

# MIACA – Minimum Information About a Cellular Assay

## Standardized description of cell-based functional assay projects

Working draft (version 070330)

### Status of this draft

This draft describes the Minimum Information About a Cellular Assay (MIACA) project initiative, the rationale of this initiative, and gives an introduction to the outline of the proposed reporting guideline. This draft is supposed to make public the MIACA initiative and its goals, and to invite for comments and other contributions. Additional contributors who wish to join the initiative are welcome. Any comments and discussions are welcome, and may be sent e.g. to [s.wiemann@dkfz.de](mailto:s.wiemann@dkfz.de) Distribution of this draft is unlimited.

### Copyright Notice

Copyright © (2006-2007) MIACA Standards Initiative. All Rights Reserved. Full Copyright notice at the end of this document.

## Table of Contents

Changes .....	2
Abstract .....	3
Acknowledgements .....	3
Background .....	4
Scientific objectives .....	5
The MIACA standard .....	5
General features .....	7
The MIACA information modules & minimum information .....	8
A MIACA data model .....	14
MIACA is modular and flexible .....	14
Note on the nomenclature .....	15
Invitation to new contributors .....	15
Next steps .....	16
Current Contributors: .....	19
References .....	20
MIACA Full Copyright Notice .....	21

## Changes

There have been three major changes made to the schema compared to the last version of the MIACA draft:

1. Major changes have been made in the description of the individual modules, especially in the data processing and analysis concept.
2. The Ontology for Biomedical Investigations (OBI) ontology work group (<http://obi.sourceforge.net/>) has been added as a major contributor to controlled vocabularies within the MIACA project.

## Abstract

Recent advances and developments in genomics and functional genomics have enabled large scale analysis of gene and protein function by means of high throughput cell biology analysis. Cells are often perturbed *in vitro* and induced effects are recorded and analyzed. Perturbations can be triggered in several ways, for instance with molecules (siRNA, expression construct, small chemical compound,...) or with other stresses cells are subjected to (temperature shift, serum starvation, ...). The cellular responses to such perturbations are analyzed in order to identify players in the biological processes addressed and to understand biological principles. Often such analyses are carried out in high-throughput, and a huge amount of data is collected. Given the availability of such data sources there is a growing need to compare and integrate such data that originate from different, however often complementary approaches, and to elucidate higher order principles. These processes are the basis of Systems Biology. Minimum information on the rationale, materials, the conditions prior to, during, and after the perturbation, as well as all experimental processes need to be recorded and documented in order to fully describe the set-up and progress of a cellular assay project, and to be able to understand and follow the data analysis and knowledge generation processes. Only then an efficient integration of data is possible. Standardized nomenclatures/ontologies and data models should be employed wherever possible. In conjunction, this allows researchers to independently evaluate the results and conclusions obtained from such assays, and is prerequisite for the establishment of repositories (databases) that take and disseminate data from cellular assays.

MIACA is neither intended to become such a data repository nor does it intend to prescribe the experimental layout of cellular assays. Instead, “merely” the minimum reporting requirements that are supposed to annotate the data from such assays shall be defined. This draft shall make public the MIACA Standards Initiative, and attract experts in the fields to contribute and actively participate in the MIACA standard development process.

## Acknowledgements

The process borrows heavily from the Proteomics Standards Initiative (PSI)<sup>1</sup>, and the Global Grid Forum document process which in turn is based on the Internet Engineering Task Force Request for Comments document process. This project is funded in part by the German Federal Ministry of Education and Research within the NGFN program (grant 01GR0420).

## Background

Large-scale efforts are under way generating genome-wide resources of siRNAs, shRNAs, full-length cDNAs, cloned ORFs, and other derivatives that cover a growing number of organisms (e.g. *H.sapiens*, *M.musculus*, *D.melanogaster*, *C.elegans*). These resources are utilized in a rapidly increasing number of laboratories for the systematic functional analysis of proteins and RNAs<sup>2,3</sup>. High-throughput experiments and assays are established and systematically applied to screen for effects induced by systematic perturbations of cells *via* the underexpression or overexpression of the respective proteins/RNAs in living cells. Protein over- and under-expression are, however, by far not the only means to perturb cells. Small chemical compounds, physicochemical parameters (temperature, humidity, supplements to the medium), and other treatments are utilized and their effects on cellular systems are recorded.

While the exploitation of siRNA, cDNA, and ORF resources in high-throughput methodologies is still mostly under development, an increasing need for standards has become obvious by which the processes leading to the collection of data can be accurately described, and by which this information can be stored, interpreted, and exchanged between bench researchers and data repositories<sup>4,5</sup>. If a standard for the description of a cellular project is to be developed, however, it should in the ideal case be comprehensive and cover any perturbation and the detection of induced effects that are possible.

In the fields of gene expression profiling<sup>6</sup> and proteomics<sup>7</sup> the definition of standardized experimental descriptions has greatly facilitated the interchange of data between laboratories<sup>8</sup>. Use of standardized ontologies<sup>7,9</sup> was key for the successful integration of such data in common repositories<sup>10</sup>. In these repositories the data can be accessed, downloaded and utilized by scientists with common interests in the same field of work (i.e. the biological question a respective assay addresses), but as well by scientists interested in other more specialized fields who want to obtain more information about their proteins/genes of interest. While databases that would disseminate data from cellular assays are currently not established, the growing interest especially in the field of Systems Biology is likely to make this a pressing issue in the near future. Reporting guidelines will then be mandatory to allow for a systematic and controlled data entry and processing as prerequisite for subsequent querying and meta analysis. In such databases the information should be provided in sufficient detail to allow for the interpretation of the validity of the project and its outcome, to allow for comparisons with other projects, and in principle permit replication.

## Scientific objectives

A number of closely related working parties covering both large scale cell-based assays and bioinformatics for data analysis are currently working on the definition of the minimal requirements of a cell based assay project. We aim at developing standards to describe the conditions of cell based projects in sufficient detail to allow researchers from other laboratories to evaluate the principles of the applied assay, the usefulness and applicability of the collected data, and the value of the inferred conclusions. Assays and their data that have been described with an accepted and common standard will be the basis for the establishment of data repositories that take and disseminate such information (e.g. similar to ArrayExpress for RNA expression profiling).

The MIACA guideline consists of three parts: 1.) The minimum requirements to describe a cell based assay project, 2.) a consensus object model integrating this information with an accompanying XML interchange format that will be compatible with the already existing formats developed by the HUPO Proteomics Standards<sup>11</sup> and by the Functional Genomics Experiment (FuGE) initiatives<sup>12</sup>. 3.) The standard will benefit from and contribute to appropriate controlled vocabularies (e.g. OBI, GO<sup>9</sup>, FuGO), which will enable the user to describe related information in a consistent way. Assays and their data that have been described with an accepted and common standard will be the basis for the establishment of data repositories that take and disseminate such information (e.g. similar to ArrayExpress for RNA expression profiling). Here we describe the three parts of the MIACA guideline and a sample implementation utilizing a whole genome RNAi screen for proteins modulating cell adhesion.

In the pre-planning phase of this new standard contributors are invited to identify themselves and express their interest in participation. In the research phase the individual modules are identified and worked out with sufficient specificity, however avoiding too much and unnecessary detail. Standard nomenclatures/ontologies will be utilized when available, novel nomenclatures will be implemented only when necessary.

## The MIACA standard

As a guide to minimum reporting requirements, the participating parties are producing a draft document entitled “[Minimum Information About a Cellular Assay](#)” (MIACA)<sup>5,13</sup>. This document will describe and encompass all necessary aspects of a cell-based functional genomics project. This document states the principles underlying the specification of the data and metadata that should be captured from cell-based functional genomics projects. It also describes a series of modules containing the minimum reporting requirements for a specific part of a project.

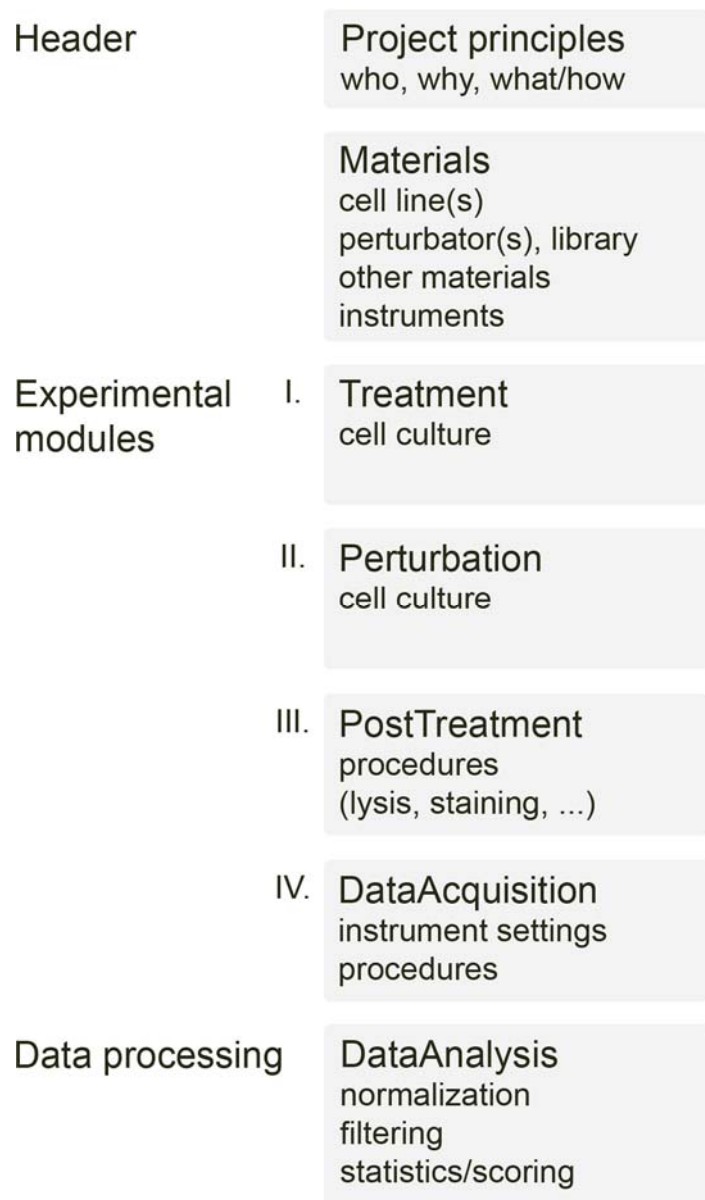


Figure 1. MIACA is structured in three major blocks: 1. the Header describing the rationale and the source of data, 2. the Experimental modules all wet-lab experimentation, and 3. the Data processing and analysis. Especially the experimental part is modular to allow for a maximum of flexibility in the planning and execution of assays. Any of these modules could be skipped to tailor the MIACA implementation to the special needs. A detailed listing and description of the individual components that are described in the individual modules is given in the supplementary text.

It is anticipated that these modules will evolve over time to take account of changes in technologies and applications. The MIACA standard currently consists of a general part, outlining the amount of information sufficient to allow a reasonable description and an assessment of data and interpretation quality within a cellular assay project, supported by a number of technology-specific parts. These parts are prepared by expert committees and then user-validated before they are fully adopted. Care has been taken to avoid too much specificity in terms of the types of assays and read-out systems to make this standard

comprehensive for the description of any cell-based project. Also the data analysis is currently not covered here, as this must be directly tailored to the specifications of assay projects, read-out system, the applied statistics tools, and to the biological question that shall be answered. Standardized descriptions of data acquisition and data analysis may be developed by dedicated working groups at a later stage, or linked in from complementary projects.

## **General features**

1. Information should be of sufficient detail to allow for the interpretation of the validity of the project and its outcome, to allow for comparisons with similar projects, and to permit replication.
2. Information should be structured in a way that enables useful querying and automated data analysis.
3. Whenever possible, controlled vocabularies (ontologies) should be used. Recommendations are given for which parts a controlled vocabulary is required. Structures and ontologies should be used as much as possible from existing standards (OBI (<http://obi.sourceforge.net/>), OBO (<http://obo.sourceforge.net/>), MGED, HUPO-PSI), more universal features should be taken from general standards (e.g. SI-units). Close collaboration with the FuGE initiative (<http://fuge.sourceforge.net/>) is anticipated.
4. Additional ontologies need to be developed/adopted. In the meantime, information could be included as name/value/source triplets (as described e.g. for MIAME).

## The MIACA information modules & minimum information

In the following, a listing of MIACA modules and the proposed minimum information requirements (i.e. bullet points) for every module is given.

### Header / ProjectDescription

#### Source

Contact details of researcher/person (1) in charge of the project:

- SourceName
- SourceAffiliation/institution
- SourceDepartment
- SourceAddress
- SourceCity
- SourceCountry
- SourceEmail

#### Project

Description (text) of the project (1) within a larger context. Could be e.g.

- ProjectProcess: biological process that is addressed
- ProjectEffect: description of measured effect and the 'reporter' used for the measurement (e.g. antibody)
- ProjectControls (1..n): Controls are relevant in data analysis to determine the robustness and the dynamic range of the assay, and to produce the list of significant hits.

e.g. measuring the effect of 'Perturbation' on the induction of apoptosis (biological process), by measuring the level of activated caspase 3 (measured effect), with a antibody specific for the activated enzyme (reporter). One project consists of experiments ( $\geq 1$ ) which are carried out in arrays (1 to n positions). A proposed definition of "experiment" is given in "The experiment concept" above.

Note: if more than one biological process are to be addressed and/or more than one (independent) effects shall be measured in subsequent measurements, these processes and effects should be described as separate, though dependent assays. However, 'effect' does not equal 'parameter', and thus several different parameters may be recorded in one data acquisition process within one assay.

#### Application

Description (text) of the specific application (1) of this project. Short description (standardized terminology) of:

- ApplicationAbstract
- ApplicationReference to (a) publication
- ApplicationProtocol reference to an accessible protocol/SOP describing the specific assay in greater detail

## Header / Materials

Description of all materials that are utilized in the project.

### ArrayType (1..n)

reaction and culture vessels that are used during the project

- ArrayName
  - ArrayID
  - ArrayType (e.g. 96, 24 well, single flask, or glass slide for cell arrays)
  - ArrayVendor or manufacturer
  - ArrayOrderNumber
    - ArraySize (surface area) could be extracted as well from the manufacturer's web-site, however is helpful e.g. to calculate the concentration of cells/mm<sup>2</sup>

### CellLine (1..n)

Description with:

- Name of cell line
  - CellLineID
  - ATCC\_number or <other>
    - CellLinePassageNumber
    - Mycoplasma test (Y/N) and other validation
  - details required in case of <other>, e.g for primary cells or cell lines: Cell line requires validation prior to its utilization in cellular assay (proven relevance<sup>1</sup>).
    - CellLineSpecies
    - CellLineStrain
    - CellLineTissue
    - CellLineOrgan
    - CellLineCellType
    - CellLinePassageNumber
    - CellLineContact (when from different lab) equivalent to <source>
    - CellLineReference to publication
- modifications (optional, if for instance transgene/genetic manipulations had been made, e.g. stably transfected, induced resistance,...) – reference to database record (“modification of cell line”), and/or protocol, and/or publication where modifications are described. Cell line requires validation prior to its utilization in cellular assay (proven relevance).

Note: if one and the same “assay protocol” is supposed to be applied to two or more cell lines, these assays are still handled independently in the data model.

### Reagent (1..n)

Reagents are media, media supplements, kits, buffers and solutions, water,...

- ReagentName
  - ReagentID
  - ReagentVendor or manufacturer
  - ReagentOrderNumber
  - ReagentLotNumber (maybe optional for database entry, but should be recorded for own sake anyway – good laboratory practice should be to use the same one lot throughout one project)
  - ReagentStockConcentration

## Perturbator (0..n)

Description of materials/conditions that are used in the project to perturb the cells. Includes information on type (e. g. cDNA, siRNA, small chemical compound), name, external references (e.g. gene/protein identifiers/order numbers). The number of 'Perturbators' is commonly  $\geq 1$ , however, 'Perturbation' is not mandatory.

if is material

- PerturbatorName
  - PerturbatorID
  - PerturbatorType
  - PerturbatorSequence/Composition of siRNA/shRNA; structure/'smile' if compound
  - PerturbatorDatabaseID – if available (e.g. in PubChem or ChEMBL)
  - PerturbatorVendor or manufacturer
  - PerturbatorOrderNumber
  - PerturbatorLotNumber (maybe optional/dispensable for database entry, but should be recorded for own sake anyway - good laboratory practice should be to use the same one lot throughout one project)
  - PerturbatorTargetSpecies (e.g. if siRNA)
  - PerturbatorTargetID (e.g. geneID of target gene in case of siRNA)
  - PerturbatorStockConcentration
  - PerturbatorSolvent

if Perturbator part of / is 'library':

- LibraryID as a reference to <library>
- Type of library: (gain-of-function/overexpression (cDNA, ORF, other); loss of function (siRNA, shRNA, miRNA, viral, other); compound; other reference on library (if available))
- LibraryFormat (single/pooled - e.g. siRNA pool)
- LibraryVendor or manufacturer (incl. URL for description)
- LibraryPreparation methodology
- LibraryFeatures: number of unique features (e.g. genes, compounds) represented in library, reference to where a list of these features can be found
- LibraryFeature PlateNumber and well position; mapping of individual perturbators in library format.
- LibraryFeatures of perturbators (length of siRNA/shRNA, type of hairpin, other)
  - if perturbator is virus:
    - VirusType
    - VirusTiter
  - if perturbator is compound:
    - Solvent (will this always be same in whole library?)
    - % of compounds with >80% purity
    - % of compound comply with Lipinski rule-of-5
    - Average molecular weight
- transfection/transduction efficiency of <cell type> (average, cv) with 'Perturbator' (library)
- if perturbator is other stress/condition (details can be also specified in cell culture conditions of perturbation if e.g. temperature):
- ConditionName
  - Type
  - Specifications (may be sequence, composition, ...)

## CompositeReagents (0..n)

Individual reagents are often combined to make a complete cell culture medium, a staining mix, or e.g. make up the final 'Perturbator' [DNA prep, complex formation, production of chip-array in case of cell array]

- CompositeReagentID
- CompositeReagentList (references) of single Reagents used with:
  - ReagentVolume
  - ReagentConcentration
- CompositeReagentList (references) of other Material (e.g. ArrayType)
- CompositeReagentTreatment: References to the respective treatments used to create the composite reagent for subsequent application in  $\geq$  one of the experimental modules with:
  - CompositeReagentTemperature
  - CompositeReagentDuration
  - CompositeReagentReference to protocol

Note: Reagent is part of **Reagents** from above

## Instrument (0,1)

Description of the data acquisition station and any other instruments utilized in the project (e.g. for transfection): name, type, manufacturer. The diversity of data acquisition stations is huge. Detailed descriptions of such instruments and their settings could be part of specialized documents, or link to other standardization efforts (e.g. MIFACE for fluorescence activated cell sorting [FACS] <http://flowcyt.sourceforge.net/miface/>).

- InstrumentName
  - InstrumentID
  - InstrumentVendor or manufacturer
  - InstrumentModel
  - InstrumentSpec (e.g. objective lenses, numerical aperture, filters)

## Experimental modules

### Treatment

Description of the conditions (1..n) that are applied to CellLine during culturing of cells. Treatment precedes a (optional) 'Perturbation', and may be applied also downstream of a 'Perturbation'.

- TRConditionName
  - TRID
  - TRTimeStampStart
  - TRTimeStampStop
  - TRArrayType
  - TRReagentName (0..n)
  - TRReagentVolume
  - TRCompReagentName (1..n)
  - TRCompReagentVolume
  - TRCellLine
  - TRPassageNumber
  - TRSeedingDensity
  - TRCellDensity at TRTimeStampStop (if applicable)

- TRTemperature
- TRCO2Content
- TRHumidity
- TROther

Treatment may also be a trypsinization step, a split of cell culture...

### **Perturbation**

Description of the 'Perturbation' **(0..n)**. Perturbation is a special case of 'Treatment', and allows for the administration of perturbator(s). More than one perturbator can be applied in the same Perturbation. More than one Perturbation may be consecutively applied. These are then treated as individual Perturbations that can be further separated by 'Treatments'. Perturbation can be with transfection (siRNA, expression clone), treatment with small compound, temperature shift, etc.

- PECondition\_name
  - PEID
  - PESTimeStampStart
  - PESTimeStampStop
  - PEArrayType
  - PEReagentName **(0..n)**
  - PEReagentVolume
  - PECompReagentName **(0..n)**
  - PECompReagentVolume
  - PEPassageNumber
  - PESeedingDensity
  - PECeildensity at PESTimeStampStop (if applicable)
  - PEPerturbatorName **(0..n)**
  - PEPerturbatorVol
  - PETemperature
  - PECO2Content
  - PEHumidity
  - PEother

Notes:

concentration of perturbator(s) may be calculated from the information given in reagents and compositeReagents, and the volume given here.

### **PostTreatment**

Description of the conditions **(0..n)** that are applied after the perturbation and optional subsequent 'Treatment' (continued cell culture), and prior to data acquisition, i.e. lysis, fixation, staining, antibody incubation,....

- POConditionName
  - POID
  - POTimeStampStart
  - POTimeStampStop
  - POArrayType
  - POReagentName **(0..n)**
  - POReagentVolume
  - POCCompReagentName **(1..n)**
  - POCCompReagentVolume
  - POTemperature

- POOther

### **DataAcquisition**

Detection of the effect(s) induced by the perturbation **(0..n)**:

- DAID
- DATimeStampStart
- DATimeStampStop
- DAInstrument: reference to instrument above
- DAInstrumentSettings (e.g. excitation and emission wavelengths with filter sets, lamp energy/Laser settings, time for data acquisition)
- DAReference to a protocol/publication

The conditions of data acquisition could be described in a supplemental document (=reference to a protocol), or be described using a dedicated standard (e.g. MIFACE for FACS; <http://flowcyt.sourceforge.net/miface/> or MIAME<sup>2</sup> for gene expression profiling). Data acquisition is optional (value=**0**) when effect is not immediately measured but cells/lysates/... are further utilized as material in other conditions. Data acquisition may be **>1** in cases where for instance samples are taken and measured at different time points.

### **DataProcessing**

Description of the processes applied to analyze the raw-data in order to generate a hit list **(0..n)**. Experimental data from cellular assays will commonly be analyzed, in order to generate a “hit-list” and to define “candidates” that are scored positive in the biological question the particular assay addressed. Data analysis is usually carried out sequentially, applying a number of algorithms and analysis tools. A large variety of analysis methods already exist, more methods will be developed and become available with time. With increasing sensitivity and specificity of analysis tools, a re-analysis of (raw)-data will potentially help to improve the candidate lists.

### **RawData (0..n)**

Description of raw data.

- RDType (number, image, movie)
- RDFormat (ASCII, TIFF, WMA)
- RDLink (link to raw data)

### **ProcessedData (0..n)**

The raw data can be processed either manually or through automated computational methods. Processing commonly generates numbers and/or phenotypic descriptions. This process is commonly split in two consecutive actions:

1. Procedures applied for the filtering and normalization of data, describing also which controls this is based on.
2. The scoring schema and statistical analysis procedures for the production of a “hit-list”.

Both actions are described.

- PDProcessName
  - PDSoftware

- ODReference (reference to a publication)
- PDVersion
- PDAlgorithm
- PDParameters
- PDLink (link to processed data)

### **PhenotypicDescription (0..n)**

A phenotypic description is usually obtained by manual observation of images and/or interpretation of processed data.

- PHProcessName
  - PHProtocol (Protocol for the characterization and classification of phenotypic observation)
  - PHDescription (phenotypic description)
  - PHTimeSpace (Time and space parameters)
  - PHTimeStamp
  - PHLink (link to phenotypic description)

Finally, a link to the raw, the processed and the interpreted data is required.

### **A MIACA data model**

is currently under development. Assuming a proper database structure to be available, the completion of all points of the minimum information above should be a simple clicking from selection lists that are generated from tables (ArrayTypes, CellLines, Reagents, CompositeReagents, etc.). Selection of perturbators should be linked to mapping tables and bar coded plates, to avoid errors and unnecessary typing. A MIACA-compliant data model and database structure that have been implemented on a sample application are under development. This shall be part of a MIACA publication that has been submitted in late 2006.

### **MIACA is modular and flexible**

The modules of MIACA, the identity and content of information modules is subject to:

- discussion in the community – additions and changes can be made at any time in this phase of the project. The community is explicitly invited to communicate comments and concerns.
- the individual project that shall be described – depending on the project outline, some modules may not be required to comprehensively describe a given project. For example, a perturbation is not compulsory to take place in order to utilize the MIACA standard for the description of a cellular assay. Instead, cells can just be grown and harvested after time. Nevertheless, also in such case a standardized description of cell line and growth conditions, as well as of the post-treatment is useful.

## Note on the nomenclature

Some key terms are often used differently by scientists and require standardization:

A **project** describes 1.) the conditions that have been established to measure the effects which are induced in cells in response to a perturbation. 2) is reference to data that have been acquired in these measurements in order to address the biological question this project was designed for.

The term **assay** is commonly utilized in two distinct ways: 1. it describes the data acquisition process, where the effects induced by a perturbation are monitored. 2. it describes the experimental processes carried out to address a biological question. The biomedical community appears to favor the latter definition.

The term **experiment** is commonly used with different meanings in the literature (one part of a project – e.g. the transfection procedure; treatment of one array; several arrays (replicates); a whole project, ...). We propose to use this term to describe the fate of one feature of an array, e.g. one well in a microtiter plate, or one spot of a chip. Alternatively, this term should not be used at all to avoid confusion.

## Invitation to new contributors

Thus far, MIACA has been developed by a small community of scientists working in wet-lab experimentation and in bioinformatics analysis of cellular assays. However, the fields covered by these persons are rather small in view of the plethora of biological questions and applications of cellular assays possible.

Understanding the need for a standardized description of cellular assays is a first prerequisite for new contributors to enter the community of developers, to contribute and review, and to establish this initiative as a driving force.

Scientists, wet-lab, bioinformatics,... working in a field related to cellular assays are welcome to identify themselves and to actively contribute in order to make another valuable contribution to the scientific community, instead of awaiting some top-down definition of such guidelines that would likely not achieve general acceptance<sup>14</sup>. Mailing lists have been installed at <https://sourceforge.net/projects/miaca> where individuals interested in being updated are invited to subscribe. Working groups shall be established where individual modules, ontologies and nomenclatures will be discussed and agreed upon. Dedicated units shall be established to discuss and agree upon standardized descriptions for specific technologies (e.g. perturbation with siRNAs or small compounds; data acquisition with plate readers, FACS, or microscopy).

## Next steps

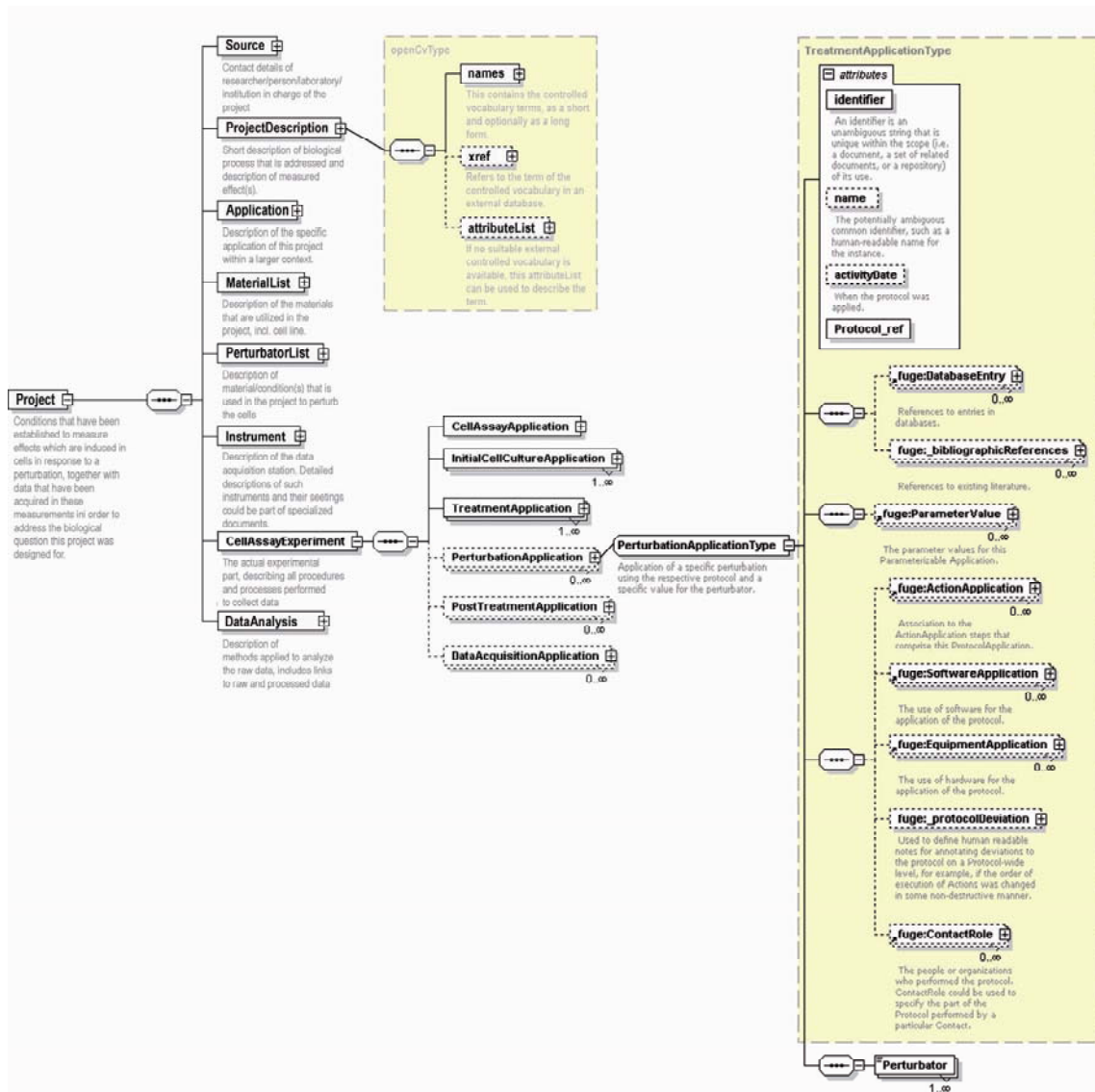
Future plans for the group include a common query interface which will access data held by all participating databases – the group is currently in the process of collecting user requirements for such an interface and would welcome any input. Sample implementations of databases that take data and information from cellular assays and that are compliant with MIACA standards are under development.

06/2007 Such databases will be made public through the miaca sourceforge pages.

The MIACA initiative aims at stimulating accepted database providers (e.g. nucleotide sequences, gene expression profiling) to implement also databases that shall take and disseminate data from cellular assays, and the accompanying annotation. MIACA is supposed to be a stimulator for these databases and to aid in the design and set-up of such databases, that the data is made available to the community in a form that renders this data useful. MIACA is not intended to become a data repository as such.

(incomplete) list of features that could/should be expressed through controlled vocabularies

Qualifier	example values	Ontology	Comment
1 Source	name	MGED, PSI	
	affiliation		
	address		
	contact details		
2 projectDescription:biological effect	Cell survival	GO ?	different layers of description possible, either a global effect is measured, or a specific one
	Cell adhesion		Phenomenon
	Chemo-resistance		
projectDescription:biological process	MAPK signaling	GO	pathway
	Apoptosis		
	Cell proliferation		
projectDescription:measured effect	turnover of marker (e.g. WST or MTT)	GO ?	reference to the respective specific cellular process that is measured
	Cell number in supernatant		
	turnover of marker (e.g. WST or MTT)		
	Erk1/2 activation	GO	
	caspase-3 activation		
	DNA replication - BrdU incorporation		
3 Materials			
cellLine:name	NIH3T3	obo, Brenda ?	
	HeLa		
	HEK293		
cellLine:organism	mouse	NCBI	
	human		
cellLine:tissue	embryo	GO, eVOC ?	
	fibroblast		level needed or just link to ATCC ?
	adult		level needed or just link to ATCC ?
	cervix		
	epithelial		level needed or just link to ATCC ?
	adenocarcinoma		level needed or just link to ATCC ?
	kidney		
	epithelial		level needed or just link to ATCC ?
	transformed with adenovirus 5 DNA		level needed or just link to ATCC ?
cellCharacteristics	adherent growth	??	which others ?
cellCultureArray:type	96well	??	
	24well		
	eppendorf tube		
cellCultureMedium	DMEM	??	
cellCultureMedium:supplement	Calf Bovine Serum, streptavidine	??	for small molecules could be: obo/ontology/chemical/chebi.obo
perturbatorType	siRNA	??	
	expression construct	??	
	small molecule	??	
	time	??	
	temperature shift	??	
4 pretreatment/perturbation/posttreatment			
treatment of cells during project	seeding and cell growth	??	
	perturbation		perturbator reference to materials
	trypsination	?	
	wash steps	??	
	antibody staining		antibody reference to materials
physicochemical values	temperature, volume, concentration	PSI	
commonly used buffers & solutions	PBS		
5 data acquisition	FACS	??	new ontology needed?
	microscope		
	plate reader		
	Mass spectrometer		
6 data analysis	commercial tools		possible to standardize??
	web-accessible tools		
	home made tools		



Subset of an example MIACA XML-file.

## Current Contributors:

Stefan Wiemann<sup>1</sup>, Alexander Mehrle<sup>1</sup>, Florian Hahne<sup>1</sup>, Henning Hermjakob<sup>2</sup>, Rolf Apweiler<sup>2</sup>, Dorit Arlt<sup>1</sup>, Stephanie Bechtel<sup>1</sup>, Wolfgang Bielke<sup>3</sup>, Amanda Birmingham<sup>4</sup>, Queta Boese<sup>4</sup>, Peer Bork<sup>5,6</sup>, Frank Buchholz<sup>7</sup>, Xun Cai<sup>8</sup>, Evelyn Camon<sup>2</sup>, Piero Carninci<sup>9</sup>, Sumit K. Chanda<sup>10</sup>, Zhu Chen<sup>8</sup>, Michael Cusick<sup>11</sup>, Christophe Echeverri<sup>12</sup>, Jürgen Eils<sup>13</sup>, Roland Eils<sup>13</sup>, Ronald Frank<sup>14</sup>, Andrew Fraser<sup>15</sup>, Andreas Girod<sup>5</sup>, Bettina Haedrich<sup>3</sup>, Michael Hannus<sup>12</sup>, Midori Harris<sup>2</sup>, Yoshihide Hayashizaki<sup>9</sup>, Winston Hide<sup>16</sup>, Oliver Hofmann<sup>16</sup>, Peter Huber<sup>17</sup>, Antony Hyman<sup>7</sup>, Lars Jensen<sup>5</sup>, Olli Kallioniemi<sup>18</sup>, Anastasia Khvorova<sup>4</sup>, Liisa Koski<sup>12</sup>, Brian Kraybill<sup>19</sup>, Joshua LaBaer<sup>20</sup>, Ulf Landegren<sup>21</sup>, Chris Lawerenz<sup>13</sup>, Hans Lehrach<sup>22</sup>, Fei-Fei Liu<sup>23</sup>, Meher Majety<sup>1</sup>, Bernard Mathey-Prevot<sup>24</sup>, Gordon B. Mills<sup>25</sup>, Nobuo Nomura<sup>26</sup>, the OBI working group<sup>27</sup>, Norbert Perrimon<sup>24,28</sup>, Angel Pizarro<sup>29</sup>, Annemarie Poustka<sup>1</sup>, John Quackenbush<sup>30</sup>, Heiko Rosenfelder<sup>1</sup>, Oezguer Sahin<sup>1</sup>, Kourosh Salehi-Ashtiani<sup>11</sup>, Florenz Sasse<sup>14</sup>, Mamatha Sauermann<sup>1</sup>, Aaron D. Schimmer<sup>23</sup>, Christian Schmidt<sup>1</sup>, Laura Selfors<sup>19</sup>, Caroline Shamu<sup>19</sup>, Birte Sönnichsen<sup>12</sup>, Chris Taylor<sup>2</sup>, Gary Temple<sup>31</sup>, Marc Vidal<sup>11,32</sup>, David J. Wrobel<sup>19</sup>

---

<sup>1</sup> Division of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany

<sup>2</sup> European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

<sup>3</sup> Qiagen GmbH, Qiagen Strasse 1, 40724 Hilden, Germany

<sup>4</sup> Dharmacon Inc, Thermo Fisher Scientific, 2650 Crescent Drive, Lafayette, CO80029, USA

<sup>5</sup> European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

<sup>6</sup> Max-Delbrück-Centre for Molecular Medicine, D-13092 Berlin, Germany

<sup>7</sup> Max-Planck-Institut of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

<sup>8</sup> Shanghai Institute of Hematology and Shanghai Center for Systems Biomedicine, Rui-Jin Hospital, 197 Rui-Jin Road, Shanghai 200025, China

<sup>9</sup> RIKEN, Genome Science Laboratory, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

<sup>10</sup> Division of Genomics, Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive La Jolla, CA 92121, USA

<sup>11</sup> Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, 44 Binney Street, Smith 858, Boston, MA 02115, USA

<sup>12</sup> Cenix BioScience GmbH, Tatzberg 47, 01307 Dresden, Germany

<sup>13</sup> Division of Theoretical Bioinformatics, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany

<sup>14</sup> Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany

<sup>15</sup> The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

<sup>16</sup> South African National Bioinformatics Institute, University of Western Cape, Private Bag X17, Bellville 7535, South Africa

<sup>17</sup> Department of Radiation Oncology, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

<sup>18</sup> Turku Centre for Biotechnology, P.O.Box 123, BioCity, Turku, FIN-20521, Finland

<sup>19</sup> ICCB-Longwood Screening Facility, Harvard Medical School, 250 Longwood Ave., Seeley G. Mudd Building, Room 604, Boston MA 02115-5731, USA

<sup>20</sup> Harvard Institute of Proteomics, 320 Charles St., Cambridge, MA 02141-2023, USA

<sup>21</sup> Dept. of Genetics and Pathology/Molecular Medicine, Rudbeck Laboratory, Dag Hammarskjöldsvag 20, Se-75185 Uppsala, Sweden

<sup>22</sup> Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany

<sup>23</sup> Princess Margaret Hospital, Ontario Cancer Institute, University of Toronto, Canada

<sup>24</sup> Drosophila RNAi Screening Center, Dept. of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston MA02115, USA

<sup>25</sup> M. D. Anderson Cancer Center, 7435 Fannin Street, Unit 950, Houston, TX 77054, USA

<sup>26</sup> Biological Information Research Center, AIST, Aomi 2-41-6, Koto-Ku Tokyo, 135-0064, Japan

<sup>27</sup> OBI is the Ontology for Biomedical Investigations project, <http://obi.sourceforge.net>

<sup>28</sup> Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

<sup>29</sup> Institute for Translational Medicine and Therapeutics, University of Pennsylvania, 806 BRB II/III, 421 Curie Blvd., Philadelphia, PA 19104-6160, USA

<sup>30</sup> Department of Biostatistics, Dana-Farber Cancer Institute, Mayer 232, 44 Binney Street, Boston, MA 02115, USA

<sup>31</sup> Mammalian Gene Collection, National Human Genome Research Institute, National Institutes of Health, 5635 Fishers Lane, Suite 4079, Bethesda, MD, 20892, USA, USA

<sup>32</sup> Department of Genetics, Harvard Medical School, 44 Binney Street, Smith 858, Boston, MA 02115, USA

## References

1. The Human Proteomics Standards Initiative. <http://psidev.sourceforge.net/>
2. Rual, J.F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173-1178 (2005).
3. Arlt, D. *et al.* Functional Profiling: From Microarrays via Cell-Based Assays to Novel Tumor Relevant Modulators of the Cell Cycle. *Cancer Res.* **65**, 7733-7742 (2005).
4. Quackenbush, J. Data standards for 'omic' science. *Nat Biotechnol* **22**, 613-614 (2004).
5. Brazma, A., Krestyaninova, M. & Sarkans, U. Standards for systems biology. *Nat Rev Genet* **7**, 593-605 (2006).
6. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genet.* **29**, 365-371 (2001).
7. Orchard, S. *et al.* Common interchange standards for proteomics data: Public availability of tools and schema. *Proteomics* **4**, 490-491 (2004).
8. Brazma, A. On the importance of standardisation in life sciences. *Bioinformatics* **17**, 113-114 (2001).
9. Harris, M.A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258-261 (2004).
10. Parkinson, H. *et al.* ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **33**, D553-555 (2005).
11. Orchard, S. *et al.* Autumn 2005 Workshop of the Human Proteome Organisation Proteomics Standards Initiative (HUPO-PSI) Geneva, September, 4-6, 2005. *Proteomics* **6**, 738-741 (2006).
12. Functional Genomics Experiment (FuGE). <http://fuge.sourceforge.net/>
13. Orchard, S. *et al.* Proteomics and Beyond A report on the 3(rd) Annual Spring Workshop of the HUPO-PSI 21-23 April 2006, San Francisco, CA, USA. *Proteomics* **6**, 4439-4443 (2006).
14. Quackenbush, J. *et al.* Top-down standards will not serve systems biology. *Nature* **440**, 24 (2006).

## MIACA Full Copyright Notice

Copyright © MIACA Standards Initiative (2006-2007). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the MIACA Standards Initiative or other organizations, except as needed for the purpose of developing MIACA Standards Initiative Recommendations in which case the procedures for copyrights defined in the MIACA Document process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the MIACA Standards Initiative or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE MIACA STANDARDS INITIATIVE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.