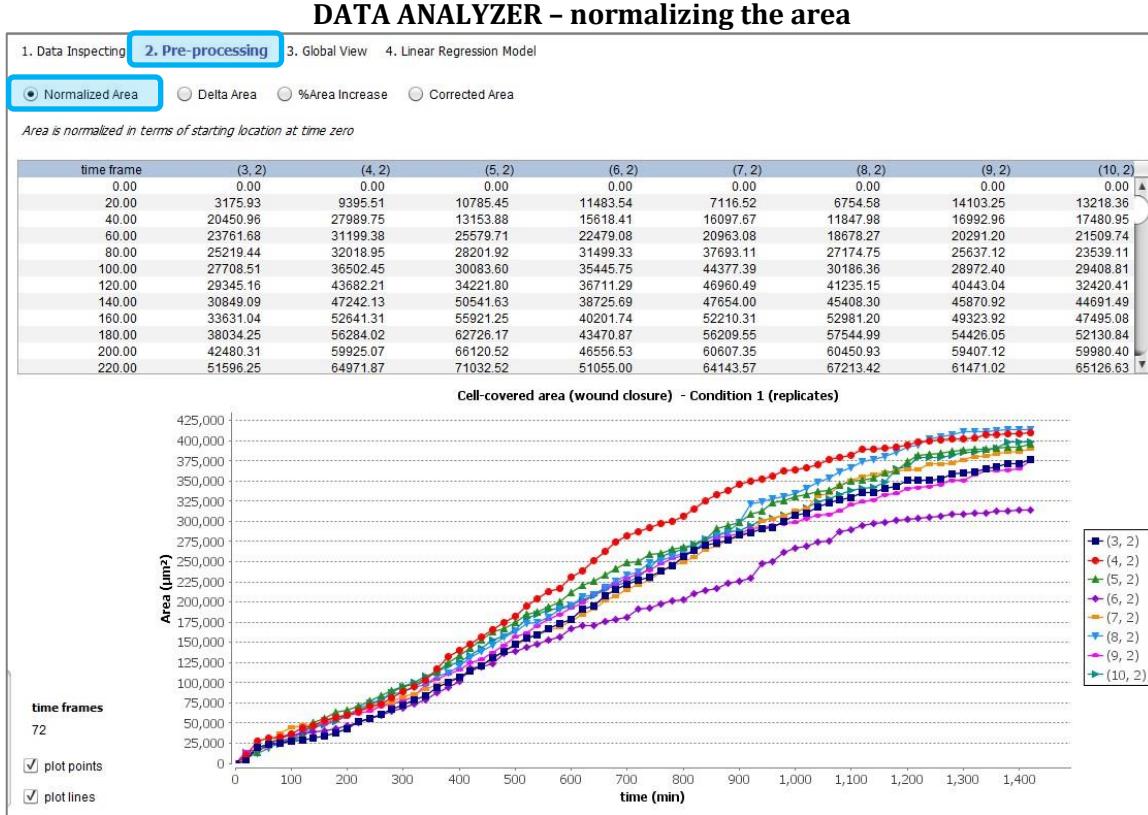
Area values are retrieved from the DB and thus explored through different steps, as detailed below and shown in the next screenshots.

- 1. Data Inspecting.** The first view encountered is simply a data inspector, where the area values are shown, together with time information and well's coordinates ((column, row), e.g. (3, 2)).

DATA ANALYZER - inspecting the raw data (area values)				
1. Data Inspecting    2. Pre-processing    3. Global View    4. Linear Regression Model				
Area values from database				
Column	Row	Time point	Area ( $\mu\text{m}^2$ )	
3	2	0	151010.42	▲
3	2	1	154186.35	
3	2	2	171461.38	
3	2	3	174772.10	
3	2	4	176229.85	
3	2	5	178718.93	
3	2	6	180355.57	
3	2	7	181859.50	
3	2	8	184641.46	
3	2	9	189044.67	
3	2	10	193490.73	
3	2	11	202606.67	
3	2	12	206556.00	
3	2	13	211866.56	
3	2	14	218099.86	
3	2	15	223595.55	
3	2	16	229228.94	
3	2	17	234768.73	
3	2	18	244381.41	
3	2	19	250884.71	
3	2	20	258428.49	
3	2	21	266072.95	
3	2	22	272332.46	
3	2	23	282425.23	
3	2	24	289758.92	
3	2	25	298133.08	
3	2	26	306671.58	
3	2	27	310457.41	
3	2	28	318464.64	
3	2	29	324264.03	
3	2	30	329701.48	
3	2	31	342434.34	
3	2	32	346156.93	
3	2	33	358616.05	
3	2	34	366317.08	
3	2	35	372322.41	
3	2	36	378102.24	
3	2	37	382314.49	
3	2	38	389461.80	▼

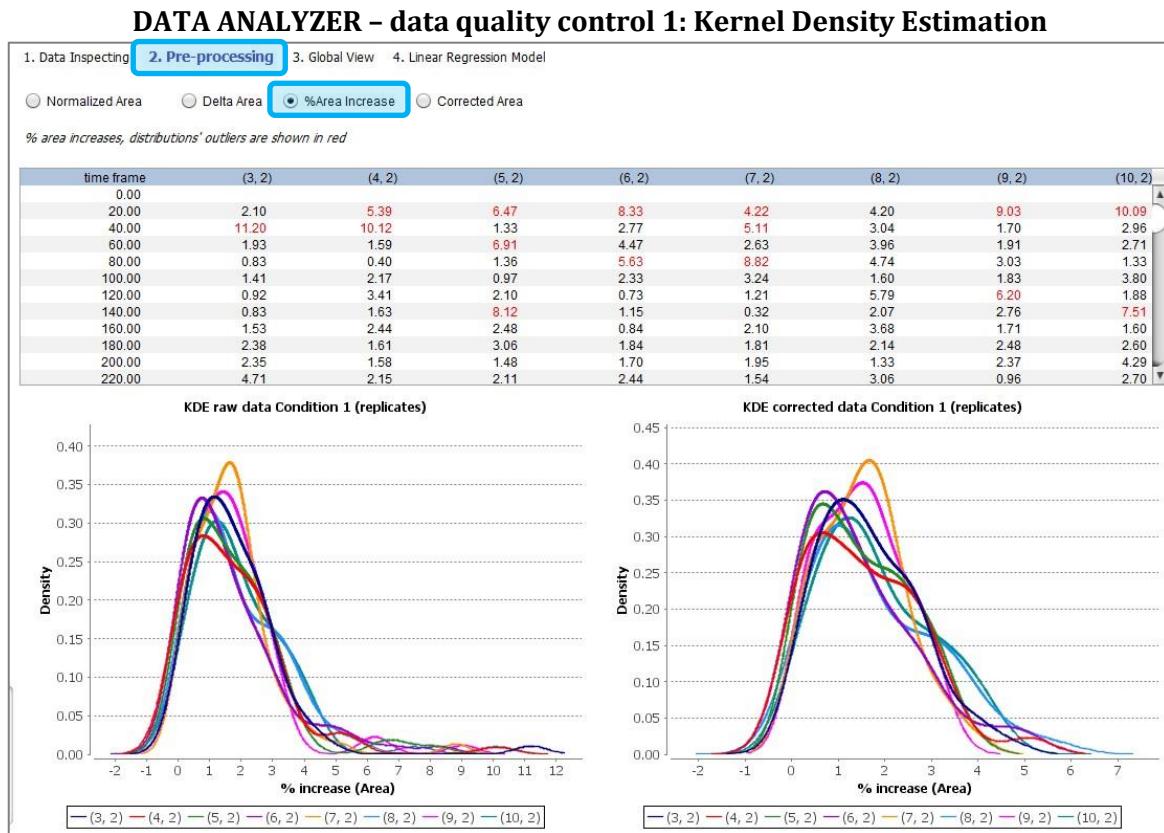
- 2. Pre-processing.** Under this second step pre-processing of the data is performed:

- Normalization. The raw data are first normalized relative to the measured area at time zero. If you have loaded open area (wound/gap), the start area (time 0) is set to 100 and values are normalized to this, and then expressed in percentage of cellcovered area (complementary to 100). If you loaded cell-covered area, the start area is set to 0 and values are normalized to this. Here you also see the plot in time of the normalized area values for each technical replicate of the selected biological condition.



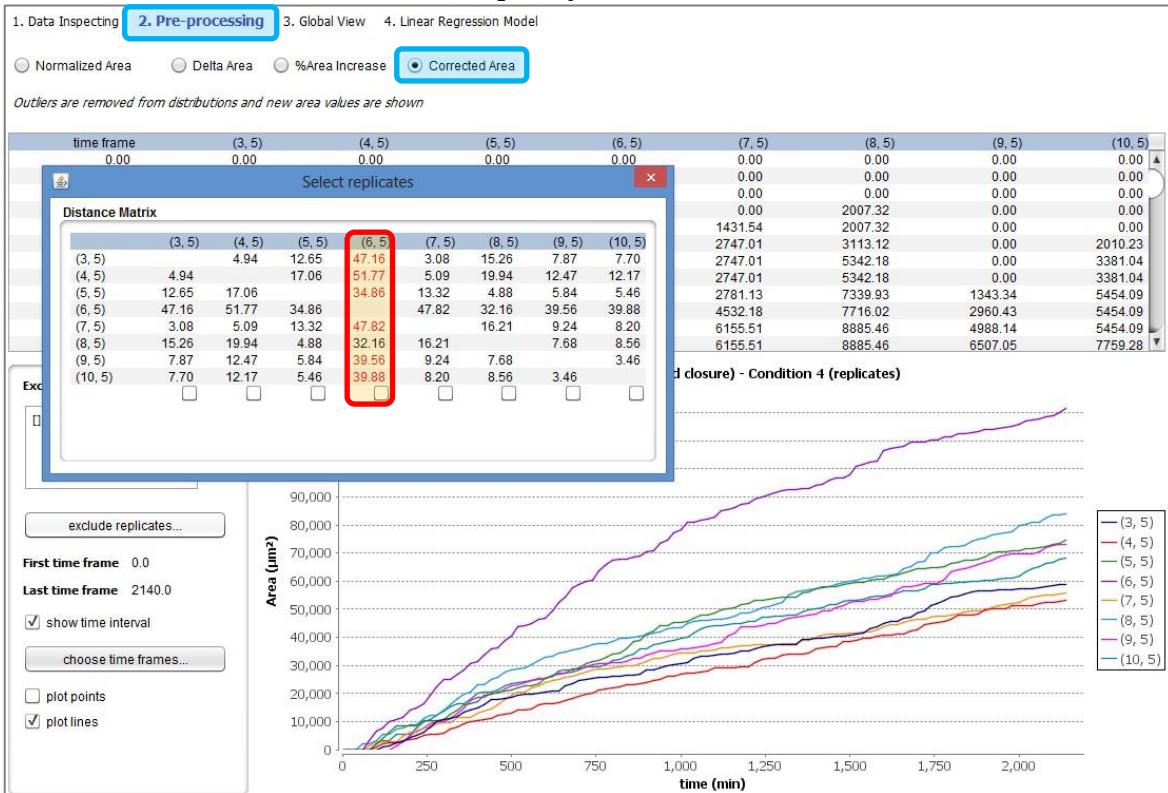
#### □ Data quality control.

- Delta area increments between time frame  $t(n)$  and  $t(n+1)$  are computed and shown, and percentages of area increase between consecutive time points are moreover calculated. Here, a first level of quality control takes place. For each well/replicate, area increases that are likely artifacts (possibly occurring due to experimental errors (e.g. cells or non-cell-objects in the cell-free zone) or due to false segmentation (e.g. when using fully automated image processing) are detected as outliers and corrected. This procedure is visualized using a Kernel Density Estimator (KDE) (see screenshot below, data quality control 1), plotting the probability density functions of all % area increases between consecutive time points for the different replicates. The left chart always presents raw data, and outliers are highlighted in red in the data table, while the right chart depicts corrected data. Note that this KDE is performed using a Gaussian kernel function, but CellMissy gives the possibility to extend this to other implementations.



- The corrected data are better visualized in this next step, where also the technical precision between replicates is examined (second level of quality control) using Euclidean distance (or any other distance metric you have selected in the analysis preferences list) as the similarity metric. The table here indicates for a specific replicate to what extent it resembles or deviates (indicated in red) from the other replicates in the biological condition.
- At both levels of quality control, you can either accept or decline the suggestion made by **CellMissy**. E.g. by unchecking the box beneath the replicate in the table you can decide to keep in a replicate designated as outlier (see screenshot below, data quality control 2).

## DATA ANALYZER - data quality control 2: Distance Matrix table

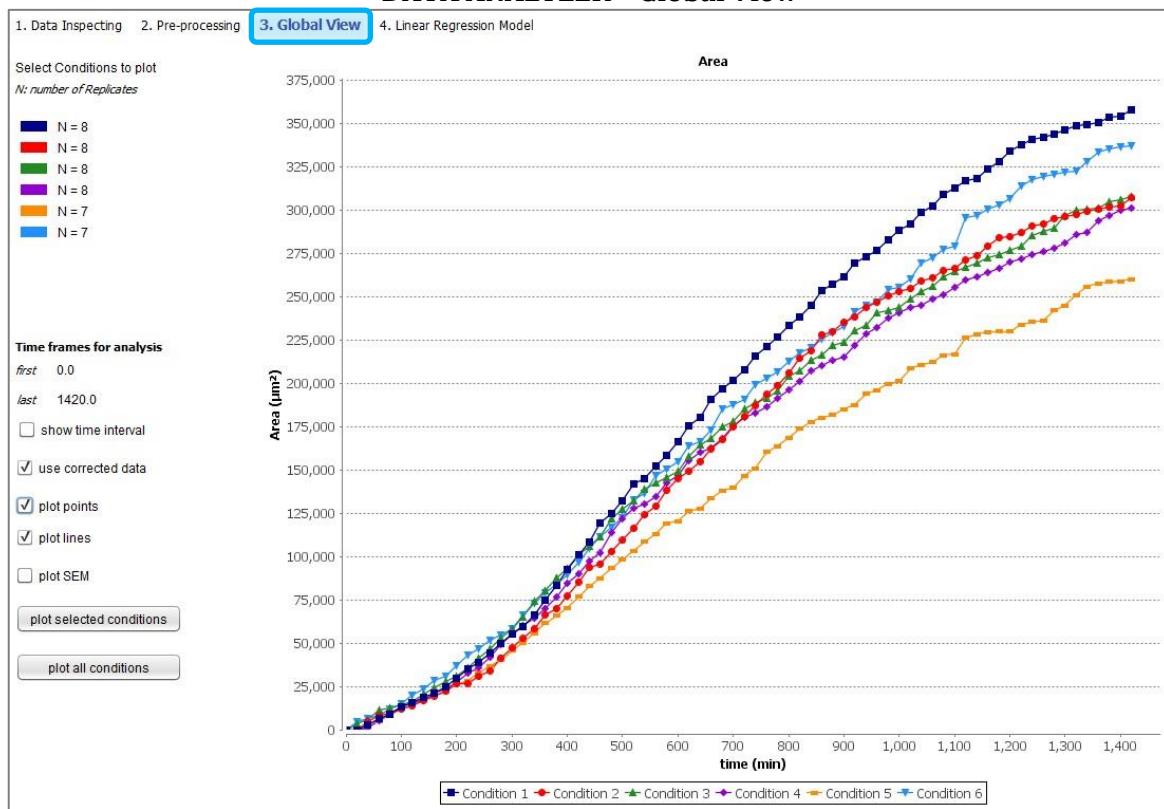


□ Finally, you can adjust the time interval in which you wish to perform data analysis.

Since data analysis implies the use of a linear regression model (see below), you optimally choose the time range where the area evolution is linear. You can change both the start and end time point (e.g. select the range 20-1000 minutes instead of 01600 minutes). Note that setting a limitation on the time range in one biological condition, automatically implies that only this time range will be considered in the subsequent analysis steps for all biological conditions in the experiment.

**3. Global view.** In the Global View step you pass from looking at one of the biological conditions to all the tested conditions (thus the entire experiment), therefore the list with the conditions is now disabled and you cannot interact with it anymore, unless you go back to previous steps of the analysis. You see now area over time plotted for each biological condition. This is based on plotting the median area across all the replicates for each biological condition. You can either use corrected data or retain the original raw data, and you can also select a subset of conditions you want to include into the plot. Standard error of the median can be plotted on top as well (by checking the corresponding box).

## DATA ANALYZER - Global View



### 4. Cell migration velocity calculation (based on Linear Regression Model) and statistical comparison of biological conditions.

Here, **CellMissy** currently makes use of a linear regression model to extract a slope and  $R^2$  of the area over time plot for each replicate in a biological condition. The median slope of the replicates provides then the median velocity of cell migration/wound closure for the biological condition. The bar chart in this view presents this median velocity computed for each biological condition (with standard error of the median). Again, you can also select a subset of conditions you want to include into the plot, just selecting the rows in the table.

Moreover, you can select a set of biological conditions on which you want to perform statistical analysis and provide a name for this defined group (see screenshot below).

## DATA ANALYZER - Linear Regression Model

1. Data Inspecting   2. Pre-processing   3. Global View   4. Linear Regression Model

**Linear Regression Table: slope + R<sup>2</sup>**

Cond	Repl 1	Repl 2	Repl 3	Repl 4	Repl 5	Repl 6	Repl 7	Repl 8	median	MAD
1	288.526 (0.984)	318.794 (0.961)	288.8 (0.985)	223.236 (0.989)	292.319 (0.988)	320.213 (0.992)	260.629 (0.983)	290.069 (0.992)	289.88	23.24
2	292.861 (0.984)	290.886 (0.981)	276.759 (0.986)	235.925 (0.986)	222.757 (0.977)	193.004 (0.965)	234.866 (0.991)	246.423 (0.968)	241.17	40.03
3	239.193 (0.985)	201.155 (0.977)	241.802 (0.989)	281.963 (0.988)	255.716 (0.95)	309.129 (0.988)	169.192 (0.947)	220.292 (0.966)	240.50	44.14
4	208.065 (0.954)	267.839 (0.973)	242.348 (0.996)	246.939 (0.991)	196.862 (0.969)	219.463 (0.971)	253.424 (0.983)	199.784 (0.987)	230.91	33.62
5	excluded	199.706 (0.987)	223.628 (0.995)	185.454 (0.993)	198.650 (0.977)	205.346 (0.988)	208.846 (0.991)	166.051 (0.958)	199.71	13.55

**Median Velocities**

**Median Velocity**

Velocity ( $\mu\text{m}^2/\text{min}$ )

Condition 1 Condition 2 Condition 3 Condition 4 Condition 5 Condition 6

**Statistics**

Type a name for the group:

Current analysis groups (name, conditions):

**Time frames for analysis**

first: 0.0  
last: 1420.0

Corrected data? YES

**Perform Statistical Analysis...**

**Create & Save PDF Report**

If you click the button “*Perform Statistical Analysis...*” a statistics dialog will pop up, with a summary statistics for all the conditions of the current group, and a table containing all the p-values generated by all pair-wise differences in median velocity using a MannWhitney U test. To correct for multiple hypothesis testing, you can either choose a Bonferroni or Benjamini-Hochberg correction. Both the statistical test and the multiple hypotheses correction methods can be easily extended with new implementation (for details, see section below 2.5). **CellMissy** provides three standard significance levels (namely 0.01, 0.05 and 0.1), and significant differences are highlighted in green in the pvalues table, as you can see in the following screenshot.

## DATA ANALYZER – Statistics: choosing statistical test, significance level and multiple comparison correction method

The screenshot shows the 'Statistics' window of the CellMissy application. At the top, it says 'Current analyzed group: 1'. Below that is a 'Statistics Summary' table:

	Max	Min	Mean	N	SD	Variance
Cond 1	320.2134	223.236	285.4358	8	31.4082	986.472
Cond 2	292.8611	193.0042	249.1851	8	35.153	1235.7353
Cond 3	309.1285	169.1918	239.8052	8	44.2419	1957.346
Cond 4	267.8393	196.8615	229.3404	8	26.7789	717.1103
Cond 5	223.6281	166.0515	198.2416	7	18.3005	334.9092
Cond 6	272.3479	231.8794	255.0184	7	15.2525	232.6376

Below the summary is a 'Mann-Whitney U test' table:

	Cond 1	Cond 2	Cond 3	Cond 4	Cond 5	Cond 6
Cond 1	-	-	-	-	-	-
Cond 2	0.1237	-	-	-	-	-
Cond 3	0.0823	0.7527	-	-	-	-
Cond 4	<b>0.0229</b>	0.4699	0.6917	-	-	-
Cond 5	<b>0.0134</b>	<b>0.0409</b>	0.0859	0.0921	-	-
Cond 6	0.0859	0.7805	0.5224	0.1239	<b>0.0262</b>	-

Below the tables are dropdown menus for 'Statistical Test' (set to 'mann\_Whitney\_Statistics'), 'Significance Level\*' (set to '0.05'), and 'Multiple Comparison Correction Method' (set to 'benjamini'). A note states: '\*p values smaller than significance level are shown in green'. A tooltip explains the Benjamini-Hochberg correction.

Note that every plot generated in **CellMissy** can be exported as a PNG file and can be easily customized through the *Properties* menu (just right click on a chart).

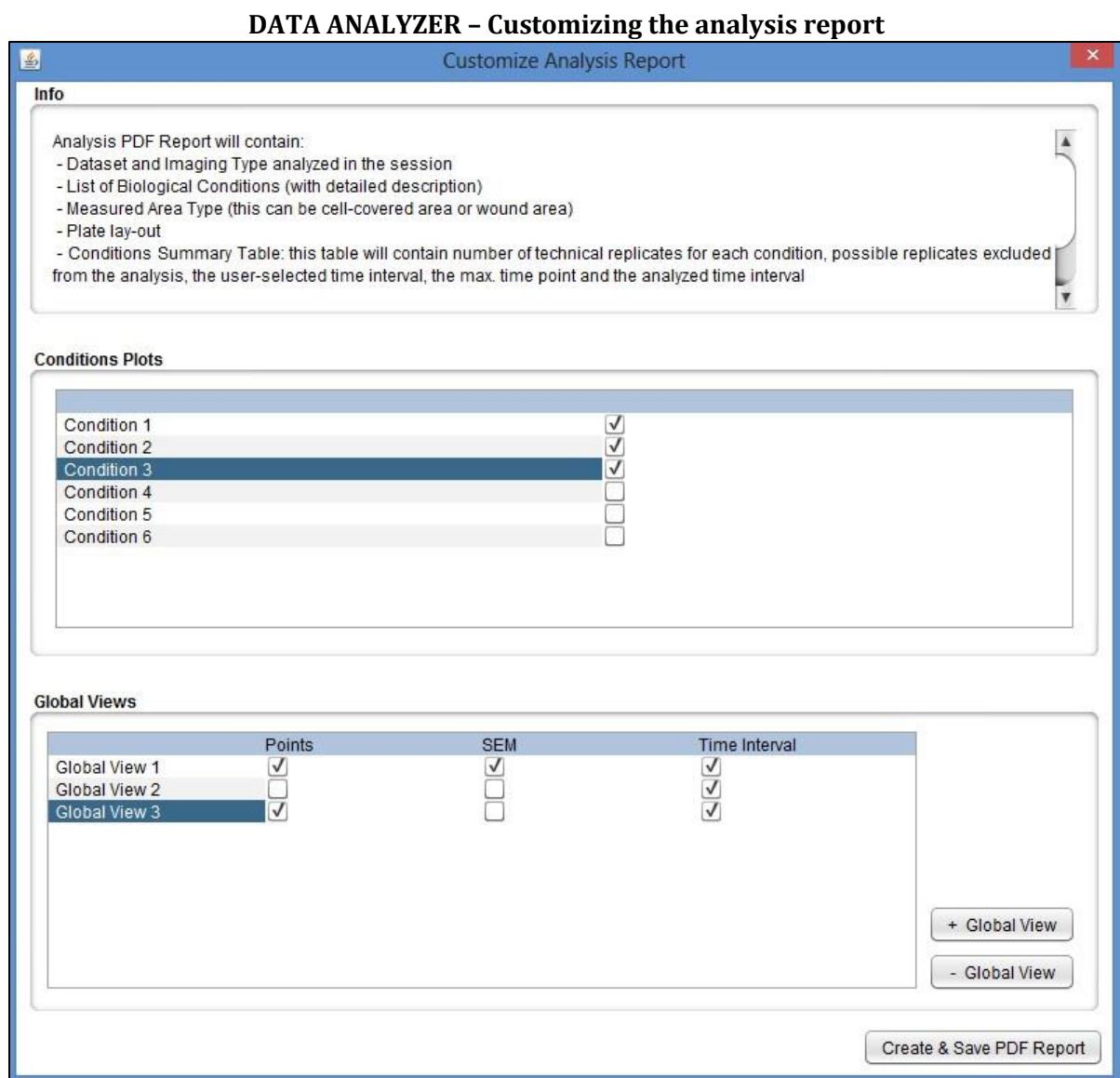
Finally, you can export a detailed *Analysis Report* in PDF that summarizes the experimental set-up, the data, results and statistics using text, tables and graphics (an example of this document is further shown in section 3.C).

This PDF will list the following features:

- Dataset and Imaging Type analyzed in the current session
- List of Biological Conditions, along with a detailed description
- Measured Area Type (this can be cell-covered area or wound area)
- Plate lay-out reporting the experimental set-up
- Conditions Summary Table: this table will contain number of technical replicates for each condition, possible replicates excluded from the analysis, the user-selected time interval, the max. time point and the analyzed time interval

You can furthermore customize the report through a dialog (see following screenshot) that will allow you to add:

- Conditions for which you want to add area-plots
- Global Views, i.e. plots reporting the overall set of biological conditions associated to the current experiment/analysis session



### 2.3.2 Single-cell analysis

Tracking data are retrieved from the DB and can be explored in many different ways, as detailed below and shown in the next screenshots. This addition to **CellMissy** has been described in Masuzzo et al., An end-to-end software solution for the analysis of high-throughput single-cell migration data. *Sci Rep. 2017.*

- 1. Data Inspecting.** The first view encountered is simply a data inspector, where the track numbers and lengths are shown, together with the well's coordinates ((column, row), e.g. (3, 2)). For every timepoint of the condition's tracks, x- and y-coordinates are shown after selecting a track in the upper table.

#### DATA ANALYZER – inspecting the raw data (single cell tracks)

1. Inspecting Data 2. Cell Tracks 3. Displacement/Speed 4. Turning Angle/Directionality 5. Distance 6. Filtering/QC 7. Conditions Analysis

Tracks from DB

Column	Row	Track Length	Track Number
2	2	12	1
2	2	12	2
2	2	16	3
2	2	21	4
2	2	18	5
2	2	40	6
2	2	15	7
2	2	32	8
2	2	16	9
2	2	11	10
	2	19	11

Track Points for selected track

Track	Time Index	x (pixels)	y (pixels)
1	0	341.71	1004.00
1	1	343.26	1012.31
1	2	344.38	1015.75
1	3	341.96	1018.96
1	4	343.05	1015.43
1	5	341.10	1012.08
1	6	341.36	1011.96
1	7	341.33	1013.25
1	8	337.23	1012.17
1	9	335.51	1009.43
1	10	335.19	1010.28

Info

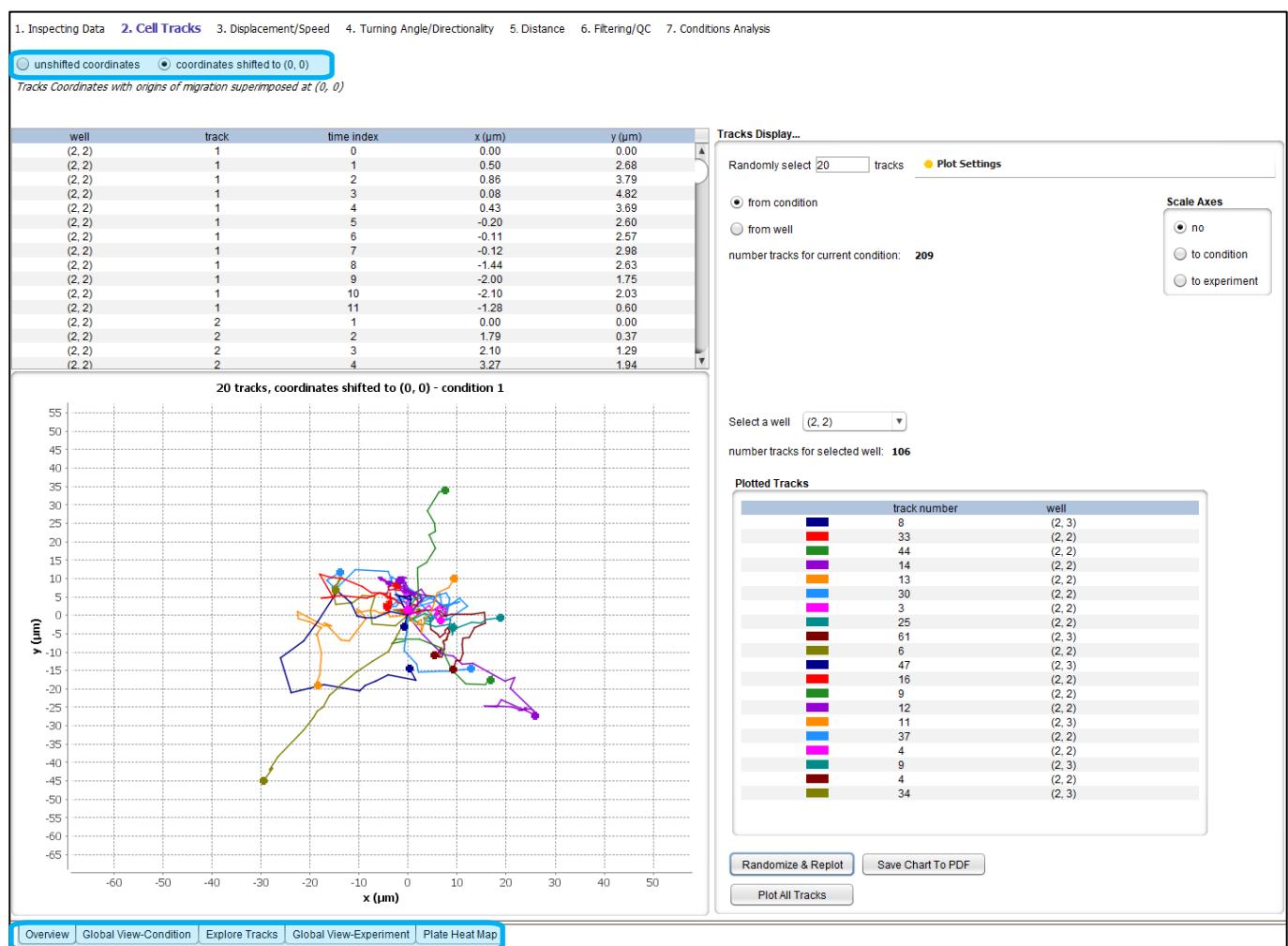
```

2019-08-28 13:52:31 - points: (146.254, 0.323), (146.416, 0.323), (146.577, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (303.474, 0.323), (309.763, 0.323), (309.924, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (303.474, 0.323), (309.763, 0.323), (314.117, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (146.577, 0.323), (146.738, 0.323), (146.738, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (146.577, 0.323), (146.738, 0.323), (146.738, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (146.577, 0.323), (146.738, 0.323), (146.899, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (270.74, 0.323), (275.416, 0.323), (277.19, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (270.74, 0.323), (275.416, 0.323), (285.091, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (270.74, 0.323), (275.416, 0.323), (288.155, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - points: (240.909, 0.323), (242.844, 0.323), (242.844, 0.323) are collinear on the plane.
2019-08-28 13:52:31 - generating instantaneous displacements...
2019-08-28 13:52:31 - generating directionality ratios...
2019-08-28 13:52:31 - generating median directionality ratios...
2019-08-28 13:52:31 - generating track displacements...
2019-08-28 13:52:31 - generating cumulative distances...
2019-08-28 13:52:31 - generating euclidean distances...
2019-08-28 13:52:31 - generating track speeds...
2019-08-28 13:52:31 - generating track end-point directionality ratios...
2019-08-28 13:52:31 - generating turning angles...
2019-08-28 13:52:31 - generating median turning angles...

```

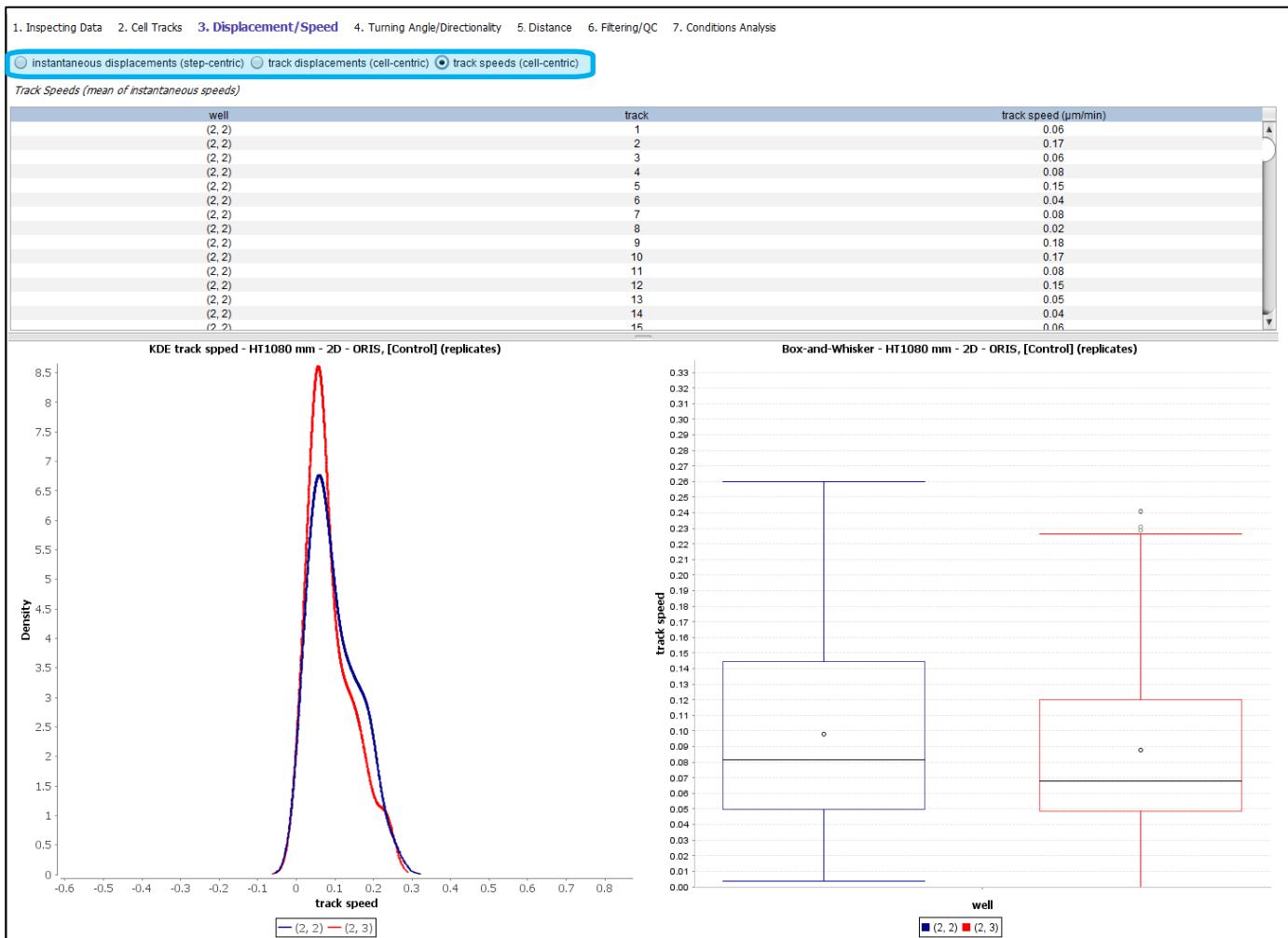
**2. Cell Tracks.** Here, single-cell trajectories imported in CellMissy can be visualized for each replicate (i.e., each well), or across all replicates for a specific biological condition. These plots can either use all trajectories, or only a random subset (with a user-defined sample size) in case of very large data sets. The visualization can either be based on the raw cell trajectories, or on re-centered trajectories where the starting point of each track is located at the origin of the coordinate system (usually depicted in the literature as a ‘rose plot’). The visualization of rose plots for all the biological conditions of an experiment, with identical axis scaling as provided by the module, provides a fast yet powerful way to visually compare migration behavior in diverse settings. Furthermore, the multiwell plate of the experiment under analysis can be visualized as a heat-map where each well in the plate color encodes one of the four parameters: (i) number of cell trajectories, (ii) mean or median cell speed, (iii) mean or median cell directionality, and (iv) robust z\*-score.

### DATA ANALYZER – exploring the cell tracks

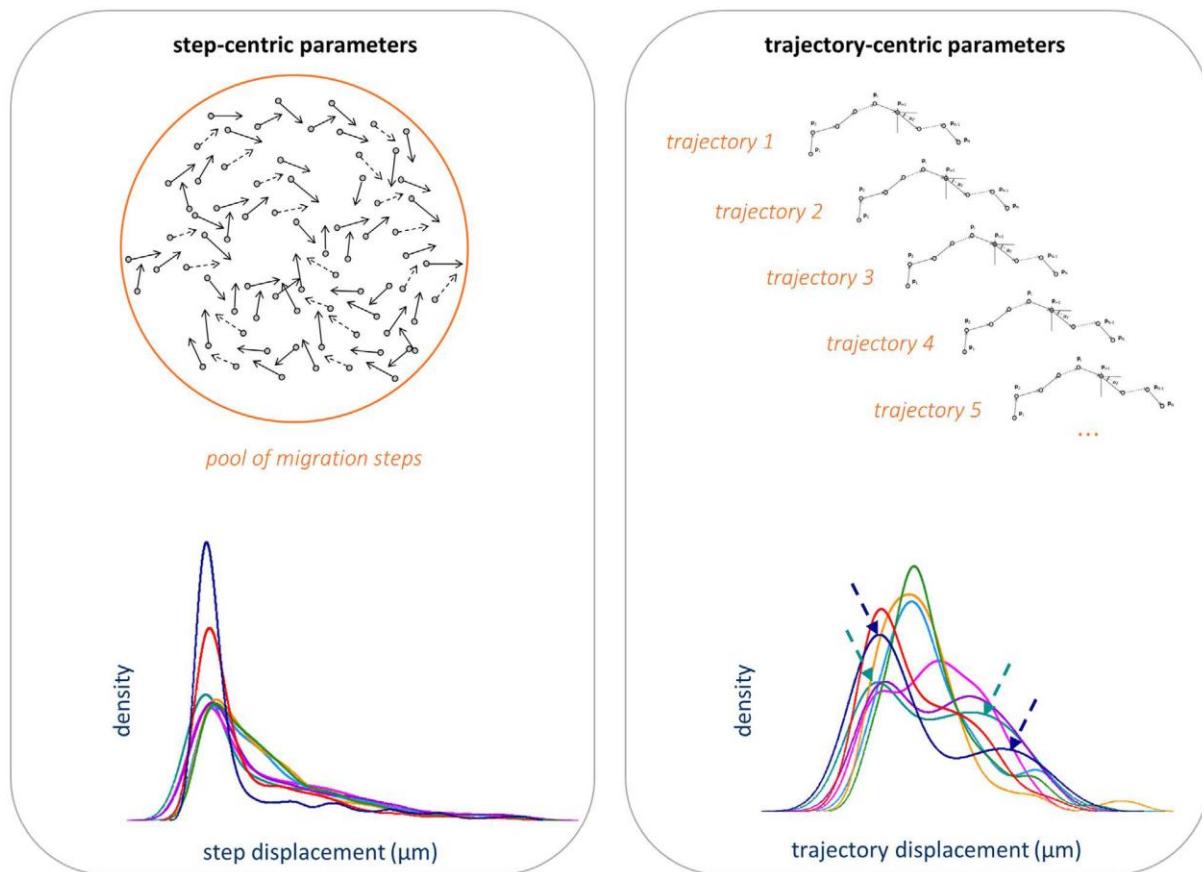


**3. Displacement/Speed.** There are generally two different ways to compute cell migration parameters and therefore conduct data analysis, as schematically shown in the figure below. On the one hand, you can first carry out a calculation of a specific parameter for each tracked cell separately (these are ‘trajectory-centric’ parameters), and then compute an aggregated value from these distributions (mean or median). On the other hand, a distribution and an aggregated value (mean or median) of all separate migration steps in a data set can be calculated, independent of which cell trajectory migration steps belongs to (these are ‘step-centric’ parameters). Both these quantification methods have their advantages and disadvantages and can therefore complement each other. Consequently, it is useful to combine these in a complete cell migration data analysis pipeline. In **CellMissy**, the distributions of the step-centric and trajectory-centric displacements, and of the cell speed for each condition at the replicate level are reported using boxplots and kernel density estimations. The latter are particularly useful for differentiating between unimodal and multimodal distributions.

## DATA ANALYZER – displacement and speed



**'Step-centric' and 'trajectory-centric' cell migration parameter computation:  
complementary analysis approaches.**



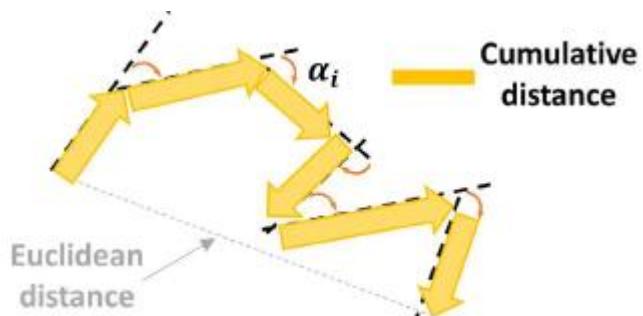
Top panels: schematic representation of the two approaches. Bottom panel: kernel density estimation (KDE) plots of step-centric and trajectory-centric cell displacements  $d_i$ : the possible presence of cell subpopulations (shown by arrows for two distributions in lower right panel) is overlooked in the step-centric approach. The density plots estimate the probability density function of the variable, i.e., the relative likelihood for the variable to take on a given value. The probability of the variable to fall within a particular range of values is given by the integral of the variable's density over that range, corresponding to the area under the curve.

**4. Turning angle/Directionality.** Cell turning angles  $\alpha_i$  between successive time frames are computed (both step-centric and trajectory-centric) as a speed-independent indicator for directionality, and are presented through traditional as well as angular histograms. Furthermore, ep\_dr, end-point directionality ratios (also known as confinement ratios or meandering indices) are computed trajectory-centrally as ratios between the net distance and the cumulative path traveled by the cell. These can be visualised as histograms, polar-, rose- and compass plots.

## DATA ANALYZER - turning angles and directionality



**5. Distance.** In this panel, you can explore boxplots of the Euclidean and cumulative (accumulated) distance travelled by the cells. These features are inherently trajectory-centric and both give meaningful information about the migration characteristics of cells. The figure below shows a schematic representation of Euclidean and cumulative distance travelled by the cell. The turning angles  $\alpha_i$  between successive time frames are shown.

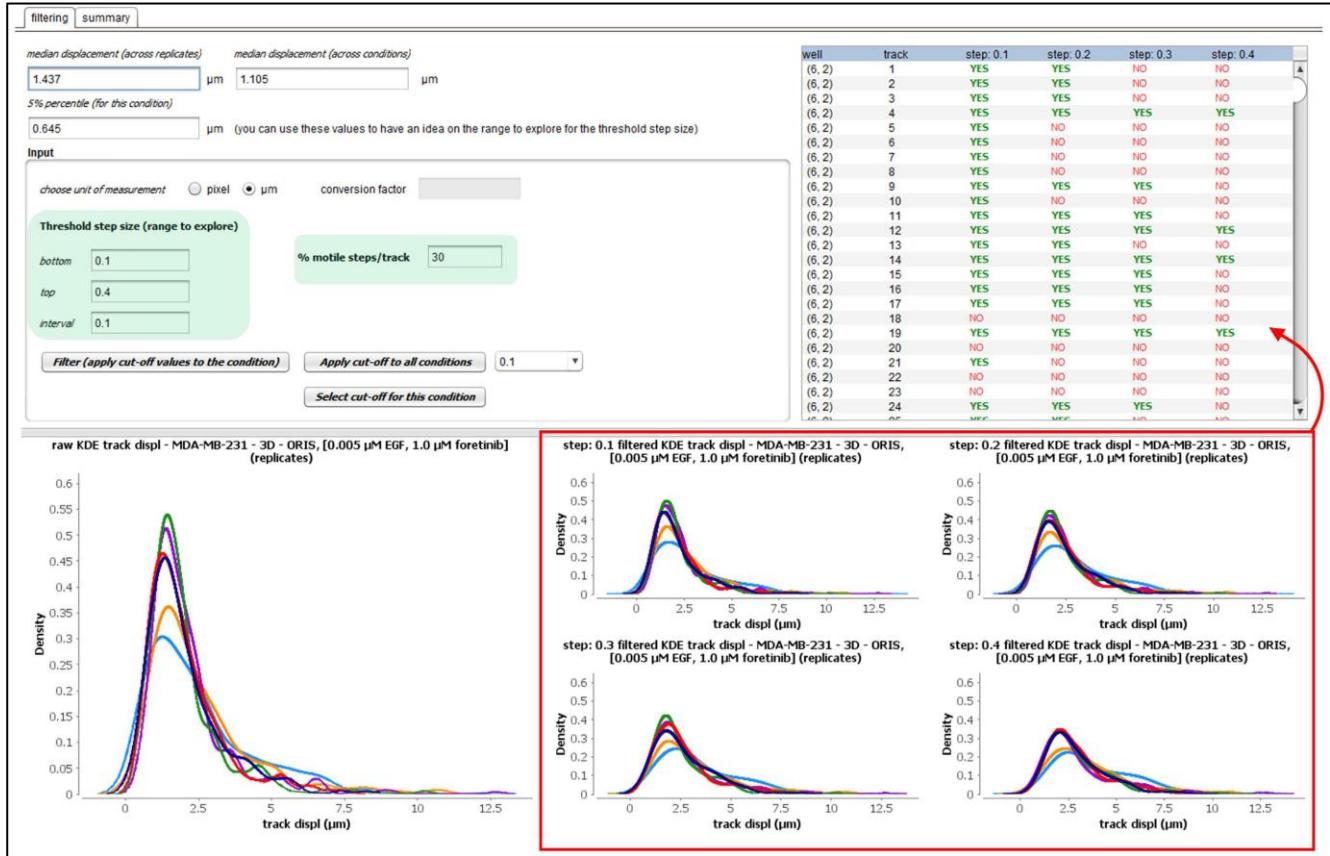


**6. Filtering/Quality Control** Here, you can filter your data using **CellMissy's** two-step filtering approach for quality control. It is implemented with a minimal-motility filtering: only trajectories that meet a tuneable two-step filter for minimal cell motility are retained for final analysis. The first criterion in this two-step filter is step-centric: cell steps are either labelled as ‘true’ or ‘false’ if these are larger or smaller than a minimal translocation, respectively. This minimal translocation can be specified in pixels or  $\mu\text{m}$ . The second filter criterion is trajectory-centric: trajectories are only retained if they meet the first criterion in at least a specified percentage of their steps (e.g., 30% motile steps/trajectory). Based on our experience with different cell lines and inhibitors, a minimal translocation/step of 0.1–0.5  $\mu\text{m}$  (for immune cells) and 1.6–2  $\mu\text{m}$  (for cancer cells), and between 20 and 30% of motile steps/trajectory appear suitable defaults for these two criteria. Of course, these criteria may vary depending on the effect of a treatment, e.g., when using a strong inhibitor of cell migration.

This flexible, two-step filtering approach can outperform a single, low cut-off-value for either mean trajectory or cell displacement when it comes to excluding true artifacts. In fact, this two-step filter allows the frequently observed phenomenon of cell pausing during migration to be taken into account, and will retain living cells that are only motile in part of the time sequence. Nevertheless, the module also provides the simpler, single-value thresholding option, where all trajectories of cells whose median displacement is below a user-defined threshold are discarded (this single threshold is then applied across all the data sets for the overall experiment).

Evaluation of the different settings can be performed by visually inspecting the kernel density functions of the corresponding displacement distributions, and population trajectories. The module presents an overview of these trajectories, highlighting which ones would be retained (marked ‘yes’ in green), and which would be removed (marked ‘no’ in red) (see screenshot below). This allows you to qualitatively judge if filtering is useful, with the goal of removing artifacts and dead cells. A desired level of filtering can be applied for each condition separately, or be set globally for all conditions. Furthermore, a summary is provided for each condition, detailing the number and the percentage of retained cell trajectories, to aid the selection process.

## DATA ANALYZER – single cell filtering and quality control



**The quality control view of the software: two-step thresholding.** The user can input a range for the threshold step size (i.e., displacement) to be tested (box on the left). The choice is guided by displaying the median and 5% percentile of the displacement distribution for the condition (across replicates) and the experiment (across conditions). Second, the user inputs a percentage of desired motile steps per cell trajectory (box on the right). The effect of each tested threshold value for the displacement can be evaluated through the density plots of the track displacement distributions (red box). For each tested value (columns), the table reports the list of cell trajectories (rows) marked in green (YES) if retained, and in red (NO) if not retained. For instance, in the example in the screenshot, a minimal step threshold of 0.4 μm, together with a minimum of 30% motile steps, removes more than 30% of the tracks for the untreated cells, which indicates over-filtering.

**7. Conditions Analysis** At this stage, you can still choose to analyze the raw data (i.e., without filtering), or proceed with the filtered data. First, you can explore the cell tracks for the entire experiment. Then, for each biological condition, statistics on speed, turning angle, and directionality ratio are reported. Subsequently, the user can select a set of conditions on which they wish to perform statistics. The user can choose the parameter to test (currently speed, directionality ratio, accumulated- or Euclidian distance), and can furthermore opt to correct for multiple hypotheses testing with either Bonferroni, or Benjamini-Hochberg correction. The available tests are a t-test and MannWhitney U test. A table reporting p-values is then shown, significant p-values are shown in green. Finally, the purpose of the normality tests tab is to allow the user to get a general idea about the distribution of the variables you can test for in the statistics tab (for each condition separately, select one on the left-hand side). An Anderson-Darling test assesses normality on the data and the skewness and *excess kurtosis* value are shown, together with a general conclusion that can be drawn from that value (e.g. not normally distributed or left skewed). Some information about these values and conclusions can be found when the user clicks on the question mark icon. At the bottom of the view a QQ-plot is displayed. The QQ-plot is important to assess the normality of the variables given that normality tests themselves can sometimes be unreliable, especially for large data sets.

## DATA ANALYZER – condition-wide overview of (filtered) tracks



## DATA ANALYZER - statistics on chosen (filtered) conditions

1. Inspecting Data 2. Cell Tracks 3. Displacement-Speed 4. Turning Angle/Directionality 5. Distance 6. Filtering/QC 7. Conditions Analysis

Cell Tracks  Cell Speeds/Angle/Directionality  Statistics  Normality Tests

**Input**

Choose conditions to add to group:

- HT1080 mm - 2D - ORIS, [20,1]
- HT1080 mm - 2D - ORIS, [25,1]
- HT1080 mm - 2D - ORIS, [30,1]
- HT1080 mm - 2D - ORIS, [35,1]
- HT1080 mm - 2D - ORIS, [40,1]
- HT1080 mm - 2D - ORIS, [50,1]
- HT1080 mm - 2D - ORIS, [60,1]

Add Group >> << Remove Group

name for the group:

**Statistics Option**

Parameter to test for: cell speed

Statistical test: mann\_Whitney...

Significance level: 0.05

Multiple comparisons correction: benjamini

Perform Statistics

Current Group Analyzed: increments

**Statistics**

	Max	Min	Mean	N	SD	Variance
Cond HT1080 mm - 2D - ORIS	0.26	0.0144	0.1004	190	0.0577	0.0033
Cond HT1080 mm - 2D - ORIS	0.3792	0.0186	0.1535	234	0.082	0.0067
Cond HT1080 mm - 2D - ORIS	0.4185	0.0162	0.1447	335	0.0837	0.007
Cond HT1080 mm - 2D - ORIS	0.3861	0.0188	0.1461	426	0.0873	0.0076
Cond HT1080 mm - 2D - ORIS	0.3319	0.0175	0.122	269	0.0841	0.0071
Cond HT1080 mm - 2D - ORIS	0.3682	0.0155	0.1278	490	0.0808	0.0065

**Pairwise Comparisons**

	Cond HT1080 mm - 2D - ORIS					
Cond HT1080 mm - 2D - ORIS	-	-	-	-	-	-
Cond HT1080 mm - 2D - ORIS	0	-	-	-	-	-
Cond HT1080 mm - 2D - ORIS	0	0.2215	-	-	-	-
Cond HT1080 mm - 2D - ORIS	0	0.2217	0.9706	-	-	-
Cond HT1080 mm - 2D - ORIS	0.1948	0	0.0008	0.0005	-	-
Cond HT1080 mm - 2D - ORIS	0.0021	0.0001	0.005	0.0032	0.2655	-

## DATA ANALYZER - normality assessment of migration features

1. Inspecting Data 2. Cell Tracks 3. Displacement-Speed 4. Turning Angle/Directionality 5. Distance 6. Filtering/QC 7. Conditions Analysis

Cell Tracks  Cell Speeds/Angle/Directionality  Statistics  Normality Tests

Accumulated Distance Euclidian Distance Directionality Speed

Condition: HT1080 mm - 2D - ORIS, [Control]

Anderson Darling: p-value: 6.263086420046537E-6 outcome: Not normally distributed

Skewness: value: 2.384223539504708 outcome: Right skewed

Kurtosis: value: 7.736710436241237 outcome: Leptokurtic

**QQPlot**

Tests on the shape and normality of the parameters

The following tests try to give an indication about the distribution of the data.  
The Anderson-Darling test is a normality test.  
The skewness and kurtosis give an indication about the shape of the data.

The different outcomes for the kurtosis are mesokurtic, leptokurtic and platykurtic.  
Mesokurtic means that the distribution has a kurtosis around zero, matching that of the normal distribution.  
This means that the given data has a similar distribution of extreme values as the Gaussian distribution.  
Positive excess kurtosis means that the distribution is leptokurtic, which means that it has fatter tails than the Gaussian distribution.  
Negative excess kurtosis means that the distribution is platykurtic, which means that it has thinner tails than the Gaussian distribution.

Please note that these tests only give an indication about normality. Inspection of the QQ-plot is necessary to conclude (non-)normality.

### 2.3.3 Dose-response analysis

This type of analysis can be carried out on two types of input: either a collective cell migration experiment as stored in **CellMissy**'s relational database, or a tabular file with compound doses and measured responses (does not have to be migration data). This choice in input data is of functional relevance, as it allows processing of e.g. toxicity or proliferative effects of potential drugs, alongside their migratory effects, all in a single tool.

**1. Data import (only applicable to generic data)** You can upload a simple 2+ -column datatable to analyze, the first column being doses and the second and following columns being the response values. This data will not be saved to the database because the database is setup for migration data only and these should be imported via the proper methods in CellMissy. You can (and should) provide additional information about the data and experiment to include in the analysis report at the end.

The screenshot shows the CellMissy software interface for 'Dose-Response' analysis. The window title is 'CellMissy - Cell Migration Invasion Storage System'. The menu bar includes 'File', 'Export', 'Import', 'Edit', and 'Help'. The main area has tabs '1. Data Loading' (selected) and '2. Dose-Response'. A sub-instruction 'Load the dose-response data you want to analyze' is present. A 'Choose a File' section contains a 'choose file' button next to a text input field for 'File to import dose-response data from'. Below it, instructions state: 'Choose a file that CellMissy can use for the import of dose-response data (CSV, TSV, XLS, XLSX)' and 'No data will be saved to the database. Please make sure to create an analysis report before closing CellMissy.' A 'General Information' section asks for a 'general overview for the experiment'. Fields include 'Project/Experiment number' (1), 'Experiment Title' (First test), 'Purpose' (Screenshot for CellMissy manual), 'Dataset' (TestExp011.txt), and a checked checkbox for 'Log-transform doses on fitting for better visual spreading (recommended)'. An 'Experiment Details' section asks for metadata: 'cell line' (HT1080), 'assay type' (proliferation), 'treatment' (PBS), and 'plate format' (12). At the bottom right are 'Cancel' and 'To Analysis>>' buttons.

**2. Input data** On this view, you are shown an overview of your data at the top. Using a menu, you can select which conditions you want to include in the dose-response analysis. These conditions will reappear in a table at the bottom of the screen for easy review. If different conditions have been treated with different compounds (not including control) you will be asked to choose which compound to analyze. When used in conjunction with area analysis, the module will also ignore any data previously excluded by CellMissy's quality control.

### DATA ANALYZER – dose-response input data screen

This table contains all conditions and their respective slopes

Condition number	Treatment	Concentration	Unit	Repl 1	Repl 2	Repl 3	Repl 4
1	Control	0.0	μM	67.575	63.495	54.852	27.787
2	Y27632	0.5	μM	67.775	37.534	40.330	44.838
3	Y27632	1.0	μM	41.328	49.493	excluded	24.772
4	Y27632	1.5	μM	25.655	33.098	23.429	35.932
5	Y27632	2.0	μM	27.935	54.857	10.586	excluded
6	Y27632	2.5	μM	20.956	7.069	32.859	18.965
7	Y27632	3.5	μM	22.566	20.796	23.695	11.644
8	Y27632	5.0	μM	excluded	31.29	19.762	excluded
9	Y27632	7.0	μM	23.84	excluded	42.258	6.585
10	Y27632	9.0	μM	39.822	excluded	45.981	25.386
11	Y27632	12.0	μM	excluded	30.517	7.415	14.447
12	Y27632	15.0	μM	41.58	37.249	22.34	36.863
13	Y27632	20.0	μM	23.26	26.198	18.933	19.948
14	Y27632	25.0	μM	17.148	18.946	19.963	19.066
15	Y27632	30.0	μM	16.263	32.611	22.288	24.121

Select Conditions to analyze  
N: number of Replicates

Conc of Y27632	Unit	Repl 1	Repl 2	Repl 3	Repl 4
0.0	–	67.575	63.495	54.852	27.787
0.5	μM	67.775	37.534	40.330	44.838
1.0	μM	41.328	49.493	excluded	24.772
1.5	μM	25.655	33.098	23.429	35.932
2.0	μM	27.935	54.857	10.586	excluded
2.5	μM	20.956	7.069	32.859	18.965
3.5	μM	22.566	20.796	23.695	11.644
5.0	μM	excluded	31.29	19.762	excluded
7.0	μM	23.84	excluded	42.258	6.585
9.0	μM	39.822	excluded	45.981	25.386
12.0	μM	excluded	30.517	7.415	14.447
15.0	μM	41.58	37.249	22.34	36.863
20.0	μM	23.26	26.198	18.933	19.948
25.0	μM	17.148	18.946	19.963	19.066
30.0	μM	16.263	32.611	22.288	24.121

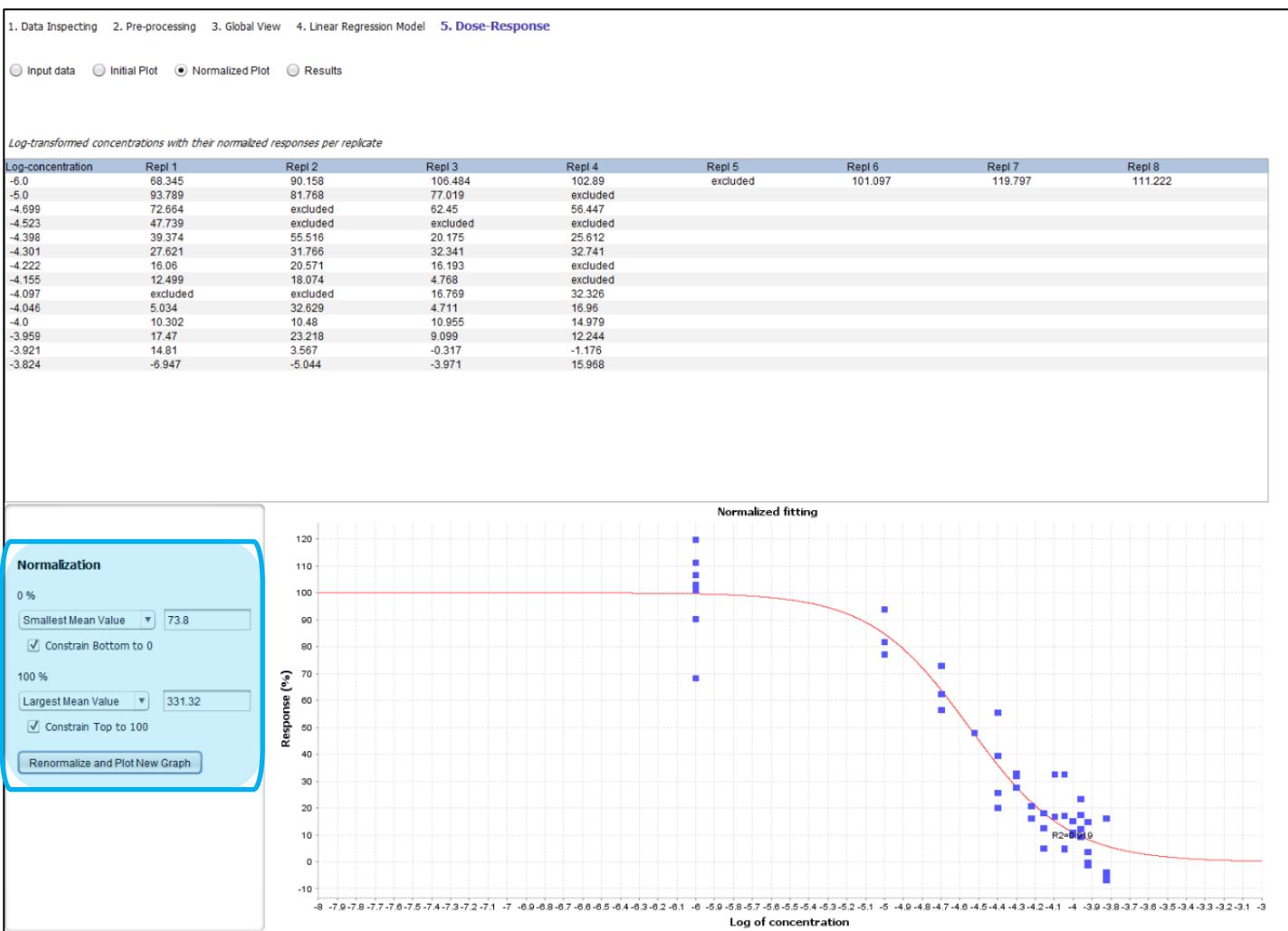
Type of experiment

Stimulation  
 Inhibition

**3. Initial and normalized fit** Cellmissy fits the data to a 4-parameter model that represents the typical the dose-response relationship. These four parameters represent the estimated minimum and maximum response (referred to as 'Bottom' and 'Top' in the software), the slope of the fitted line (Hillslope) and the concentration at 50% effect (EC50%). In both the initial and normalized view, you are show a table of the fitted data and a plot with the best fit superimposed on the data. In both views, you have the option to constrain the Bottom and Top parameter to a certain value, fixing these parameters in the model while fitting. This can be useful if you have knowledge about minimum or maximum response, e.g. from literature or previous experiments.

The initial fit is performed on the raw data (except the doses are log-transformed to provide a better visual spreading for the plots), for the normalized view the data is normalized to be between 0-100. The smallest mean value of responses of the conditions is set to 0, the largest mean value is set to 100. You can customize and redo the normalization using the menu in the bottom-left. Possibilities are mean, median or a value of your own choosing. In the presence of a control condition, its dose is set to one log-value below that of the smallest treatment concentration.

### DATA ANALYZER – dose-response normalized data fit



## 4. Results

In the results table, you are provided with a summary of both data fits. The table contains the best-fit values of the dose-response model parameters, their standard error and 95% confidence interval, as well as the  $R^2$  of the fitted curve. You can save all results in a detailed analysis report, which provides a general overview of the experiment conditions, all plots and statistics, and information on any applied normalization.

### DATA ANALYZER – dose-response results

1. Data Inspecting   2. Pre-processing   3. Global View   4. Linear Regression Model   5. Dose-Response

Input data    Initial Plot    Normalized Plot    Results

Statistical values from the curve fit of the initial and normalized data.

	Initial fitting	Normalized fitting
Best-fit value		
Bottom	63.565	0.0
Top	333.513	100.0
Hillslope	-1.5	-1.657
LogEC50	-4.531	-4.551
EC50	29.45E-06	28.14E-06
$R^2$ (goodness of fit)	0.92	0.919
Standard error		
Bottom	21.908	–
Top	10.826	–
Hillslope	0.332	0.164
LogEC50	0.064	0.032
95% Confidence interval		
Bottom	20.625 to 106.504	–
Top	312.295 to 354.731	–
Hillslope	-2.15 to -0.849	-1.98 to -1.335
LogEC50	-4.656 to -4.406	-4.614 to -4.487
EC50	22.00E-06 to 39.30E-06	24.31E-06 to 32.58E-06

**Initial fitting**

Velocity ( $\mu\text{M}/\text{min}$ )

Log10 of concentration

$R^2=0.92$

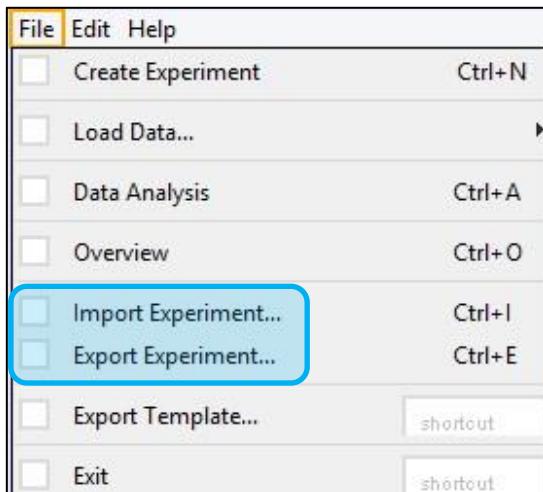
**Normalized fitting**

Response (%)

Log10 of concentration

$R^2=0.919$

## 2.4 DATA EXCHANGE IN CELLMISSY

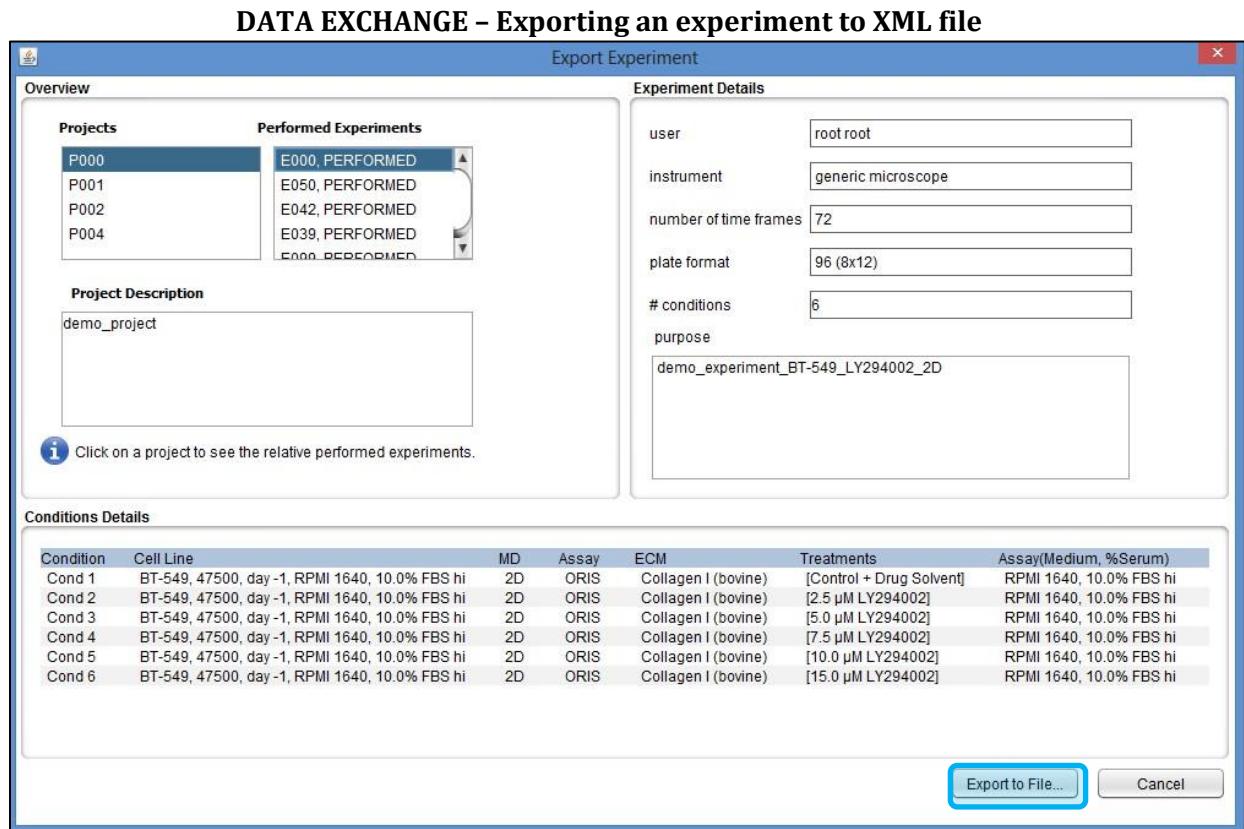


**CellMissy** provides a means to export an entire experiment, complete with the cell migration data, on top of the set-up settings, to an XML file. This XML file can be directly re-imported into **CellMissy**, thus enabling automatically import of an experiment to a different **CellMissy** DB, or to exchange data between different **CellMissy** environments.

### 2.4.1 Export an experiment

From the *File* menu of **CellMissy**, you can export an experiment to an XML file: for a project, you see all the experiments stored in the DB that are *PERFORMED*, i.e. experiments for which the area values have already been loaded. When you click on an experiment, some details are shown, along with a table reporting the biological conditions of this experiment, as shown in the next screenshot. Clicking the “*export to file...*” button will then make you choose a directory to save the file in and **CellMissy** will automatically assign a name to the file, of the type: “*experiment\_ExXX\_PYY*”, with X and Y to be replaced with the number of the experiment and its project, respectively.

At <https://github.com/compomics/cellmissy/tree/master/supportFiles> you can find an example of such XML file.



## 2.4.2 Import an experiment

Still from the *File* menu of **CellMissy**, you can also import an experiment from an XML file; such file can be generated using the “*Export experiment*” functionality of the software illustrated in section 2.4.1. **CellMissy** shows here a dialog, where you can choose an XML file to be imported: once the file has been correctly parsed and imported, the details of the experiment are shown (including the number of algorithms and imaging types), together with a table containing all the conditions details, as shown in the following screenshot (part 1). Clicking the “*next*” button will lead you to another view (in the same dialog, see screenshot part 2), where you can choose a project and an instrument for the experiment to import. Here you also see a detailed description of what is needed in the currently **CellMissy** DB for the importing of the chosen experiment, meaning that if new parameters are associated to the imported experiment, these will be stored to the DB as well. Finally, through the “*save*” button you can proceed to save the experiment.

Please note that you can only import an experiment to a certain project if this experiment (represented by its unique number) is not present in the DB yet for that particular project! This will notably lower the possibility of errors or duplication of data.

## DATA EXCHANGE – Importing an experiment from XML file – part 1

**Import Experiment**

**Choose a File**

File to import experiment from

**i** Choose a file that CellMissy can use for the import of an experiment.  
This has to be an XML file generated through the Export Experiment functionality of CellMissy.

**Experiment Details**

exp number	E039	instrument	generic microscope	# algorithms	2
purpose	BT549 RI 2D, CytD 3D	plate format	150 (10x15)	# imaging types	1
		# conditions	10		
		# time frames	108		
		duration	36.0		
		interval	20.0		

**Conditions Details**

Condition	Cell Line	MD	Assay	ECM	Treatments	Assay(Medium, %Serum)
Cond 1	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	2D	ORIS	Collagen I (bovine)	[Control]	DMEM, 1.0% FBS hi
Cond 2	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	2D	ORIS	Collagen I (bovine)	[5.0 $\mu$ M ROCK]	DMEM, 1.0% FBS hi
Cond 3	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	2D	ORIS	Collagen I (bovine)	[10.0 $\mu$ M ROCK]	DMEM, 1.0% FBS hi
Cond 4	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	2D	ORIS	Collagen I (bovine)	[15.0 $\mu$ M ROCK]	DMEM, 1.0% FBS hi
Cond 5	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	2D	ORIS	Collagen I (bovine)	[20.0 $\mu$ M ROCK]	DMEM, 1.0% FBS hi
Cond 6	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	3D	MCTS	Collagen I (rat tail) (2.0 mg/ml)	[Control]	DMEM, 1.0% FBS hi
Cond 7	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	3D	MCTS	Collagen I (rat tail) (2.0 mg/ml)	[5.0 $\mu$ M CytD]	DMEM, 1.0% FBS hi
Cond 8	BT-549, 50000, day -1, DMEM, 10.0% FBS hi	3D	MCTS	Collagen I (rat tail) (2.0 mg/ml)	[15.0 $\mu$ M CytD]	DMEM, 1.0% FBS hi

## DATA EXCHANGE – Importing an experiment from XML file – part 2

**Import Experiment**

**Choose a Project and an Instrument**

Select an instrument

Select a magnification

Select a project

**i** Choose an instrument and a magnification for the experiment.  
Choose also a project to assign the experiment to.

**Details**

The following new parameters will be stored in the currently used CellMissy DB:

plate format	<i>no new parameters to add</i>
cell line	<i>no new parameters to add</i>
migration assay	<i>no new parameters to add</i>
bottom matrix	<i>no new parameters to add</i>
ecm composition	<i>no new parameters to add</i>
ecm density	<i>no new parameters to add</i>
treatment	<i>no new parameters to add</i>

**Please wait, experiment is being saved to DB!** 

## 2.5 EXTEND CELLMISSY: IMPLEMENT NEW ANALYSIS ALGORITHMS

The architecture of **CellMissy** is fully pluggable for analysis algorithms. This means that different distance metrics, statistical tests, kernel density estimators, outlier detection methods, and multiple hypothesis correction methods can be added by any interested developer. Such new methods can be plugged dynamically into the tool, and will subsequently be immediately available to the user from the **CellMissy** interface.

Since **CellMissy** uses the *Spring* framework for the logic layer, new implementations can be added to the software simply specifying a bean name for each new implementation, along with the fully qualified class name (i.e. including the package) in the Spring configuration XML file. Then, you just have to add the CLASS file(s) for the new implementation(s) to the “ext” subfolder in the **CellMissy** folder, again with the complete package structure. Finally, just run CellMissy: the new implementations will be automatically picked up and shown in the GUI.

### Example - how to add a new multiple testing correction algorithm

This example will show you how to add a new multiple testing correction algorithm. By default, **CellMissy** implements two different ways of correcting for multiple hypothesis, namely the Bonferroni and the Benjamini-Hochberg correction methods; these are presented through a drop down list and you select one or the other while analyzing a certain group of biological conditions. If you want to add an extra algorithm, you need to implement the **CellMissy** *MultipleComparisonsCorrector* interface.

Let us call this new implementation (class): **MyCorrectorImpl**, and let us assume its fully qualified class name is: **com.compomics.cellmissy.cellmissytest.MyCorrectorImpl**.

Thus, in the Spring Configuration file present in the **CellMissy** folder (*mySpringXMLConfig.xml*), we will add a line containing the information about the new implementation, as shown below:

```
<!-- analysis implementations beans -->
<!-- specify a bean name for each implementation, and the fully qualified classname, i.e.: including package -->
<bean id="benjamini" class="be.ugent.maf.cellmissy.analysis.impl.BenjaminiCorrector">
</bean>

<bean id="bonferroni" class="be.ugent.maf.cellmissy.analysis.impl.BonferroniCorrector">
</bean>

<bean id="my_corrector" class="com.compomics.cellmissy.cellmissytest.MyCorrectorImpl">
</bean>
```

Next, we will put our **MyCorrectorImpl.class** file in the “ext” folder, along with the entire package structure, i.e. we will add **com.compomics.cellmissy.cellmissytest**, with **cellmissytest** ultimately containing the new **MyCorrectorImpl.class** file.

**NOTE:** make sure you add both a name for the new bean and the fully qualified class name for the implementation in the Spring file, and be sure as well you build the new implementation with the same JAVA version in which CellMissy has been built (by default, this is JAVA 1.7 or above).

Any other analysis algorithm can be easily extended in **CellMissy** in the same way; below is a list of the extensible feature and the correspondent interface to implement:

Outliers Detection: ***DistanceMetricOperator*** interface

Kernel Density Estimation: ***KernelDensityEstimator*** interface

Distance Metric: ***DistanceMetricOperator*** interface

Statistical Test: ***StatisticsCalculator*** interface

### 3. EXAMPLE DATA

#### A. EXAMPLE AREA DATA TO USE WITH GENERIC INPUT FORMAT

Experiment type: scratch wound assay performed in a multiwell (12-well format) (*Tondeleir et al. PMID: 22448045 and unpublished*): 2D migration on a coating of monomeric collagen Type 1.

Biological Conditions:

Condition	Cell line	Treatment	Number of replicates
1	MEF*	None/control	3
2	MEF <sub>beta-actin-/-</sub>	None/control	3
3	MEF	10µM Y27632**	3

\*MEF: mouse embryonic fibroblast

\*\*Y27632: Rock inhibitor

Data acquisition: 25 measurements during 12.5 hours, 30 minutes time interval Image processing: ImageJ (NIH), manual delineation of area, output in pixel.

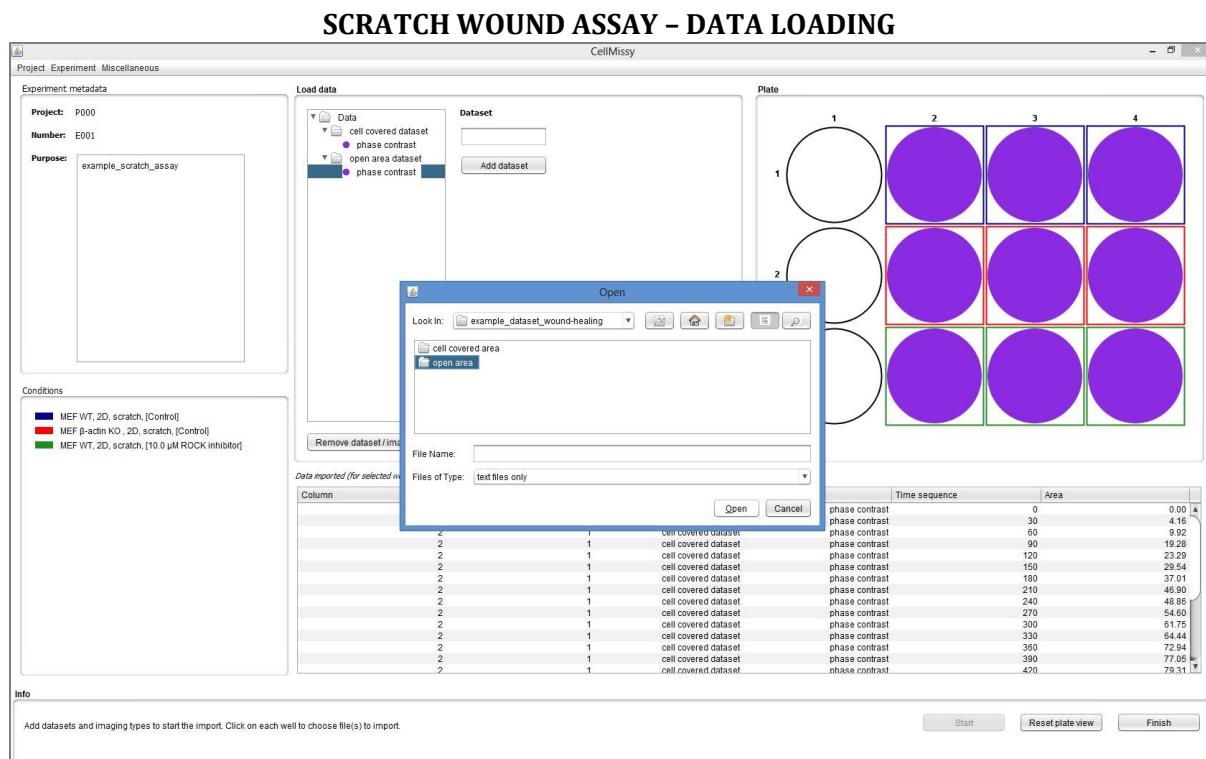
The following screenshot shows how the set-up of this experiment looks like.

**SCRATCH WOUND ASSAY - EXPERIMENTAL SETUP**

The screenshot displays the 'SCRATCH WOUND ASSAY - EXPERIMENTAL SETUP' window from the CellMissy software. The window is divided into several panels:

- Project:** Experiment\_Miscellaneous
- Experiment metadata:**
  - Project: P000
  - Number: E001
  - Purpose: example\_scratch\_assay
- Conditions:**
  - Add condition (blue square)
  - Remove condition (red square)
  - Condition 1 (blue)
  - Condition 2 (red)
  - Condition 3 (green)
- Plate:**
  - Select a plate format: 12 (3x4)
  - Plate layout: A 3x4 grid of wells. Wells 1, 2, and 3 in each row are grouped by color (blue, red, green) and outlined in the same color.
  - Buttons: Randomize wells, Clear last selection, Clear all.
- Conditions Setup:**
  - Choose a Cell Line: MEF β-actin KO
  - Seeding Density: 47500 cells/well
  - Seeding Time: day -1
  - Growth Medium: DMEM
  - Serum: FBS hi
  - Serum Concentration: 15 %
  - Add a new Cell Line: If the cell line you want to use is not present, add it. Cell line name: [input field] Add Cell Line button.
- Info:** Add conditions and select wells for each condition. Conditions details can be chosen in the right panel.
- Buttons at the bottom:** PDF Report, Previous, Next, Finish.

At <https://github.com/compomics/cellmissy/tree/master/supportFiles>, you can download the “example\_dataset\_scratch” folder, where you find two sets of text files you can use for generic import in **CellMissy**. One set of data contains the area values obtained measuring the open area in this experiment, while the other one contains the cell-covered area values in turn produced from the experiment. The next screenshot shows the data importing for this experiment.



The output files generated by **CellMissy** are also present in the same folder, namely in the “output” subfolder. Please note that each selection of a specific time interval produces in **CellMissy** a different analysis session, consequently generating a different PDF report. In the analysis reports folder of the scratch dataset you will find for the cell-covered dataset two different PDF reports, one generated in a session where the full time interval was taken into account, and another one generated in a session where only a subset of the time interval was considered. The open area dataset has instead been analyzed in the full time frame, providing one PDF report.

## B. EXAMPLE AREA DATA AUTOMATICALLY LOADED IN THE SOFTWARE

Experiment type: Cell exclusion zone assay performed in a 96-multiwell plate, 2D migration on coating of monomeric collagen Type I.

Biological Conditions:

Condition	Cell line	Treatment	Number of replicates
1	BT-549*	None, only drug solvent f.c. solvent: 0.2% DMSO	8
2	BT-549	2.5 µM LY249002**	8
3	BT-549	5 µM LY249002r	8
4	BT-549	7.5 µM LY249002	8
5	BT-549	10 µM LY249002	8
6	BT-549	15 µM LY249002	8

\*BT549: breast cancer cell line

\*\*LY249002: PI3K inhibitor, drug solvent DMSO, fc in assay 0.2%

Data acquisition: 72 measurements during 24 hours, 20 minutes time interval

Image processing: Customized automated software developed in our group (CELLMIA, collaboration with DciLabs) (Van Troys et al., *Methods Mol Biol. 2018.*) which measures the cell-covered area (in  $\mu\text{m}^2$ ).

At <https://github.com/compomics/cellmissy/tree/master/supportFiles>, you can download the “example\_dataset\_ORIS” folder, where you find a SQL script (the “example\_dataset\_ORIS” file) to automatically set up and load the data for this experiment in **CellMissy**. Running this SQL script will store the area data to the database, allowing you to visualize, explore and analyze the data with the Data Analyzer module. The output file generated by **CellMissy** is also present in the same folder, namely in the “output” subfolder.

### C. EXAMPLE SINGLE-CELL DATA TO USE WITH GENERIC INPUT FORMAT

Experiment type: Single cell migration of cells embedded in Matrigel (2 mg/ml), cells enriched in one focal plane. Performed in a multiwell (48-well format) (Masuzzo et al., *Sci. Rep. 2017*).

Biological Conditions:

Condition	Cell line	Treatment	Number of replicates // well coordinates
1	Ba/F3, Bcr-Abl p190a*	None/control	2 // (2,2); (2,3)
2	Ba/F3, Bcr-Abl p190a	10µM Y27632**	2 // (2,4); (2,5)
3	Ba/F3, Bcr-Abl p210	None/control	2 // (4,2); (4,3)
4	Ba/F3, Bcr-Abl p210	10µM Y27632**	2 // (4,4); (4,5)
5	Ba/F3, Bcr-Abl p210mut	None/control	2 // (6,2); (6,3)
6	Ba/F3, Bcr-Abl p210mut	10µM Y27632**	2 // (6,4); (6,5)

\* *Bcr-Abl* oncogene variants

\*\*Y27632: Rock inhibitor

Data acquisition: Phase-contrast xyt imaging during 6 hours, 1.5 min interval, 20x objective.

Image processing: Customized automated software developed in our group: CellMIA, output is x-y trajectories.

At <https://github.com/compomics/cellmissy/tree/master/supportFiles>, you can download the “example\_generic\_singlecell” zip folder, which contains separate files (one for each technical replicate), in the typical structure layout required by the software for the generic import of single-cell migration data in **CellMissy**. This data contains the assigned track id, timeframe and x and y coordinates.

## D. EXAMPLE DATA FOR GENERIC DOSE-RESPONSE ANALYSIS

**CellMissy** has the ability to import and analyse generic tables containing doses and responses, even though it is focused on cell migration data. This option can be very useful for proliferation experiments, which are often performed alongside migration assays when testing compounds.

To demonstrate this functionality, we have opted to include some freely available dose-response datasets, from a free R package called ‘drc’ (for a full description of the contents of this package, see: <https://cran.r-project.org/web/packages/drc/drc.pdf>). The table below shows the ‘glymet’ dataset which has nine tested concentrations and 5-14 replicates per concentration. Next to the generic data from the drc package, we have also added three cell migration dose-response datasets that we converted to a simple table. These are three collective migration experiments, two ORIS (2D) and one spheroid (3D) assay. Here, the responses are the average speeds of the cell sheets.

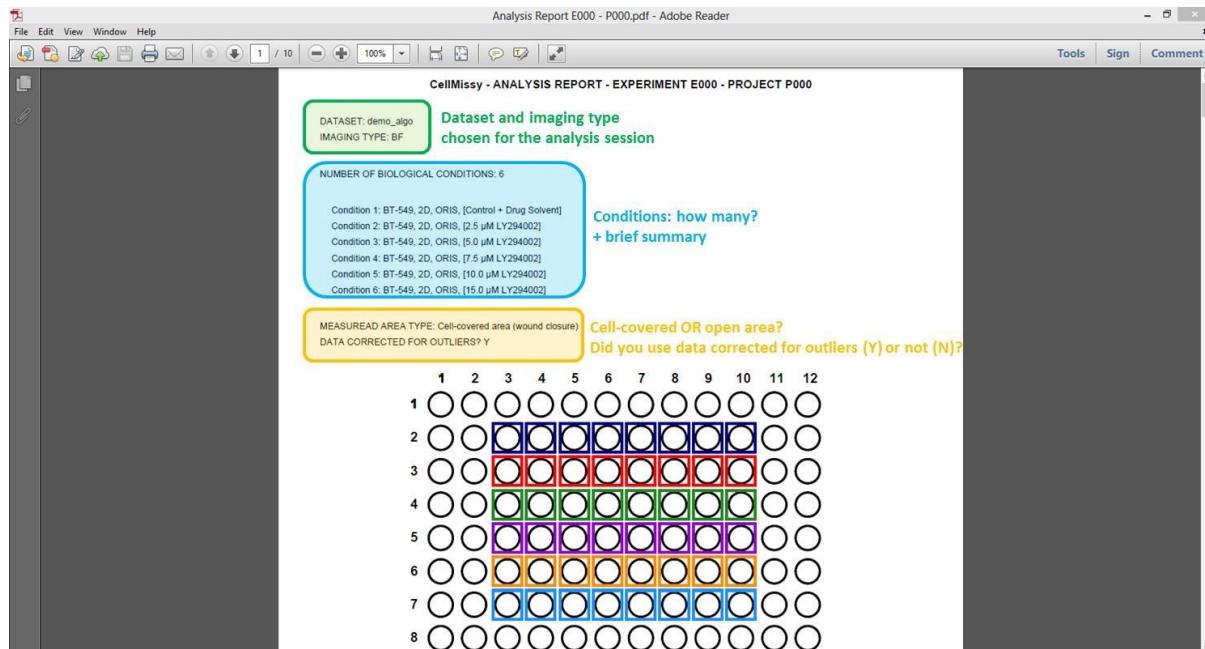
The data can be found at:

<https://github.com/compomics/cellmissy/tree/master/supportFiles/doseResponseData>.

Dose	rgr1	rgr2	rgr3	rgr4	rgr5	rgr6	rgr7	rgr8	rgr9	rgr10	rgr11	rgr12	rgr13	rgr14
0	1.69	1.69	1.44	1.46	1.58									
5937.5	1.56	1.77	1.60	1.55	1.75	1.57	1.76	1.59	1.73	1.49	1.52	1.62		
11875	1.70	1.51	1.75	1.50	1.54	1.37	1.62	1.39	1.70	1.52	1.69	1.35	1.15	
23750	1.61	1.59	1.76	1.44	1.41	1.36	1.36	1.26	1.43	1.12	1.41	1.12	1.19	
47500	1.38	1.33	1.55	1.33	1.35	1.16	1.37	1.47	1.39	1.30	1.40	1.35	1.37	1.08
95000	1.17	0.94	1.32	1.13	1.06	0.90	0.99	0.90	1.06	0.92	1.15	1.09	1.07	0.90
190000	0.47	0.48	0.56	0.59	0.98	0.75	0.63	0.61	0.55	0.49	0.68	0.68	0.64	0.57
380000	0.39	0.29	0.55	0.27	0.45	0.49	0.42	0.51	0.35	0.29	0.51	0.51	0.59	0.57
760000	0.09	0.17	0.05	0.05	0.16	0.16	0.08	0.16	0.10	0.18	0.31	0.28	0.27	0.25

Rgr: relative growth rate. Changes from original: additional data not pertaining to dose or response have been removed, values are rounded up to two decimals and reformatted so that replicates are shown in separate columns.

## E. EXAMPLE OF AN AREA ANALYSIS REPORT (ANALYSIS REPORT E000 - P000\_OPEN AREA, ALSO AVAILABLE ONLINE)



CONDITIONS SUMMARY					
CONDITIONS	# TECHNICAL REPLICATES	TECHNICAL REPLICATES EXCLUDED?	USER SELECTED TIME INTERVAL	MAX. TIME POINT	ANALYZED TIME INTERVAL
	8	N	(0, 71)	71	(0, 40)
	8	N	(0, 71)	71	(0, 40)
	8	N	(0, 71)	71	(0, 40)
	8	N	(0, 71)	71	(0, 40)
	7	Y [(3, 6)]	(0, 40)	71	(0, 40)
	7	Y [(3, 7)]	(0, 47)	71	(0, 40)

Number of replicates kept in the analysis

Technical replicates were excluded (Y) or not (N)?

Time interval selected by the user

Suggested last time point: from this point on, all the conditions have non-null area values

(first, last) time points used for the analysis

