UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
DIPARTIMENTO DI SCIENZE DELL'AMBIENTE E DEL TERRITORIO

DOTTORATO DI RICERCA IN BIOLOGIA - CICLO XXIV

QUANTITATIVE EVALUATION OF *IN VITRO* TRANSFORMATION BY ANALYSIS OF MORPHOLOGICAL AND BIOCHEMICAL MARKERS AND STATISTICAL DESCRIPTORS

Claudio Procaccianti
Matr. 725269

Coordinator: Prof. Enzo Wanke

Tutor: Dr. Chiara Urani

External Tutor: Prof. Federico M. Stefanini

Academic Year 2010/2011

Acknowledgements

First and foremost, I want to thank Dr. Chiara Urani of Università degli Studi di Milano Bicocca, Dipartimento di Scienze dell'Ambiente e del Territorio and Prof. Federico M. Stefanini of Università degli Studi di Firenze, Dipartimento di Statistica "G. Parenti", for their invaluable insight, drive and guidance, both on an academic and on a personal level throughout these three years.

I want to thank all the Molecular Biology and Genomics Unit of the Institute for Health and Consumer Protection, Joint Research Centre (Ispra, VA) and in particular Dr. Laura Gribaldo, for the kind collaboration on all the genomic experiments, and Dr. Raffaella Corvi of the *In vitro* Methods Unit of the European Centre for the Validation of Alternative Methods, Institute for Health and Consumer Protection, European Commission Joint Research Centre (Ispra, VA) and her unit, for the scientific material made available for the generation of the *foci* images archive.

I want to thank all my colleagues of Bicocca, which made this experience as wonderful as it was. Thank you for all the knowledge you shared with me, and for making every working day a pleasant day as well.

Finally, I want to thank my family for all the support they showed during these years. Thank you for believing in me, for pushing and supporting me…and for having all the patience in the world to deal with me. I would have never made it this far, if it wasn't for you.

Table of Contents

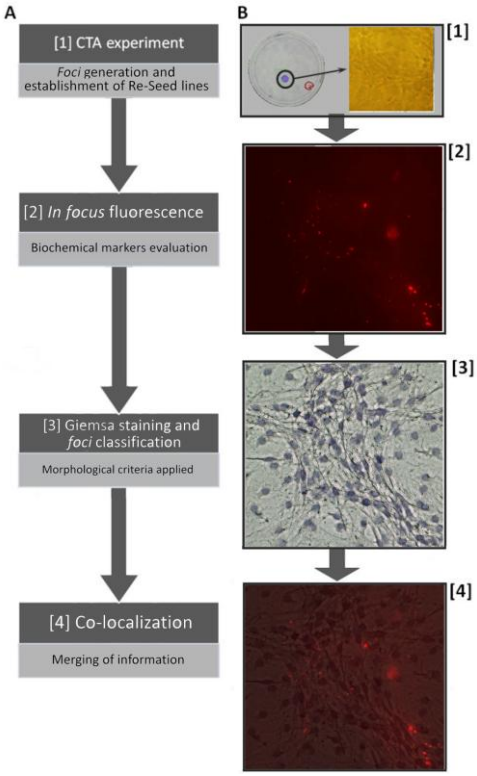
Abstract

The EC directive REACH (EC1907/2006) sets, amongst others, the need for all the chemicals to be tested for their carcinogenic potential. *In vitro* assays can provide a fast and reliable tool for screening purposes. The Cell Transformation Assay (CTA) is one of the *in vitro* assays in the most advanced phase of the validation process and the only one able to evaluate both the genotoxic and the non-genotoxic potentials.

The evaluation of results of the CTA is based on the scoring of transformed colonies (*foci*) by a trained expert on the basis of their morphological features. Levels of cell packing and multilayered growth, as well as fibroblastic shape of cells, criss-crossing and invasion of the surrounding monolayer features are evaluated for classification. While the decision making process is based on standard criteria, their interpretation is potentially biased, especially in borderline cases, due to a certain degree of subjectivism inherent in the evaluation of qualitative features. This aspect is critical towards the international validation of the CTA assay: subjectivity driven error might in fact result in under or over estimation of the carcinogenic potential of tested compounds.

In this work, different approaches were used to develop an objective method to give decisional support to the operator in the classification procedure. Biological markers related to the transformation process (p53, cx43), and to a general cell stress (Hsp70) were analyzed. A novel technique for the *in focus* localization of biological markers of transformation was developed **(Flowchart and Graphical Abstract – *In-Focus* Co-localization)**. RNA whole genome screening was used to set the conditions for future molecular characterization of *foci*-derived cell lines. A novel, Quantitative Index of Dissimilarity has been obtained by statistical descriptors capturing morphological features and employing an unsupervised image analysis approach, in order to help the operator in the decisional process of scoring the borderline cases **(Flowchart and Graphical Abstract – *In Silico* Approach)**.

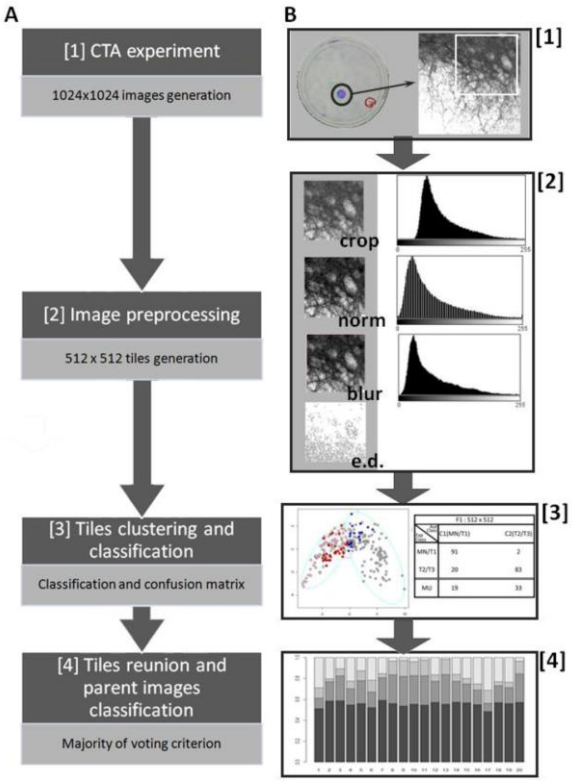
Keywords: biochemical markers of transformation, cadmium, carcinogenesis, cell transformation assay, cluster analysis, quantitative index of dissimilarity, RNA fingerprinting, unsupervised learning.



Flowchart and Graphical Abstract – *In-Focus* Co-localization.

A. *Foci* obtained from a cell transformation assay [1] were either collected for establishment of a Re-Seed cell line, or fixed and used for fluorescence experiments [2]. Re-Seed cell lines were established for future molecular and biochemical characterization. Preliminary studies were performed on HepG2 cell lines to standardize the protocol and investigate the effects of cadmium on the whole genome of tumoral hepatic human cells (not shown in figure). Petri dishes were then Giemsa stained and the standard process of scoring was carried out [3]. Images obtained from both fluorescence and Giemsa staining were studied for superimposition in order to evaluate the co-localization of biochemical markers with characteristic morphological structures of *foci* [4].

B. 100mm petri dishes were followed for all the CTA duration and screened for *foci* formation since the 3rd week. Transformed colonies were recognizable even before fixing and staining (A Giemsa stained petri dish is shown for exemplificative purposes). *Foci* were harvested under sterile conditions and seeded in new dishes. Growth of the new lines was daily monitored and images taken to demonstrate the *focus-like* patterns of growth [1]. Non-collected *foci* were fixed, isolated by pap-pen tracing and tested for Cx43 localization. Images were taken at 400x magnification [2]. Standard scoring process was carried out and images of Giemsa stained *foci* corresponding to areas tested for Cx43 taken at 400x magnification [3]. Computer assisted superimposition of fluorescence and Giemsa staining was evaluated [4].



Flowchart and Graphical Abstract – *In Silico* Approach.

A. 1392x1040 images generated from CTA experiments [1] were pre-processed and tile cropped [2]. Obtained tiles were clustered by means of the PAM algorithm, and expert *versus* classifier agreement evaluated by means of confusion matrices [3]. Parent images classification was inferred from the cluster assignation of derived tiles by applying a majority of voting criterion. **B.** Images from intensely Giemsa stained colonies were taken at 25x magnification at an original resolution of 1040x1329, then resized to 1024x1024 [1]. Images were then tile cropped ([2], crop), normalized over a 0-255 arc ([2], norm) and filtered by Gaussian blurring ([2], blur). Descriptors were obtained from the latter step as summary statistics of the gray levels distributions. Tiles were also filtered by Canny's edge detection algorithm to extract an edge-specific descriptor ([2], e.d.). Obtained descriptors were used by the algorithm to cluster the images in *k* groups based on the number of biological classes ([3], left panel. *k* parameter is set to 2 in the example, as per positive and negative biological classes). Expert scores and automatic classification scores are compared in a confusion matrix to assess agreement, sensitivity and specificity of the corresponding step of classification ([3], right panel). Majority of voting criterion was applied to infer the parent images scoring from the attribution classes of the derived tiles. Finally, the Quantitative Index of Dissimilarity was calculated for each Morphologically Undefined parent image to each of the four classes (MN, T1, T2, T3) [4] **(Procaccianti et Al., 2011).**

Work Overview

The evaluation of the carcinogenic potential of chemicals is a matter of first importance. The only assay validated at regulatory level for the assessment of the transforming potential is a two year *in vivo* bioassay to be performed in rodents. Such tool is not suited for screening purposes due to its costs, both monetary and in terms of animal lives and hours of work. Moreover, there is a still hot debate in the scientific community regarding the validity of its results when applied to human carcinogenesis. In 2006, the European Community released a directive (REACH - EC 1907/2006) that stated that all the chemicals, both those already available and all the future ones to be released on the market, must be evaluated for their potential to induce cancer in human. This document states as mandatory the use of *in vitro* methodologies for screening purposes and only when positive results arise, the use of *in vivo* testing is allowed.

Due to both the economic costs and the debate on the validity of results of rodents assays, and given the recent EC directive resulting in a huge need for fast and reliable screening tests, a growing interest of the scientific community has been focusing on the research of *in vitro* alternatives to the *in vivo* bioassay. The Cell Transformation Assay (CTA) is one of the *in vitro* methods in the most advanced phase of the validation process. A strong advantage of the CTA, besides the cheap costs in comparison with the *in vivo* bioassay and the strong reliability of

its results, is its flexibility in testing both genotoxic and non-genotoxic transforming potential of chemicals. This assay is based on the evaluation of the morphological features of cellular colonies (*foci*), obtained after treatment with a test compound in suitable cell lines, by an expert operator. The *foci* obtained in the C3H cell transformation assay are scored and divided in classes (Type I, Type II and Type III) on the basis of set criteria and corresponding to different levels of transformation.

Nevertheless, for how well trained the operator might be, this system of evaluation is theoretically prone to errors due to the subjective nature of the evaluation of the morphological features process. This critical point is even more important when scoring *foci* that don't strictly fall in the canonic classes, but exhibit an intermediate or a mixed phenotype. This subjectivity in scoring leading to a possible over or under estimation of the transforming potential of a compund is one of the most critical factors hampering the process for the international validation of the CTA for the assessment of the chemical carcinogenesis. The experimental grounds set for this work were the effects of cadmium salts treatment on the C3H cell line. Though cadmium has widely been accepted as a human carcinogen, most of the molecular mechanisms underlying its carcinogenic potential remain elusive.

Aim of this work was to develop a system based on objective criteria to give decisional support to the operator in the critical phase of *foci* scoring.

In order to achieve this goal, the work focused on an integrated approach of different methodological approaches.

1. A new technique allowing to study the presence of biochemical markers of transformation *in-focus* without altering the morphological structures needed for the classification has been developed.

2. In perspective of the molecular characterization of cells transformed by a chemical as obtained at the end of a CTA, a method for the isolation and establishment of a stable cell line obtained from *focus*-derived cells was developed (Re-Seed lines). Information obtained by means of whole genome microarray analysis can help shedding some light on the molecular mechanisms underlying cadmium-induced transformation. Moreover, RNA profiling can be used as a candidate gene approach for the choice of biochemical markers of transformation for the *in-focus* analysis. The preliminary work for Re-Seed screening has been carried on in human hepatoblastome HepG2 cell line, established from an *in vivo* target organ of cadmium toxicity.

3. Image analysis techniques were employed to transform qualitative-subjective features such as the morphology of *foci* in quantitative-objective data. Different *ad hoc* algorithms were developed in order to extract statistical descriptors from the features of *foci*. A classifier of the unsupervised type was built to exploit the extracted descriptors for a bias-free classification approach. A Quantitative Index of Dissimilarity was developed as a measurement of the distance/difference between

the morphology of undefined, border line *foci* and each of the canonic classes, in order to give decisional support to the operator.

The integration of *in vitro* and *in silico* techniques might in future result in a multi level, high content analysis of *foci*, in order to make an objective gold standard for the classification process.

List of Used Abbreviations

3Rs – Reduction, Replacement and Refinement
B[a]P – Benzo[a]pyrene
BER – Base Excision Repair System
BME – Basal Medium Eagle
C1, C2 – F1 Classes of Attribution (hierarchical approach)
C'1, C'2, C'3, C'4 – F2 Classes of Attribution (hierarchical approach)
C3H – C3H10T1/2 Cl8 Mouse Embryo Cell
Cx43 – Connexin 43
DSB – Double Strand Brakes
GJIC – Gap Junction Intercellular Communication
GSH – Reduced Glutathione
HR – Homologous Recombination Repair System
HSP – Heat Shock Protein
IARC – International Agency for Research on Cancer
IRE – Iron Responsive Element
IRE-BP – Iron Responsive Element Binding Protein
MAPK - Mitogen-Activated Protein Kinase
MD – Morphologically Defined
MMR – Mismatch Repair System
MT – Metallothionein
MU – Morphologically Undefined
NER – Nucleotide Excision Repair System
NHEJ – Non Homologous End Joining Repair System
OECD – Organisation for Economic Co-operation and Development
PAH – Polycyclic Aromatic Hydrocarbon
PAM – Partitioning Around Medoids
PBS – Phosphate Buffer Saline
PE – Plating Efficiency
PKA – Protein Kinase A
PKC – Protein Kinase C
QIoD – Quantitative Index of Dissimilarity
ROS – Reactive Oxygen Species
S – Survival
TF – Transformation Frequency
TPA – 12-O-Tetradecanoylphorbol-13-Acetate
XP - Xeroderma Pigmentosum

Chapter 1: Introduction

1.1 Cancer as a Genetic Disease

It is possible to define carcinogenesis as a “genetic pathology”, which final product is the disease (or rather, the set of diseases) called cancer. However, unlike most of the traditional genetic pathologies, cancer is not associated to a single gene which determines the phenotype. For example, in the case of the Duchenne dystrophy, a direct link between the disease and genetic alterations of the genes coding for dystrophin is known **(Hoffman et Al., 1987)**. Complex and extremely various as the causes for tumoral transformation might be, comprehending environmental factors as well as viruses and, in some cases, a genetic predisposition transmitted within the family line, the final result is the appearance of mutations among the genes that regulate the control of the fundamental processes that lie at the bottom of the regulation of the cell cycle.

In homeostasis conditions, proliferation and programmed cell death (apoptosis) processes are tightly related and in reciprocal balance. On the other hand, in tumoral pathologies this balance is lost in favor to one or another part. Genes involved in these processes are defined *gatekeeper* genes, and their loss brings alterations in the homeostatic balance, with uncontrolled proliferation of mutated cells. Another category of genes strongly involved in the transformation are the genes linked to the DNA system repair, the so called *caretaker* genes. Loss of

such functions brings an increase in the frequency of DNA mutations and their fixing in the host genome.

Regarding the transmission of the disease, two forms of cancer can be distinguished: sporadic and familiar. In sporadic forms, mutations are acquired casually. In familiar forms it is not possible to define a classic mendelian transmission of the disease. Instead, it is the inheritance of a complex genetic frame that brings predisposition to develop a particular form of cancer. In fact, human healthy diploid cells have a double copy of every gene. In familiar forms, usually just one of the two alleles involved in the tumoral process is transmitted **(Voegelstein et Al., 2002)**. For example, in the case of familiar forms of colon carcinoma linked to the APC gene, hereditary forms of the pathology sum up to just the 1% of cases. This is because, much as all the cells in the individual with a diseased allele are at risk of becoming tumoral, this predisposition is not sufficient for the development of tumor itself. A second event is needed to inactivate even the second allele, as explained in Knudson's “double hit” theory **(Knudson, 1971)**. However, the high number of cells at risk makes patients with a genetic predisposition not develop just one tumor, as in the more frequent sporadic cases, but a multitude of tumors and at an average younger age. More examples of hereditary forms are represented by retinoblastoma (RB gene) or Li-Fraumeni syndrome (TP53 gene). In all of those cases, involved genes codify for proteins with an oncosuppressant function. In fact, once that both genes are

deactivated, their function is compromised and the cell starts to proliferate in an abnormal manner, within a process called “clonal expansion”. Given the abnormal proliferation speed and the lack of operating caretaking functions, those cells are characterized by a strong predisposition to the occurrence of more genetic mutations (tumor progression).

In both familiar and sporadic forms, mutations that fix themselves in the host's genome are needed. As a matter of fact, the genetic pool of a cell is constantly stressed by a series of endogenous and external factors. UV radiations, alkylating agents, polycyclic aromatic hydrocarbons (PAHs) as well as endogenous processes as oxidative stress induced by radicals generated by losses in the electron transport chain during cell respiration provoke damages and modifications in the genetic material of the cell. In those cases, a healthy cell faces a split decision: repair the damage or, if unable, start the programmed death pathways. Initially, a primary block intervenes, stopping the progression of the cell cycle waiting for decisions; subsequently, enzymes involved in the damage repair pathways intervene. One of the most important roles in this step relies on the p53 protein **(Ozaki et Al., 2011; Levine et Al., 2011)**. Given the importance of genome preservation, numerous repair systems were developed throughout evolution and are based on several different molecular mechanisms, depending on the type of damage:

- *nucleotide excision repair system* (NER) recognize structural alterations of the double helix structure and removes whole aberrant nucleotides
- *base excision repair system* (BER) removes just the bases damaged by, for example, alkylating of oxidizing agents, leaving the DNA sugar-phosphate frame intact.
- *mismatch repair system* (MMR) recognizes and solves problems related to a wrong bases pairing.
- *homologous recombination repair system* (HR) and their *non homologous* counterparts (NHEJ) deal with reparations in case of double strand breaks (DSB)

Genetic alterations typically fall under five typologies: small alterations (e.g. Frame shift mutations, missense mutations), variations in the number of chromosomes (trysomy, aneuploidy), chromosomal translocations (chronic myelogenous leukemia with the t(9;22) translocation and the chimerical Philadelphia Chromosome formation), and amplifications and insertions of external sequences (oncogenic viruses). **(Vogelstein et Al., 2002)**

1.2 Chemical Carcinogenesis

A carcinogenic substance can by definition induce the formation of tumors or increase their malignancy or the frequency of their insurgence. Mechanisms of actions of cancerous substances fall in two classes: genotoxic mechanisms and non-genotoxic mechanisms. Substances

belonging to the first class, as, for example, PAHs like methylcholantrene and benzo[a]pyrene exercise their transforming potential by inducing direct genotoxic damage to the host's genetic material, such as strand breaks, modification of the chemical structure of the bases and nucleotides, modifications and alterations of number and chromosomes integrity. Compound belonging to the second class, instead, are substances that don't show a potential for direct interaction with DNA, but rather act through alternative mechanisms, such as oxidative stress, inhibition of intracellular communication or alterations in the genetic material damage repairing systems. In fact, the occurrence of damages to the genetic material is part of the physiological events that happen during the life of a cell. One example might as the generation of reacting oxygen species (ROS) due to the mechanisms of electron transportation at the base of oxidative respiration. A logical consequence of a failure of the systems assigned to repairing such damage is the genome destabilization. We can find an example of the importance of the repairing systems in the *Xeroderma pigmentosum* (XP) genetic disease. Genes involved in the *nucleotide excision repair system* (NER) mechanisms of DNA damage repair are responsible for this pathology. In fact, patients affected with XP present an extreme tendency to develop lesions of the retina, keratosis and skin tumors due to formation of pyrimidin dimers as induced by UV rays, even in case of simple unprotected exposition to solar rays.

Carcinogenesis is a very complex biological process, that can be virtually divided in different phases (multi-step process) and that is strongly influenced by a variety of factors such as age, diet, hormones and environmental factors. This great number of possible interactions makes realizing a unique experimental model to evaluate this biological event extremely complex. At the present state of international regulations, the normative in force promulgated by OECD (*Organization for Economic Co-operation and Development*) regarding the study of carcinogenesis is the *test guideline 451* (TG451). In this document, first published in 1981, the standard reference tests to assess the carcinogenic potential of substances are defined. Such guidelines allow the administration of chemicals to be studied in selected animals only in three possible ways (oral, cutaneous, inhalatory), depending on the physical and chemical characteristics of the test substance and the routes of exposure in humans. Housing of the animals must be kept in controlled humidity conditions with 12 hours cycles of artificial light and darkness. It is mandatory to use for tests animals of both sexes, always young and in healthy conditions. At least 50 animals per testing group for each sex must be used, for at least 3 dose testing groups plus controls. Duration of the study normally is of 24 months for rats and 18 months for mice and hamsters. The issue of these guidelines was an important moment in the history of *in vivo* experimentation. In fact, even if normally used since the 60s, this document, published only 20 years later, represented the first

regulation on animal experimentation, detailing not only experimental protocols, but animals welfare conditions (nutrition, housing) as well **(OECD, 1981)**. The normative TG451, together with the normative TG452 (long term toxicity studies) and TG453 (combined studies about chronic toxicity and carcinogenesis) constitute the whole of norms and provisions to regulate animal experimentation and rationalize animal usage at an international level.

Animal models are not only used in the assessment of the carcinogenic potential of a compound. In fact, they also have an important role in the study of molecular mechanisms and pathways involved in genetic diseases and in the development of tumoral forms, as well as in the study on drug and cosmetic products toxicity. Some of the more widespread model use transgenic mice knock-out for gatekeeper genes. Other models are based on the induction of the diseased phenotype through chemical agents, like diabetes induction in obese mice though the destruction of pancreatic β cells with streptozotocin. All these models, however, suffer from two important limitations: simplified models as they might be, the interpretations of results can be strongly complicated by the variety of interactions that take place in an organism complex as a mammal; moreover and more importantly, there is a lot of unresolved debate in the scientific community regarding the actual validity of data gleaned *in vivo* from rodent or other laboratory animals when in relation

to possible effects of the same substances on humans **(Knight et Al., 2006)**.

The listed considerations about the effectiveness of *in vivo* models, together with ethical questions about experimentation on animals of pathological conditions that may be very painful, made so that since the end of the 80s the interest of a growing part of the scientific community focused on the research of new experimental models. More than being scientifically relevant, those new models should have an ethic value in respect to animal life, causing the less possible damage to laboratory animals. This philosophy has been outlined in the acronym “3Rs” and is based on the principles of *replacement*, *reduction* and *refinement*. At the actual state of the art, there is no model which enables the complete *replacement* of animals in the study of carcinogenesis. In fact, as affected by intrinsic limitations as the *in vivo* models might be, they still are the most complete model for the study of chemical carcinogenesis, as well as the only experimental model actually validated at an international level by OECD. On the other hand, models that enable *reduction* in the number of used animals and *refinement* of experimental conditions and their usage do exist, and it is considered mandatory when validated. An example, the models that will be exposed hereafter allow to reduce the number of used animals to 20-25 for group and to a period of 6-9 months, compared to the 2 years needed for the conventional trial. It is important to notice that those models can employ transgenic animals.

It is then good practice to add wild-type specimen in the controls, so to exclude the possibility that the genetic drift of the animal results in an erroneous interpretation of the results, and so in a potential under or overestimation of the transforming potential of a substance. It is also important to mention the tests performed on synthetic organs (e.g. synthetic skin) as very useful in the study of toxicity in cosmetic products. Also, *in silico* models that allow to partially foresee the behavior of a molecule on the basis of its molecular structure and affinity with already known molecules (e.g. QSAR models) are available **(Knight et Al., 2006)**.

1.3 Examples of Models Based on Transgenic Mice

An important tool allowing the reduction of the number of tested animals, in compliance with the principles of the 3Rs, is the use of transgenic mice models. Both the tg.AC and the p53+/- transgenic models, herewith described as an example of two widely used models, are not just useful to evaluate the transforming potential of chemicals, but are also able to give insights regarding the molecular mechanisms of action of cancerous substances. However, these models still are in need of more studies in order to be validated at regulatory level by the competent organs **(Maurici, 2005; Knight et Al., 2006).**

1.3.1 Tg.AC Mice

Those transgenic mice have a knock-in construct that expresses the v-H-ras oncogene under control of the zeta-globin promoter. This oncogene codifies for a constitutionally active form of the monomeric G protein with kinase activity c-Ras. This animal represents a useful model in the study of skin cancer. While wild-type animals require the co-administration of the chemical to be studied with an initiating agent for the skin, in Tg.AC mice the presence of the activated oncogene allows studying the action of the chemical without the need to initiate the skin. Thus, the transgene mimics the priming effect of the initiating agent. One of the most important critical point of this test is in the impossibility to distinguish between carcinogenetic effects due to genotoxic agents from the one induced by promoting agents.

1.3.2 Heterozygous p53 +/- Mice

p53 protein has a fundamental role in the maintenance of genome stability. A condition of heterozygosis for p53, micking the genetic background of the Li-Fraumeni syndrome, renders these transgenic mice highly susceptible to the development of tumors, compared to wild-type animals just like in the case of patients with the corresponding genetic syndrome. This model is mostly utilized in fast studies with a duration of about 26 weeks. In fact, in such a brief period, it is not possible to observe an increase in the frequency of spontaneous insurgence of

tumors in control animals. In transgenic animals, the loss of the healthy residual allele happens as a consequence of mutations induced by tested substances, resulting in the formation of tumors. This is the model of transgenic animal nowadays more utilized for *in vivo* transformation tests.

1.4 Analogies Between *In Vivo* and *In Vitro* Transformation Processes

Transformation processes, both in clinic and in the laboratory tests i*n vivo* as well as *in vitro*, are strictly related. In fact, in all of those biological processes, the transformation happens through sequential phases, through a multi-step process.

From a clinical point of view, tumoral lesions can be divided in three groups: pre-malignant lesions, primary tumors and metastasis. In the first stadium, lesions can be characterized at tissue level by dysplasia, hyperplasia, leucoplakia or adenoma. The transformed cells belonging to these lesions expand in a clonal way, mostly due to the selective advantages acquired through genetic alterations. Moreover, they can become less responsive to stimuli of negative regulation of growth and terminal differentiation. With the accumulation of those alterations and with the consequent deregulation of cell control mechanisms, cells acquire a malign phenotype, resulting in the production primary tumors. Those cells are often not responsive to signals of growth inhibition, and

proliferate in a fast and indiscriminate manner, forming local masses (*tumoral masses*). New clones tend to accumulate different mutations, as a result of their genetic instability. Those new mutations can give the tumor characteristics of more invasiveness and tendency to form metastasis. We can, then, virtually divide the process of formation of a tumor in three steps: initiation, promotion and progression. Such processes have, however, been studied only indirectly in human, mostly through the evaluations of tumoral incidence in relation to age of insurgence. Extrapolated results suggest that in a majority of cases, a number of different genetic events spanning from 4 to 6 are necessary for a malign tumor to develop. More specifically, all the data seem to point out that two independent genetic events are necessary to activate both copies of an oncosoppressor gene. It is also necessary that two or three genes involved in the regulation of the growth or in the damage repairing systems are lost for the tumoral development to kick in. This pattern of activation is coherent with the described steps: initially, the cells lose the possibility to differentiate (initiation step, with the loss of oncosoppressor gene function) and, then, start to proliferate in an uncontrolled manner and acquire a malign phenotype (promotion and progression steps, genomic instability and accumulation of mutations). Coherently with the statistical studies, in fact, genetic studies founded that among the various alterations encountered in human tumors, the inactivation of oncosoppressor gene is the most frequent. Therefore, it is

commonly accepted that the malign phenotype acquired by the tumoral cells through the various steps of the transformation process is defined by genetic alterations accumulated during the tumoral progression process.

In the carcinogenesis induced by chemical substances in animals, a group of animals in standard conditions is treated with the chemical agent in study. Other groups are utilized as negative control (e.g. vehicle only) and as positive control (e.g. treated with a known carcinogen). Even if the kinetic of absorption and distribution of the administrated molecule results in a more or less equal repartition of the substance in the whole animal, only some cells actually respond to the treatment with the chemical agent acquiring mutations. Speaking of “initiated animal” would therefore be inaccurate, given that only some of his cells acquire that characteristic, becoming actually “initiated”. Acquired characteristics permit the acquisition of selective advantages, like the loss of the possibility of terminal differentiation or a proliferative advantage compared to normal cells. Mutated cells become “immortal”, not showing neither senescence processes nor terminal differentiation or contact inhibition. This aspect correlates with the response to non-mutagenous but mitogenic stimuli, like in the case of administration of phorbol esters in skin tumor models *in vivo*. In fact, though all the cells of the animal are equally exposed to the administered promoting agent, only some of them that have already acquired stable mutations, spontaneously or caused by

an initiator agent as well, will generate tumors, coherently with the clinical data.

In the same way, even in the *in vitro* carcinogenesis it is possible to recognize a series of sequential passages that can be brought back to the multi-step nature of the transformation process. Like in the *in vivo* studies, the cells utilized in these models are seeded in standard conditions, then treated with the agent to be studied and compared with positive and negative controls. A part of the cells will show morphological changes typical to the acquisition of genetic mutations, in parallel to what happens to patients and to “initiated cells” in the treated animal. The loss of contact inhibition, the capability to grow in soft agar and the capacity of invasion of the normal surrounding monolayer represent the subsequent passages of the transformation process, coherently with the acquisition of a genomic instability and the accumulation of new mutations and invasion. The proof of the transformed nature of the cells that have acquired all those characteristics has been obtained with the subcutaneous injection in naked immunocompromised mice. In fact, injected cells form local tumors in a percentage that varies coherently with the degree of advancement of the transformation processes.

An example of this parallelism can be done if considering two events with a central role in the transformation process, like the fixation of the mutation and the expression of the transformed phenotype. Both events appear in the animal as well as in the test *in vitro*. Moreover, studies

indicate that of those two processes, the first appears with elevated frequency, while the second is much more rare. In the *in vitro* transformation, this second phase seems to be strongly influenced by the seeding density. One of the possible mechanisms at the bottom of such a behavior might be in the transfer of signaling molecules with the nearby cells: this transport appears blocked between healthy cells and transformed cells, while it is possible between healthy cells or between transformed cells. This cell communication is made possible by the proteins of the connexin family. They are membrane proteins that form canals between two contiguous cells, one semicanal each, thus bringing physically in communication cytoplasms and permitting the free passage of small molecules. It is likely that this phenomenon can be partially accounted for the so called *bystander effect*: the presence of healthy cells seems to act as a strong inhibitory signal for the transformed cells. Moreover, as a further confirmation of the closeness between transformation processes *in vitro* and i*n vivo*, the damage profiles of various substances are quite similar between the two tests: in fact, a percentage that varies between 69 and 85% of cancerous compounds *in vivo* has demonstrated an analogous capacity of transformation *in vitro*. On the other hand, while at a regulatory level, the reference test for the evaluation of carcinogenesis is the *in vivo* test performed on rodents **(OECD, 1981)**, strong discrepancies exist between the carcinogenic potential of compounds in rodent and in human **(Knight et Al., 2006)**.

Substances showing a transforming potential in *in vitro* tests could be studied *in vivo* to obtain a more complete profile of toxicity, regarding the potency of the agent and the possible target organs. The interest of the scientific community in relation to the *in vitro* methods has become even stronger after the issue by the European Community of the REACH directive **(REACH EC1907/2006)**. Due to the huge amount of substances to be tested, assays that are fast, reliable and cheap as the *in vitro* methods, might cover a lead role in the enforcement of the directive. In short, this document asserts that for every substance present or to be put in commerce there is the mandatory need for a complete dossier that evaluates its toxic or carcinogenic potential. This document has therefore generated a strong need, especially for companies, to employ new methods in order to face the high costs, both in terms of time and money, that the *in vivo* test on rodents would imply. Moreover, the problems in the transportability of the results obtained in rodents give even more emphasis to the need of more reliable assays.

1.5 *In Vitro* Cell Transformation Assay

Though there are already various tests validated at regulatory level that assess the genotoxic and mutagenous potential of substances both in eukaryotic and prokaryotic cells (micronuclei formation tests, comet assay, Ames test), no *in vitro* assay actually effective in the study of both

genotoxic and non-genotoxic agents has been validated yet **(Maurici, 2005)**.

Cell transformation assays allow to evaluate the carcinogenic potential through the study of morphological changes induced by chemical compounds *in vitro* in mammal cell systems. Typically, those tests are used both in the academic research field and for screening by pharmaceutical and cosmetic industries, as well as by environmental protection agencies (ARPA) **(Maurici, 2005)**. In brief, the standard protocol schedules the seeding of cells in petri dishes and the treatment with the agent in study for a period of 24-48 hours. After a period of more or less 6 weeks of recovery, cells are fixed and stained. Dark colonies that are told apart against the monolayer are scored depending on their morphological characteristics **(Figure 1.1, and below, Figure 1.2)**.

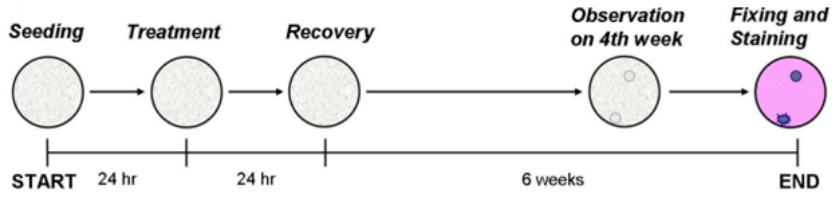


Figure 1.1. Schematic Overview of a C3H Cell Transformation Assay.
At day 0 cells are seeded at a low density. After 24 hours treatments are performed. At the end of the treatments, a recovery period of 6 weeks starts. At the end of it, cells are fixed and Giemsa stained. *Foci* are spotted as dark colonies against the normal monolayer **(Urani et Al., 2009)**.

The three CTAs are in the most advanced stadium of validation are: SHE **(LeBoeuf et Al., 1990)**, C3H10T1/2 Cl8 **(Mondal et Al., 1980)**, BALB/c 3T3 **(Aaronson et Al., 1968)**. The SHE cell line derives from hamster embryonic cells and presents a normal karyotype. This test is particularly indicated to individuate the early steps of carcinogenesis. The C3H10T1/2 and BALB/c 3T3 lines are derived from murine fibroblasts with aneuploid karyotype. The tests based on those two lines are more suitable to study the late steps of carcinogenesis **(OECD, 2007)**. Cytotoxicity of compounds in study in those tests is evaluated depending on the capacity to form colonies or on the alterations of proliferation speed. To do so, cells are to be plated and treated, and, after 10 days the survival is evaluated as a ratio between the number of formed colonies between the treatment and control groups. Cell transformation tests feature the great advantage, unique among *in vitro* tests, to have the possibility to be modified so to be fitted to study a chemical not only for its role as initiator of transformation, but also as promoter or as a molecule with protective activity, such as in case of the two-stage CTA. In this variation, cells are plated and then treated with an agent of known genotoxic activity (methylcholantrene, MCA) and then treated with concentrations of the substance whose activity as promoter or protective is under evaluation. Results are compared with sub toxic doses of a known promoter (phorbol ester, TPA).

On the other hand, a critical limit for this test lies in the study of substances that necessitate of bio-activation. The lack of some of the phase I or phase II enzymes might, in fact, give false results in the study of substances that necessitate of a metabolic activation before exercise their toxicity. Some cell lines can, however, express some of such enzymes. Alternatively, it is possible to introduce modifications *ad hoc* to the experimental protocol, typically by adding hepatic extracts S9. Even if cell transformation tests *in vitro* are normally used in various laboratories, studies about inter- and intra-laboratory reproducibility are still very rare. The European Center for Validation of Alternative Methods (ECVAM), has, in fact, promoted a series of initiatives that involved European, Japanese and USA laboratories promoting international validation **(OECD, 2007)**.

1.6 Characteristics of Cell Lines Used in Transformation Assays

The purpose of a transformation assay is to evaluate the carcinogenic potential of chemical substances in a specific manner. An optimal model, then, should be characterized by high sensitivity and specificity to chemical agents, and, at the same time, should feature a low frequency of spontaneous transformation, in order to minimize the errors in the estimation of the actual carcinogenic potential due to an excessive rate of “background mutations”. In the case of CTA in C3H cell line, cells that haven't surpassed the 21° passage are characterized by a very low

frequency of spontaneous mutation. The BALB/c cell line present a higher frequency of spontaneous transformation. However, this latter cell model presents a very high sensibility to carcinogenic chemicals. Numerous chemical compounds are not toxic by themselves, but they need activation through biochemical reactions for their toxic potential to manifest. Those enzymatic reactions (e.g. oxidation, reduction, conjugation with macromolecules) are divided in phase I reactions (modification) and phase II reactions (conjugation) and are normally effectuated by specialized cells of an organism in order to metabolize and make easier the elimination of xeno-biotics (processes of detoxification). An important propriety of the cell lines used in the *in vitro* tests is the expression of mixed-function oxidase activity enzymes, that can metabolize and bio-activate through phase I reaction various compunds. However, a substance might necessitate not only of phase I reactions before becoming active, showing its potential only after the conjugation with other macromolecules (e.g. glucuronic acid). In this case, the transformation test might give a negative result due to the lack of activation rather than an actual absence of toxicity. In the case a substance gives a negative result during a transformation test, before declaring the non toxicity, it is advisable to repeat the test in presence of S9 hepatic extract, rich in enzymes capable of performing further reactions of bio-activation **(OECD, 2007)**.

1.6.1 Characteristics of the C3H10T1/2 Cl8 Cell Line

This cell line has been selected so to be characterized by a high level of growth inhibition upon reaching confluence. This characteristic has been obtained growing the primary culture at a low density of population through a strict regime of passages effectuated every 10 days at a concentration of 0.5 x 10^5 cells per petri of 60mm of diameter. **(Reznikoff et Al., 1973; OECD, 2007)**. Cells of the C3H10T1/2 line have a fibroblastic aspect. However, they are capable to revert at their embryonic origin, showing a behavior similar to the one of stem cells, with the capacity to differentiate themselves in various different cell types, if correctly stimulated. **(Taylor et Al., 1979; Hirschi et Al., 1998)**. When very young (15-21 passages) those cells present a very low frequency of spontaneous transformation, being so optimal for their use for *in vitro* transformation assays. Moreover, given the strong level of post confluence inhibition, it is important that the cells are always kept at sub-confluence levels before starting a cell transformation assay.

This cell line presents a high level of sensibility to a vast spectrum of carcinogenic compounds, which manifest itself in the development of well defined colonies of transformed cells against the non-transformed monolayer, that go under the name of *foci* **(OECD, 2007)**. Moreover, C3H cells express phase I enzymes with mixed-function oxidase activity, being thus useful for the study of chemicals in need of bio-activation.

1.7 Protocol for the C3H10T1/2 Transformation Assay

Transformation assay done on C3H presents the need to use cells at extremely low passages (15-21) so to have the minor possible frequency of spontaneous transformation, as previously noted. It is then advisable, once obtained young cells, to expand them as fast as possible and do numerous aliquots of young cells to be stored in liquid nitrogen, so to have a congruous amount of starting material for following experiments. The maintenance in culture of C3H cells is usually effectuated in BME culture medium added with a 10% FBS. The cells must be strictly kept at sub-confluence through passages with trypsin.

The serum used for a transformation assay must be tested for its plating efficiency (PE) as the ratio of colonies formed during a cytotoxicity assay and the number of seeded cells. The minimum value to accept a batch of serum is a PE of 20%. Unused serum must be preserved at very low temperatures (< -60 °C) and can be used for at least 5 years. In fact, the serum has a fundamental role for its concentration and composition. It is therefore very important that, for all the duration of a transformation assay, the same type and the same batch of serum is used, so as to guarantee homogeneity of results. Various studies indicate that the FBS is the only suitable kind of serum for this type of tests among all the types normally in commerce. As possible for one laboratory as this might be, such dependence is a problem in terms of international standardization and inter-laboratory reproducibility. The study of a

synthetic alternative of FBS to use in transformation assays might be a solution for this critical point.

The effect of cell plating density has an important role on the test efficiency. In fact, its impact on transformation processes might be explained in relation to the number of divisions toward which a cell may go before undergoing contact inhibition. The greater the number of divisions, the easier the cells are subject to the transforming effect of the chemical agents. This hypothesis seems verified by the results of Mondal's experiment, which demonstrated that from a CTA done on a single cell results a 100% of transformation and a null toxicity. **(Reznikoff et Al., 1973)**.

A 24 to 48 hours of treatment with a known carcinogen at a standard concentration (typically, PAHs like methylcholantrene, benzo[a]pyrene, N-Methyl-N'-Nitro-N-Nitrosoguanidine) is used as positive control. The control treatment must produce an average of 2 *foci* per petri. For a transformation test based on C3H cells, is best to seed 200-500 cells for petri of 60 mm or an equivalent number on petri of 100mm, respecting the plating density as cells/ mm^2.

Where experimental protocol do not very much differ between BALB/c and C3H, differences in the valuation of results are much sharper. In fact, where in the test based on BALB/c, the obtained colonies at the end of the protocol are divided only between transformed and not transformed, in the transformation assay based on C3H, they are divided in

transformed and not transformed, and the first further distinguished in three classes, depending on growth patterns and morphology of cells and colonies. It is then separately counted the frequency of appearance of Type II and Type III *foci*. A sum of the criteria that have to be met to accept the results of the assay are a PE of at least 20% and a number of transformed *foci* in the control plates inferior to 0.15 *foci*/plate. The low number of *foci* for plate in C3H is a measure of how low the frequency of spontaneous transformation is in this model. Historically, this value seems to reach values even inferior to 0.07 *foci*/plate, where in the BALB/c there is an even higher threshold value of 0.5 *foci*/plate, nearly tenfold higher than the C3H value.

An important modification of this test is the so called two-stage CTA, which enables to evaluate the activity of a chemical to act as a promoter or protective towards the chemical transformation. In short, cells are treated with a known carcinogen (MCA usually) and then left for recovery. After a variable period to be assessed for every single agent in study in order to find the optimal value, but usually comprehended among 1 day and 1 week, the promoter or protective agent in study is added to the culture and confronted with a known promoter used in non toxic concentrations, (usually TPA). At the state of the art, the two-stage CTA is the only *in vitro* assay which allows studying the role of chemicals as promoters and protective for the transformation.

1.8 Evaluation of Results

One of the most critical points in *in vitro* cell transformation assays lies in the interpretation of its results. The evaluation of morphological alterations in cell colonies (*foci*) is done by a trained expert through the observation at the optical microscope. This phase requires extensive training and experience, and relies on the judgment done by the operator of the evaluation criteria, so much that difficulties have been encountered in finding objective canons capable to pacify all the experts, even more on the borderline cases **(Combes et Al., 1999)**. In a more specific detail, colonies obtained in a transformation assay are classified depending on their morphology in three classes:

- Type I, characterized by highly packed cells with normal morphology and forming colonies with round sides mostly of average little dimensions **(Figure 1.1, panel B)**.
- Type II, characterized by stacking and growth on multilayer, with cells moderately polarized, and formation of colonies of variable dimension and uneven edge **(Figure 1.1, panel C)**.
- Type III, characterized by highly polarized cells, extensive criss-crossing and pronounced multilayering, with colonies usually big and deeply stained, with protruding edges that invade the surrounding normal monolayer and formations similar to whirlwinds on the sides. *Foci* of this type may be recognizable even before the Giemsa staining at the optical microscope **(Figure 1.1, panel D)**.

Foci belonging to Type II and Type III may present themselves as colonies uniformly colored or with a “threaded” aspect. Moreover, there are numerous cases in the literature of reported *foci* with non-canonical morphology, on which classification the scientific debate is still very hot on. *Foci* that present characteristics intermediate among two classes of damage are usually defined as intermediate, and are assigned to the inferior class of damage. *Foci* that present different areas with morphological characteristics belonging to two different classes of damage are defined mixed or pleiomorphic, and are assigned to the superior class of damage. However, the “decisional line” is very thin and has to be traced on a case by case basis relying on the operator judgment. This process is theoretically open to errors of overestimation or underestimation by even the most well trained expert **(Figure 1.1, panels E and F respectively)**.

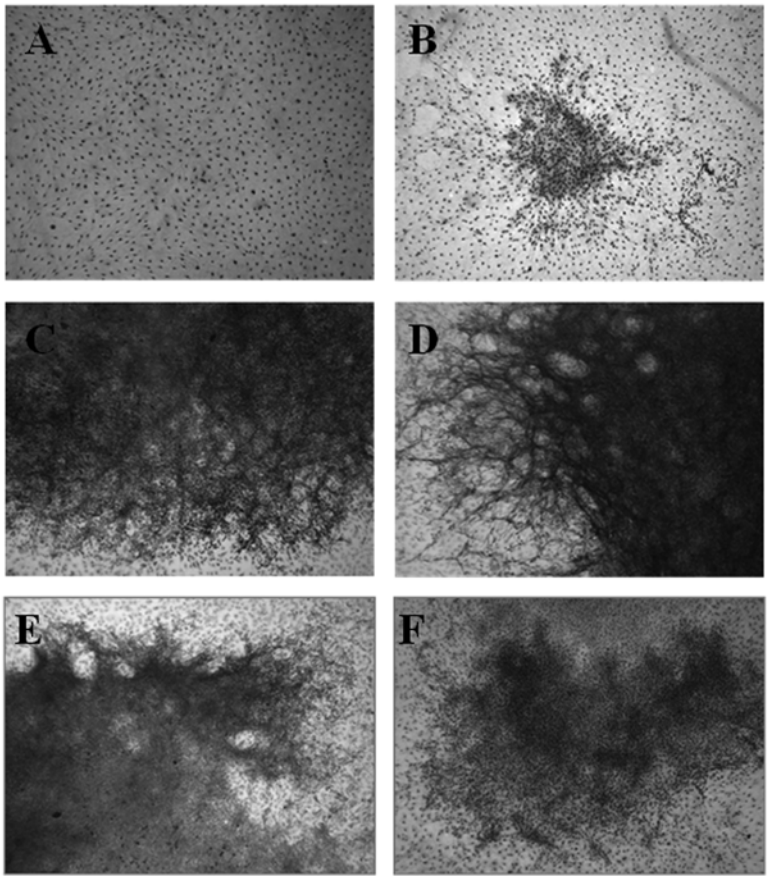


Figure 1.2. Sample images of CTA morphological classes on C3H cells.
Figure shows a panel of typical C3H *foci* images. All images were taken at 25x magnification in optical microscopy **(Procaccianti 2011).**
Panel A. Normal, contact inhibited cells; **Panel B.** Type I *focus* characterized by normal, non-piling, tightly packed cells; Type II *focus* **(Panel C)** characterized by massive piling into virtually opaque multilayers. Cells are transformed and mildly polar cells. Type III *focus* **(Panel D)** showing highly polar, multilayered criss-crossed arrays of cells. Cells are transformed and show fibroblastic aspect. Mixed type *focus* **(Panel E)** presenting characteristics belonging to two different classes. Intermediate *focus* **(Panel F)** showing features intermediate between two different classes.

Moreover, while *foci* belonging to Type I are not reported by many authors, others assume that they may represent an initial step of carcinogenesis, maybe characterized by cells still inhibited in the

progression by some factors, like high cell density or low concentrations of serum. Some authors also suggest the possibility that Type II may, in a process similar to tumoral progression *in vivo*, degenerate with time in a Type III *focus*.

From the point of view of biological behavior, cells isolated form Type I *foci* are not transformed, as they do not form colonies when seeded in soft agar and don't produce tumors when injected under the skin of immunocompromised animals. On the contrary, Type II and Type III *foci* are composed by transformed cells, producing xenografts respectively in the 50% and the 85% of cases, after a period of incubation that varies from 2 weeks to 6 months. In the 100% of cases, the tumors produced are sarcomas. Cells obtained from negative controls do not induce formation of tumors.

The performances of an *in vitro* transformation assay are evaluated in terms of positive and negative depending on a 2 x 2 table, also providing the levels of sensibility, specifity and positive as well as negative predictivity values of the test as well **(OECD, 2007)**.

1.9 Gap Junction Intercellular Communication (GJIC)

Intercellular communication is a mechanism that enables to two contiguous cells to freely exchange small molecules weighting less than 1500 Da through diffusion mechanisms. Connexin monomeric proteins oligomerize into hexamers commonly referred to as connexons, or

hemichannels. Assembly in both homopolymers and heteropolymers is possible. Two adjacent connexons join to form a complete gap junction channel. Individual channels are arranged in hexagonal arrays, referred to as gap junction plaques. Connexins are assembled on the endoplasmic reticulum, oligomerize in the *trans*-Golgi network and are delivered to cell surface by a system of vesicles. Connexons then form complete junction plaques. Retrieval of plaques from the membrane is mediated by the formation of double membraned vesicles called “annular junctions”. Degradation is mostly lysosomial **(Figure 1.2)**. The connexin family is ubiquitary espressed and is composed of 21 members in human, the most diffused being the connexin 26 (Cx26), the connexin 32 (Cx32) and above all the connexin 43 (Cx43), which is expressed in more than 35 different cell types. Connexin 43 has been demonstrated to be binding partner of a variety of intracellular molecules involved in survival and proliferation pathways, such as PKA, PKC, mitogen-activated protein kinase (MAPK), cdc2 kinase and caveolin-1 **(Laird, 2006; Naus et Al., 2010)**.

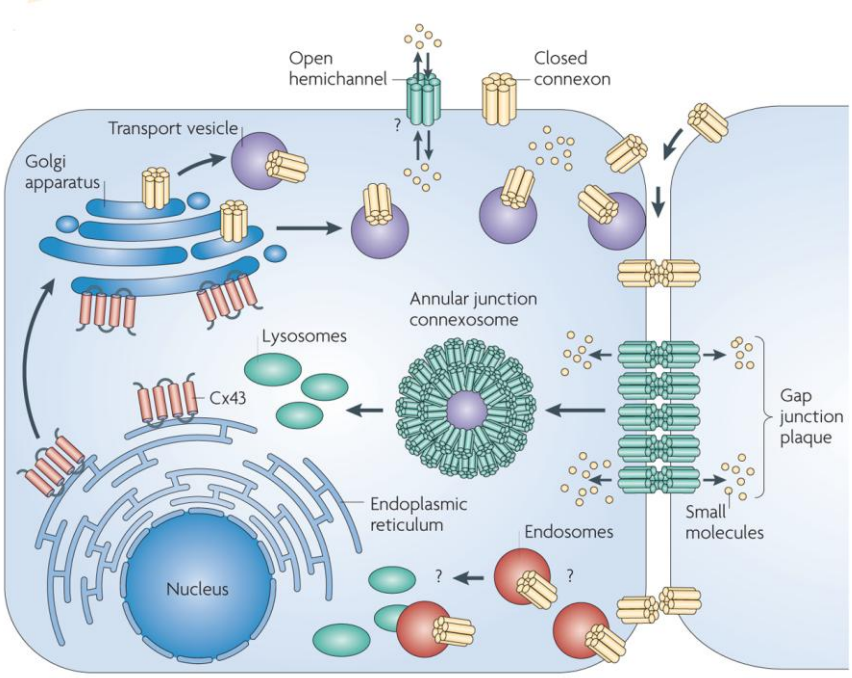


Figure 1.3 Life Cycle of connexins.
Cx43 is synthesized on the endoplasmic reticulum and transferred to Golgi apparatus, where it is assembled into homo and heteropolymers. Transport vesicles carry the connexon on the membrane where it forms junctional plaques. Plaques are sequestered from the membrane via annular junctions, and are delivered to lysosomes for degradation. Other minor pathways of degradation may coexist (e.g. monoubiquitinilation) **(Naus et Al., 2010)**.

The connexin family seem to have a very important role in regulating cells proliferation and differentiation. In fact, a number of alterations in both cellular communications and expression profiles of connexins were found in many pathologies and solid human tumors, and chemical carcinogens with channel inhibitory functions are known. Knockout mice for connexin genes show also a higher degree of induction of chemical carcinogenesis when treated with a carcinogen.

Intercellular communication allows cells to exchange protective factors like glutathione, or catch paracrine stimuli through hemichannels. Moreover, the existence of a communication independent tumor suppressant ability of gap junction has also been demonstrated by the use of mutants. In fact, recent experimental evidences seem to indicate that Cx43 is implied in the regulation of cells cycle not only as a canal proteins, but also thanks to the interaction with intracellular proteins with an anti-tumoral action, such as caveolin-1, and regulating the expression of cyclins and cyclin dependant kinases. Recent studies associate Cx43 and N-cadherin, linking the processes of cellular adhesion and migration with connexins on a GJIC-independent manner. Cx43 has also been recently demonstrated as having a role in tumoral progression and in the invasiveness of tumors in the advanced state. Controversial studies indicate a possible role for Cx43 in facilitating the formation of tumor metastasis. Recent works have also showed an association between the inhibition of GJIC and transformation, but not between the inhibition of GJIC and genotoxicity **(Langlois et Al., 2008; Naus et Al., 2010)**.
Due to all these characteristics and the emerging importance of connexins in neoplastic transformation in the recent literature, the evaluation of the levels of intercellular communication, as well as the localization and levels of a ubiquitary expressed connexin as Cx43, might result in markers suitable to study the carcinogenetic effect of

compounds, even for those not exhibiting direct genotoxic potential and thus not recognizable through other conventional methods.

The GJIC assay has been developed initially by **Fitzgerarld et Al. (1982)**, and is currently used mostly by academic bodies for research purposes. **El-Fouly et Al. (1987)** developed the *scrape-loading* technique to assess changes in the levels of intercellular communication. Briefly, cells are carved upon reaching confluence levels in the presence of a solution containing a fluorescent dye with low molecular weight. Due to capillarity, the dye migrates proportionally to the level of intercellular communication. So far, there still is no protocol formally validated at regulatory level for assays studying study intercellular communication. Nevertheless, protocols have been proposed and are under evaluation **(Maurici, 2005)**.

1.10 Notions of Image Analysis

The methods so far described for the evaluation of *foci* are based on the microscope observation by a trained operator, who assigns a score to each colony on the basis of canonized morphological criteria described in literature **(Combes et Al., 1999; OECD, 2007)**. However, this evaluation process is entirely based on qualitative features, whose evaluation is inherently subjective rather than objective. This subjectivity in the evaluation of results of a cell transformation assay renders *bona-fide* errors possible, due to either over or under evaluation of morphological features, resulting in a possibly wrong estimation of the

carcinogenic potential of the analyzed substance. It has been reported that in many cases is possible that classification experts may even disagree over *foci* class attribution **(Combes et Al., 1999).** This is even more problematic in colonies with borderline morphology features, like mixed of intermediate types. To assess the problems of subjectivity of classification, a model based on measurable features of *foci* like the diameter of colonies has been proposed **(OECD, 2007; Keshava, 2000)**. In such context, the development of a methodology based on quantitative data to support the operator in the decision making process might play a lead role for minimizing as much as possible the impact of the subjective evaluation of the morphology.

The use of image analysis techniques is well documented in the scientific literature. These methods are based on the identification of a series of structural patterns within the analyzed data, and exploiting them for dividing images into clusters, by building affiliation groups characterized by high similarities within groups (intra-class variability) and by high dissimilarities between different groups (inter-classes variability) **(Jain, 1989)**. Examples of the use of these techniques applications can be found in the recognition of latin characters, in particles scattering studies **(Crosta et Al., 2008)**, or in the classification of biological images in fluorescence **(Urani et Al., 2009)** or in visible light **(Ridder et Al., 1997)**. A digital image is characterized by a discrete structure, defined by a raster where each element is called *pixel*. Each digital image is therefore

similar in structure to a numeric matrix with rows and columns as the number of *pixel* on its two dimensions, while the value of each individual element represents the corresponding image pixel intensity **(Figure 1.3)**. Thus, an image with the dimensions 1024 x 768 is in fact represented by a matrix with 1024 rows and 768 columns. Digital images can roughly be devided in binary (black and white images), gray scales or coloured images.

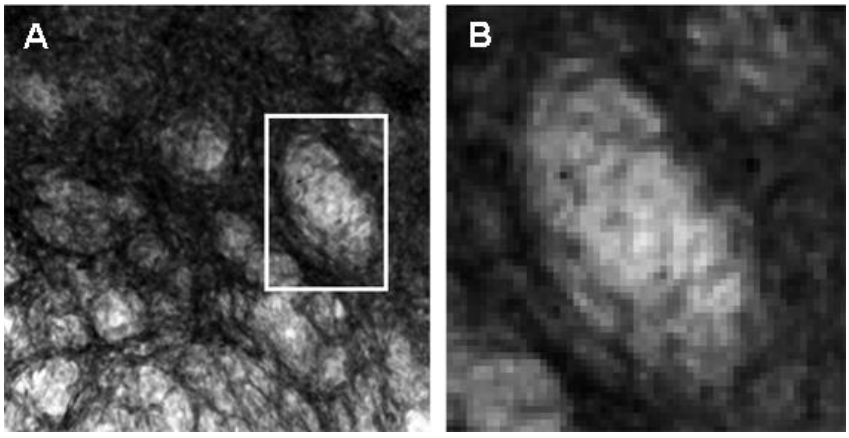


Figure 1.4. Example of a Digital Image Structure
On the left (panel A), a detail of a *focus* image. On the right side (panel B), a detail of the image was enlarged until the discrete structure of single pixels was visible. Images were edited for exemplificative purposes**.**

The gray scale 8 bit images are represented by a single matrix with values between 0 and 255 determining the exact gray's intensity, while for the black and white images, the matrix presents only two extreme values (0 and 255). On the other hand, coloured images are represented by more matrices, one for each colour channel. For example, an image in the standard format RGB (red, green, blue) is composed by three

matrices – one for red, one for green and one for blue. The value of each matrix element represents the image colour component in that exact point. The final color is then perceived as the resultant of the three channels (e.g. in the RGB format, 255-0-255 is magenta). The study of the pixel's distribution can be therefore used in order to transform the visual characteristic of an image in numeric data, thus transporting the qualitative features of the image subject (in the case of *foci*, morphological features, as described in paragraph 1.8) to quantitative information (vectors and numbers). Such quantitative data can help the expert in the *foci* evaluation process **(Ridder et Al., 1997; Crosta et Al., 2006; Urani et Al., 2009; Urani et Al., 2010)**.

1.11 Automatic Classification of Images

The notion of *algorithm* can be defined as a finite number of instructions structured in a suitable way to obtain a specific result. An algorithm that divides data (e.g. numerical data obtained from digital images) in two or more groups (or clusters) based on strucutral *patterns* is a classification algorithm, or *classifier*.

In the process of classification is possible to distinguish two steps subtending to the use of such program: a *training step* – where images with known and certain group membership are used in order to train the classifier; and a *classification step* – where new data of unknown membership are scored on the basis of the results of the training step.

The comparison between the obtained scores of classification and the reference standard allows the evaluation of the performances of a clustering algorithm. In the case of *foci* images scoring, due to the lack of an actual objective gold standard, the set reference is the judgment of a well trained expert. The outcome of the training of a classifier is strongly affected by the samples size and the quality of the information used. A low quality training dataset is usually defined as “noisy”.

Two approaches of automatic classification can be distinguished, defined as *supervised* and *unsupervised*. In the *supervised* approach, the data used in the training step have to be previously scored by the expert. Therefore, the labels of assignment and the number of classes are part of the data used for the training itself. The algorithm then exploits the given class attribution as a scaffold for searching patterns. On the other hand, in the *unsupervised* approach, the analyzed data do not contain information regarding the group membership. In some cases, even the number of clusters can be inferred during the training step by the algorithm itself. Whatever the approach, the goal is to obtain a trained classifier to be used for the scoring new data on the basis of the obtained patterns. It is noteworthy that the classification information used during the evaluation of performances don’t necessarily belong to the data submitted to the classifier itself. Thus, the evaluation of the performances of an *unsupervised* classifier is possible without biasing the *unsupervisedness* of the process itself. Finally, a classifier can be

structured either as a single unit or as a combination of different algorithms, even of different nature, either supervised or unsupervised. Actual data to be analyzed for an image analysis process are be extracted from the images by means of different approaches, spanning from a “simple” extraction of the statistical information regarding the image’s *pixel* distribution, to more complex analysis techniques derived from the signal theory or the Fourier transform analysis and its diverted functions (e.g. Fast Fourier Transform, or FFT). The extrapolated data are then structured in a matrix with a variable degree of dimensionality, and represent a quantification of the image features. Due to their usual complexity, the obtained matrices can be simplified by the application of mathematical and statistical methods. Such techniques extract most of the hidden information included in the dataset, resulting in a final simpler matrix. A typical example of a statistical technique for multivariate analysis that is widely used in this context is the principal components analysis, or PCA. Through the use of this technique, the extrapolation eigenvectors and eigenvalues from the original matrix allows to reduce the dimensionality of the dataset, while maintaining most of the variability of the original samples (percentage of variability explained on the first *n* components).

1.12 Cadmium: Chemical Characteristics and Biological Effects

Cadmium presents a carcinogenic potential recognized by IARC decades ago **(IARC, 1993)**, even if potential for direct damage to the DNA have never been registered. Cadmium is found at low concentrations in the Earth's crust, mainly in zinc-containing mineral deposits. It has been produced and used in a variety of applications as cadmium oxide (used in batteries), cadmium sulfide (used as a pigment) or cadmium stearate (used as a plastics stabilizer). Occupational exposure to cadmium and cadmium compounds occurs mainly in the form of airborne dust and fume (e.g. cadmium production and refining, nickel-cadmium battery manufacture). Although levels vary widely among the different industries, occupational exposures generally have decreased in the last two decades **(IARC, 1993)**. Absorption by man occurs mostly through inhalation (cigarette smoke or exhalations on working place) or by ingestion (contaminated food and water, especially contaminated cerals). Burning of fossil fuels is also known for releasing cadmium in the atmosphere **(Hartwig et Al., 2010)**.

It is present in various forms, both water-soluble and not. Cadmium kinetics are characterized by a tendency to form deposits in target organs, especially kidney and liver, probably accounting for part of its organ-specific toxicity **(Urani et Al., 2006)**.

Cadmium seem to operate mostly through epigenetic mechanisms, inducing a plethora of different effects, principally:

- Inhibition of DNA damage systems repair
- Increase of ROS induced damage
- Alterations of cell cycle control mechanisms

Intertwining of all these involved and closely linked pathways might give an explanation for the complex cancerous potential of this metal.

(Hartwig, 2010)

Cadmium is present in the periodic table as an element of transition in the zinc group (Group 12, or IIB, depending on the naming). It is noteworthy that elements in the same group have the same configuration in external orbitals, thus exhibiting, in most of the cases, the capacity of forming similar molecular bonds. Zinc has a very important role in Zinc Finger domains. This proteic secondary structure have the function of DNA bonding site are is widely present among proteins regulating cell cycle, most notably in the p53 *caretaker* protein structure, and genetic transcription. One of the possible mechanisms of action of cadmium on the cell cycle control might in fact be dependent from the ability of this metal to compete with zinc bonding in zinc fingers domains, with consequent alteration in the functionality of involved proteins. Even if effects of direct mutagenesis have never been found in tests performed on bacteria and mammals, a high co-mutagen potential has been noticed in co-administration regime in presence of alkylating agents or UVC radiations. Cadmium shows a strong clastogenic activity, as shown in tests on mammal cells, with the formation of micronuclei and partial

deletions, confirmed by *in vivo* experiments on rodents. Studies on cellular extracts with cadmium salts confirmed its incapacity to induce direct damage to DNA. Results seem to go in the direction of reactive oxygen species (ROS) generation, interaction with proteins and damage repair mechanisms, as well as epigenetic regulation mechanisms, like DNA methylation patterns.

Increasing in ROS levels induced by cadmium both *in vivo* and *in vitro* are widely documented. Though cadmium seems not to directly participate to the formation of oxygen radical species, the increase seems to be due to a inhibitory activity on enzymes involved in the ROS metabolism, such as catalase, superoxid dismutase and glutathione reductase. In fact, chromosomes damage induced by ROS are suppressed by co-administering of antioxidants. However, oxidative damage is not enough to fully explain the increased transformation frequency induces by cadmium, which is probably also related to an inhibitory effect on DNA damage repair systems. Moreover, altered ROS concentrations seems to be involved in tumoral proliferation steps through the activation of transcription factors sensible to cell redox state. On the whole, oxidative stress might intervene in both the transformation initiation steps, and in the promotion of tumoral growth.

Cadmium interferes with nearly all DNA repairing systems:

- blocks the recognition and incision of the DNA strand, as well as the assemblage and disassembling of the multi-proteic complexes involved in the repairing system by nucleotide excision (NER).
- low cadmium concentrations inhibit the activity of oxidized nucleotides removal enzymes, both in vitro and in vivo. Altered levels of oxidative damage induced by with low cadmium concentrations might, then, be linked to the inhibition of repair systems rather than to a direct induction of ROS production.
- cadmium inhibits MMR repair system, through an inhibition of the ATPase activity of the involved enzymes, reducing their affinity for DNA.

Cadmium alters the expression of numerous genes involved in stress response, among which genes involved in the metallothionein biosynthetic pathway (immediate response to metal presence), genes of the HSP family (generic stress response) and of GSH neo-synthesis (oxidative stress response).

Genes of primary response to cadmium treatment include also various proto-oncogenes like *c-fos*, *c-jun* and their joining product AP-1, *c-myc* and the transcriptional factor NF*k*B. Cadmium treatment induces a block of the SP1 factor and shows a suppression effect towards proliferation inhibition factors, like in case of p53. **(Levine et Al., 2011**; **Méplan et Al., 1999**).

1.13 Cadmium: Molecular Mechanisms

Independently from the way of administration of cadmium (soluble or not), experimental data show that the toxic effects are linked to the Cd^{2+} ion. Even if most of mechanistic studies have been done on water-soluble compounds, recent studies show a potential block of removal of damage induced by B[a]P with both the soluble and the non soluble forms. Moreover, conformational changes induced by both forms on the gatekeeper protein p53 seems to be similar. Levels of inhibition of repair seem to directly relate with cadmium levels in the nucleus, thus only not directly with the bio-availability of the compound itself **(Hartwig, 2010)**.

Cadmium water-soluble compounds enter in the ionized form through the L-type calcium channel (LTCC) and by proton-metal co-transporters (e.g., DMT1, DCT1, and Nramp2) **(Martelli et Al., 2006)**.

Non soluble compounds mainly enter through phagocytosis and degradation in the lysosomes, where low pH allows the degradation of non soluble particles and the liberation of ions.

One of the mechanisms hypothesized for the multitude of effects that cadmium causes at not cytotoxic concentration (block of DNA repair systems, inhibition of cell cycle controls) is based on competition between cadmium and other fundamental elements for bonding sites, among which Zn^{2+}. Cadmium can substitute zinc due to similar dimension and charge, in so modifying the behavior of various enzymes. In fact, cadmium shows a strong affinity for the thiol groups, which are

present in zinc finger motifs of protein-DNA interaction and protein-protein interaction. A characteristic example of the function of those domains is found in p53 protein, where zinc finger regions are found in DNA bonding domains. However, cadmium does not interfere with repairing systems only through p53, but also interacting with other proteins with zinc finger domains. As an example, cadmium inhibits the action of hOGG1, a glycosidases involved in base excision repair systems (BER). Even if hOGG1 does not have zinc finger domains itself, it is under the transcriptional control of the SP1 promoter, which, on the other hand, presents such domains. In fact, bonding of this transcriptional factor to DNA is inhibited by cadmium, presumably for the displacement of zinc.

Cadmiums seem to act at transcription levels, both by direct interaction with transcription factors and by induction of second messengers. In fact, cadmium ion bonds with the MTF1 transcription factor seem to be directly responsible for the increase in the MT synthesis. This mechanism is common between responses to the presence of metals, like in the case of the iron IRE-BP IRE chain of interaction for the induction of synthesis of ferritin. Other effects on transcription seem to be due to the increase of second messengers such as ROS, resulting in the activation of factors of survival and growth sensible to the ROS levels **(Valko et Al., 2006)**, or Ca^{2+}, through the activation of protein kinase C and its phosphorylation products.

More epigenetic effects of cadmium comprehend the inhibition of the DNA-(cytosine 5)-methyltransferase enzyme, thus reducing DNA methylation and then inducing an increase of levels of expression of some proto-oncogenes, as well as the inhibition of cadherin-mediated cellular adhesion signals and the increase of levels of the anti-apoptotic factor Bcl2.

Tolerance mechanisms have a fundamental role in cadmium toxicity. By increasing MT levels and ROS metabolism genes transcription, a situation of protection toward acute toxicity arises. This stress response actually favors the accumulation in the tissue with consequent long term effects of inhibition of DNA damage repair systems and of pro-apoptotic mechanisms **(Urani et Al., 2007; Hartwig, 2010)**.

The plethora of cadmium effects, many of which of only recent discovery have brought the regulation bodies to reduce the recommended limit form 7 µg/kg to 2.5 µg/kg **(EFSA, 2009)**. This threshold level is mostly based on the concentration inducing toxic effects on the kidney. However, more than the nephrotoxic effect, cadmium presents strong effects of demineralization in the bones, an increase in the insurgence of cardiovascular problems, as well as an increase in the frequency of lung, kidneys and prostatic tumors. Recent publications associate cadmium to breast tumors and endometriosis, even if such data are not yet definitive. Given that tumors that develop in the prostate, uterus and breast show hormone dependence characteristics, many groups are studying

possible interactions between cadmium and pathways dependent from steroid hormones, with, up to now, contradictory results.

Chapter 2: *in vitro* materials and methods

2.1 Cell Cultures

C3H10T1/2 Cl8 (hereon, C3H) cells were chosen for their high sensitivity to carcinogenic compounds and their low spontaneous transformation rates, as well as for the high level of information achievable by the four output classes of the CTA based on that cell line. HepG2 cell lines were chosen for their hepatic heritage, in order to study the effects of cadmium on liver cells.

Cells were routinely cultured in Basal Medium Eagle (BME, Sigma Chemical Co., St. Louis, MO). Medium was enriched with 10% heat-inactivated fetal bovine serum (FBS, Euroclone, Pero, Italy), 1% glutamine, 0.5% HEPES 2M and 25 µg/mL gentamicin (all from Sigma). Serum inactivation was performed at 56 °C for 40 minutes. Cells were kept in a humidified incubator supplied with a constant flow of 5% CO_2 in sterile air at a fixed temperature of 37 °C. Culture flasks and 100-mm-diameter dishes were all purchased from Corning (catalogue no. 430167, Acton, MA). Medium was routinely changed three times a week. Cells were daily observed to check health and confluence levels. Experiments were performed with C3H cells between 6^{th} and 22^{nd} passage. HepG2 cells were cultured in Opti-MEM medium (Gibco, Life Technologies BRL, USA), enriched with 10% heat-inactivated fetal bovine serum, L-glutamine and 1% antibiotics. Cells were kept in a sterile incubator at constant 37° and 5% CO_2 **(Urani et Al., 2005)**.

Cadmium is an ascertained carcinogen, though its mechanisms are still elusive. The stock solution of $CdCl_2$ (1mM) was prepared in ultra-pure water (0.22 µm filtered Milli-Q water, Millipore, Vimodrone, Italy) and stored at 4◦C. Benzo[a]pyrene (B[a]P) was chosen as a positive control for its well known carcinogenic and genotoxic potential. Stock solutions dissolved in DMSO (6.25 mg/mL, corresponding to 25 mM and 0.625 mg/mL, corresponding to 2.5 mM) and stored in the dark. Stock solution of 12-O-tetradecanoylphorbol-13-acetate (TPA) (Sigma) was prepared in DMSO to a final concentration of 1 µg/µl. TPA was chosen as well known promoter agent and as reference standard in 2-step cell transformation assays **(OECD, 2007)**.

2.2 Cell Viability

Cell survival was evaluated with a standard MTT assay. Briefly, this method consists in a colorimetric assay based on the quantification of the activity of the succinate dehydrogenase, an enzyme localized at mitochondrial level and active only in healthy cells. The catalytic activity of this enzyme is to reduce the water-soluble 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to formazane, that precipitates in the form of blue/violet crystals. Crystals are then dissolved in DMSO and absorbance reading performed at 560-570 nm. MTT stock solution was prepared in the dark at a concentration of 5 mg/ml and sterile filtered

with 0.22 µm diameter filters. Solution was stocked at 4 °C in the dark for maximum 4 days.

C3H cells were seeded on petri dishes, 80'000 cells each (Day 0). 24 hours after seeding (Day 1), medium was removed and petri added with medium enriched with 1 µM $CdCl_2$. Positive control was medium enriched with 1 µg/ml B[a]P dissolved in DMSO, while negative control was medium only. Every 24 hours (on Days 2, 3 and 4), an equal number of dishes for each treatment were observed under the optical microscope and processed for MTT assay. 200 µl/petri of MTT stock solution were added in the culture medium at the end of the treatment time**,** to a final concentration of 0.5 µg/ml. Dishes were incubated in the dark at 37 °C for 4 hours. Culture medium was then gently removed through aspiration with a fine-needle (22 gauge diameter) syringe and the crystals dissolved in 2 ml DMSO/petri. Absorbance was read at 560 nm. An equal volume of pure DMSO was used as blank **(Urani et Al., 2006).**

2.3 HSP70 and p53 Proteins Extraction Protocol

Cells were seeded on day 0 at a concentration of 2.5 x 10^6 cells/flask. 2 flasks were used for each treatment, to a total number of 5 x 10^6 cells/treatment. On day 1, medium was removed and replaced with fresh medium added with scalar doses of CdCl2 (0.1 µM, 0.5 µM, 1 µM, 2 µM). Positive control was B[a]P dissolved in DMSO at 0.1 µg/ml and 1 µg/ml.

Negative controls were medium only and medium with vehicle (16 µl DMSO in 10 ml medium). On day 3, culture medium was removed and cells washed with 37 °C PBS (137 mM NaCl, 2.68 mM KCl, 2.09 mM Na_2HPO_4, 1.469 mM KH_2PO_4 in MilliQ water) and harvested with tripsin. Cell suspension underwent two cycles of centrifugation (1'200 rpm equivalent to 170g at 4 °C for 5 minutes). Between cycles, supernatant was discarded and cells rinsed in cold PBS. Pellet lysis was performed in freshly prepared Sample Buffer (500 µl Tris HCl 1M pH 6.8, 1 ml SDS 10%, 500 µl glycerol, 250 µl 2-mercapto-ethanol, 250 µl phenylmethylsulfonyl fluoride (PMSF) 20 mM, 2.5 ml MilliQ water for 5 ml total of buffer). Protein extract was boiled at 100 °C for 5 minutes and homogenized with a fine-needle syringe (22 gauge diameter) **(Urani et Al., 2006).** Final volume was divided in two aliquots and stored at -80 °C. Total protein concentration was quantified by a standard Lowry assay.

2.4 Western Blotting

Total proteins extracts were diluted in Sample Buffer to reach a final concentration of 5 µg/µl. Bromophenol blue powder was added before loading and samples boiled for 5 minutes. 30 µg of total proteic extracts were loaded for each sample. Electroforetic separation was performed in 10% bis-tris precast gel for p53 and 7% bis-tris for HSP70. Both separation gels and membrane transfer were performed following producers instructions (INVITROGEN©).

Aspecific binding sites were blocked by treatment of the membrane with a blocking solution (PBS enriched with 3% serum bovine albumin (BSA) and 0.05% Tween-20) for 2 hours. Primary antibody hybridization was performed for 16 hours at 4 °C. Secondary antibody hybridization was performed at room temperature for 3 hours. Primary monoclonal mouse antibodies against p53 and against HSP70 were bought from Stressgen and used at 5 µg/ml in PBS enriched with 5% BSA and 0.1% NaN_3. Secondary anti-mouse antibodies phosphatase conjugate were diluted 1:10'000 in blocking solution (all bought from Sigma). Protein binding was visualized by the colorimetric substrate BCIP/NBT. Protein expression was measured with Gel Doc image system (Bio-Rad Laboratories, Milano, Italy). Quantitative analysis was performed with Quantity One dedicated software **(Urani et Al., 2006)**.

2.5 Indirect Immunofluorescence

Cells were seeded on 20 mm side sterilized glass slides in a 35 mm petri dish (Corning) at a concentration of 80'000 cell/petri**.** Forty-eight hours after seeding, medium was removed and cells were treated with fresh medium added with 1 µM $CdCl_2$. Positive control used was 1 µg/ml B[a]P. Negative control received medium only. After 24 hours, treatments were removed and cells washed with PBS. Cells were then fixed in 2 ml ice cold methanol for 15 minutes. Blocking solution used was PBS enriched with 1% BSA and 0.05% v/v Tween-20 for 1 hour. Primary antibody was

diluted 1:100 in blocking solution and 50 µl applied to each sample. Samples were incubated for 1 hour at 37 °C in a humidified chamber. The secondary fluorescent anti-mouse antibody Alexa Fluor 594 diluted 1:100 in blocking solution was used always in the dark. Hybridization was performed for 45 minutes in the same conditions as the primary antibody. Slides were finally washed twice with PBS, rinsed twice with MilliQ water and mounted on histological glass slides. Mounted slides were stored at 4 °C in the dark and observation under fluorescence microscope was performed after 24 hours.

All image acquisitions, both for fluorescence and normal light observations, were performed with a Zeiss Axioplan™ microscope, equipped with either 25x or 400x magnification (Carl Zeiss, Arese, Italy) and a digital camera (CoolSnap-ProColors Media Cybernetics, Bethesda, MA, USA).

2.6 Cell Transformation Assay

2.6.1 Cytotoxicity Assay

Cytotoxicity assays are performed in the same experimental conditions of cell transformation assays and are aimed to evaluate the plating efficiency of a batch of serum and the toxic effect of the treatments on the initial formation of C3H cell colonies. Plating efficiency (PE) is measured as the ration between the number of colonies formed in the control dishes and the total number of seeded cells.

$$PE = \frac{\text{Number of colonies counted}}{\text{Number of cells seeded}}$$

Survival (S) was calculated as the ratio between the number of colonies in treated dishes and the number of colonies in control dishes. This value is a measure of the toxic effects of the treatments on the cell growth **(OECD, 2007)**.

$$S = \frac{\text{Number of colonies counted in treatment groups}}{\text{Number of colonies counted in control groups}}$$

C3H cells were seeded on 100 mm petri dishes at a concentration of 200 cells/dish in 8 ml of complete culture medium. Twenty-four hours after seeding, medium was removed and 10 ml fresh medium enriched with $CdCl_2$ at doses of 1 µM and 2 µM (see paragraphs 4.3.1 and 4.3.2 for single experiments details) for 24 or 48 hours was added. Positive control was medium added with B[a]P 1 µg/ml dissolved in DMSO for 24 or 48 hours, while negative control dishes received medium only. At the end of the treatments, medium was aspired and cells washed with 37 °C sterile PBS. A volume of 15 ml of fresh medium was added for 10 days (*recovery* period), with a single change after 7 days. After 10 days, medium was removed, dishes washed with PBS and fixed with 15 ml of methanol for 10 minutes at room temperature. Dishes were then briefly rinsed with MilliQ water and Geimsa stained following producer instructions (Giemsa stain modified, Sigma-Aldrich, Milano, Italy) **(OECD, 2007)**.

Colonies formed were circled on the dish surface with a permanent marker and counted for cell numerosity at the optical microscope with a 25x magnification. Colonies counting less than 50 cells each were excluded from the final treatment group value.

Protocol was slightly modified for the evaluation of PE and S in the 2-stage cell transformation assays. Four days after the beginning of the *recovery* period, medium was removed. All groups except control received medium added with 0.1 µg/ml 12-O-tetradecanoylphorbol-13-acetate (TPA) dissolved in DMSO, diluting the stock solution 1:10'000 in the final volume. The final DMSO concentration was inferior than the 0.1% v/v threshold concentration as set by OECD **(OECD, 2007)**. Control group received medium only. Fix and staining were carried out as previously described in this paragraph.

2.6.2 Cell Transformation Assay (CTA)

C3H cells were seeded at 800 cell/dish per each 100 mm dish. Dishes made both of plastic and sterilized glass were used. Treatments were performed 24 hours after seeding: medium was removed and fresh medium enriched with $CdCl_2$ at doses ranging from 1 µM and 2 µM and times between 24 and 48 hours (see paragraph 4.3 for single experiments details) was added. Positive control was medium added with B[a]P 1 µg/ml dissolved in DMSO for 24 or 48 hours, while negative control was medium only. The same experimental conditions and the

same batch of serum of corresponding cytotoxicity assays were maintained. At the end of the treatments, the culture medium was removed and dishes were rinsed twice with 37 °C sterile PBS. Fresh complete medium was added in a volume of 15 ml for the whole rest of the CTA experiment (*recovery* period): cells were cultured for 6 weeks with a weekly change of medium. Upon reaching confluence (around the 3rd week), culture medium was changed from the normal high serum (10% FBS) to a low serum regime (5% FBS) and administered for 3 more weeks with weekly changes. Starting from the 3rd week, dishes were weekly observed at optical microscope with 25x magnification to check cells health status and *foci* formation. At the end of the 6th week, dishes, medium was removed, dishes washed with PBS. Fixing and Giemsa staining processes were carried as previously described (paragraph 2.6.1) **(OECD, 2007)**.

Protocol was slightly modified for 2-stage CTAs. After 4 days from the beginning of the recovery period, medium was removed and 15 ml of modified medium enriched with 0.1 µg/ml TPA dissolved in DMSO, diluting the stock solution 1:10'000 in the final volume of medium, were added until the 6th week with weekly changes. TPA addition was maintained throughout all the experiment. All dishes received the enriched medium except control, which received normal medium instead. The enriched medium followed the same pattern of serum concentration changes throughout the experiment: before confluence, cells were given

medium with 10% FBS; after confluence (3rd week) serum was lowered to 5% FBS. Fixing and Giemsa staining were performed as described (paragraph 2.6.1) **(OECD, 2007)**.

Giemsa stained colonies were circled with a permanent marker and were observed at the optical microscope with a 25x magnification to evaluate the standard morphological features for *foci* scoring **(Landolph, 1985)**.

Evaluated features were:

1. Multilayering
2. Polarization of the cells (rounded *versus* fibroblastic)
3. Presence of criss-crossing fibers
4. Intensity of staining
5. Invasion of the surrounding monolayer

Transformation frequencies (TF) were calculated as the ratio between the number of dishes presenting transformed *foci* and the number of observed dishes in the same treatment groups **(OECD, 2007).** Tallies were separately calculated both for Type II only, Type III only and for both of the transformed classes.

$$TF = \frac{\text{Number of transformed } foci \text{ counted}}{\text{Number of dishes counted}}$$

2.7 *In-focus* Immunofluorescence

2.7.1 Preliminary Studies

In order to elaborate a protocol allowing for the co-localization of a biochemical marker and the morphological structures needed for *foci*

classification, a preliminary model was built to demonstrate that fixing, immunostaining and Giemsa staining processes are compatible. As a test ground, C3H cells were seeded at 80'000 cells in 35 mm diameter petri dishes as for standard immunofluorescence assay (see paragraph 2.5). Anti β-Tubulin mouse monoclonal antibody (Sigma) was used at a dilution of 1:100 in PBS +1% BSA +0.1% NaN_3 for 1h at 37°C in the dark after methanol fixing **(Fumarola et Al., 2005)**. Plastic dishes and glass dishes were also tested to assess the eventual autofluorescence, not showing problematic background signal.

2.7.2 In-focus Immunofluorescence

In order to study the presence of biochemical markers inside a *focus*, a novel method was developed. Starting from the 3rd week of the CTA protocol (paragraph 2.6.2), glass and plastic petri dishes were thoroughly screened every week for *foci* formation at the optical microscope with a 25x magnification. Sterility was always conserved. Interesting colonies were highlighted on the external surface of the dishes and followed until the end of the experiment. Before fixing, both the selected colonies and an area of normal monolayer (negative control) of similar dimensions (around 20 mm side) on each petri dish selected for immunofluorescence, were highlighted with a permanent marker on the surface of the petri. Fixing was performed in 15 ml methanol as described for the CTA protocol (paragraph 2.6.2), with a particular attention in not deleting the

maker lines. Subsequently, the monolayer surrounding the marked lines out of the selected areas was carefully removed and the obtained square traced out with a normal pap-pen for histological analysis. Particular attention was placed in order to keep the colony at the center of the frame and not destroy the morphological structures of the possible *focus* in the process, especially for colonies growing close to the borders of the petri dishes.

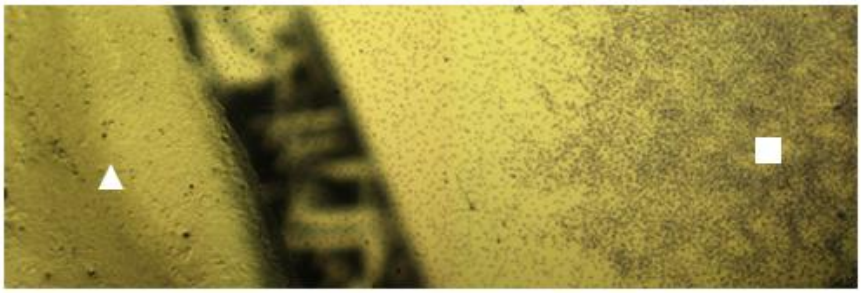


Figure 2.1. Pap-pen tracing of a *focus*
Pap pen tracing is showed on the left side of the image (white triangle). The procedure did not alter the border of the focus present on the right side (white square).

After tracing, frames were treated as normal histological slides following a slightly modified version of the protocol for indirect immunofluorescence described in paragraph 2.5. Petri dishes were washed with 15 ml blocking solution and the primary antibody, diluted 1:100 in blocking solution, was applied at the center of the frame in a volume of 50-80 µl (depending on the size of the frame) for 1 hour at 37 °C. Alexa Fluor 594 was used as secondary fluorescent antibody. An incubator with a fixed temperature of 37 °C acted as humid dark chamber for both steps of incubation. At the end of the incubation, dishes

were washed twice with PBS and twice with MilliQ water and stored in the dark at 4 °C. Observed was performed after 24 hours under the fluorescence microscope at 400x magnification.
After observation, dishes were rinsed in MilliQ water and Geimsa stained as previously described. *Foci* scoring was carried out as normal.

2.8 Isolation of *Foci* Derived Cell Lines

In order to isolate a cadmium transformed cell line for future biological and molecular characterization, transformed colonies obtained from cadmium treated dishes were scrape-harvested prior fixation and seeded in 35 mm petri dishes in high serum culture medium. Normal monolayer from control dishes (both medium only and TPA treated) as well as normal monolayer from dishes with scraped colonies were also harvested as controls. All harvests were performed in sterility conditions. Colonies were selected on order to establish cell lines derived both from obviously transformed *foci* and from *foci* of uncertain classification. Cells were cultured in normal condition and daily observed.

2.9 Gap Junction Intercellular Communication Assay (GJIC)

In order to assess if changes in the functionality of gap junctions were present in C3H cells treated with cadmium, the level of gap junction mediated cell-cell communication was evaluated by a scrape-loading assay as described by other Authors **(El-Fouly et Al., 1987)**.

C3H cells were seeded at 1.5 x 10^5 cell/dish on 20 mm side sterilized glass slides in a 35 mm diameter petri dish. After 3 days of observation, cells reached the optimal level of sub-confluence needed for experiments. Colture medium was removed and substituted with fresh medium added with 2 µM $CdCl_2$. Control group received medium only. After 24 hours, medium was removed and 1 ml PBS enriched with 0.5 mg/ml Lucifer Yellow CH dilithium salt (Sigma cat. L0259) was added in the dark to each sample, without removing the glass slides from the petri. Lucifer Yellow is a small molecular weight fluorescent dye which freely diffuses through gap junctions.

Three incisions on each sample were performed with a microtome blade perpendicular to the glass slide in presence of the dye. After 6 minutes of incubation, dishes were washed with PBS and observed at fluorescence microscope with a 400x magnification following producers instructions (Sigma). Images were taken from both sides of the cut and the number of dyed cells for observational field was counted. The presence of the cut was ascertained by light microscopy with a 25x magnification.

2.10 RNA Extraction and Whole Genome Microarray

HepG2 cell lines are human low deviation hepatoblastoma cells, presenting most of the activities of functional liver cells, with the advantage of being a stable line **(Urani et Al., 2007).** Due to cadmium hepatic toxic effects, an RNA fingerprint of this chemical might help

shedding some light in the molecular mechanisms of this metal.

HepG2 cells were seeded at a concentration of 1.5 x 10^6 cells per dish both for total RNA and miRNA-enriched RNA extraction. Treatments were performed 24 hours after seeding. For total RNA extraction, after medium removal, cells were added either with medium only (negative control) or with medium enriched with 2 µM $CdCl_2$ and 10 µM $CdCl_2$. Treatments were carried out for 24 hours. At the end of the treatments, culture medium was removed and cells were washed with ice cold PBS. Total mRNA extraction was performed in petri with 350 µl of RLT lysis buffer (Qiagen, Germantown, MD, USA) added with 20 µl of 2-mercapto-ethanol each 1000 µl of buffer, following producers instructions and stored at -80 °C. Total RNA purified extracts was analyzed on a Whole Human Genome Oligo Microarray chip (Agilent), containing 4 rows of about 41'000 unique genes and transcripts. Principal components analysis (PCA) was used to find patterns in the data.

For miRNA-enriched RNA extraction, cells were treated with 10 µM $CdCl_2$ or medium only (negative control). Extraction was performed in 750 µl mirVana buffer following producers instruction (Ambion).

RNA purification, quantification, microarray analysis and RT-PCR were performed in collaboration with the research group of Dott. Laura Gribaldo (Institute for Health and Consumer Protection, Molecular Biology and Genomics Unit, Joint Research Centre, Ispra – VA).

Chapter 3: *In Silico* Methodologies

3.1 Images Generation and Preprocessing

A Zeiss Axioplan™ microscope, equipped with a 25× lens (Carl Zeiss, Arese, Italy) and a digital camera (CoolSnap-ProColors Media Cybernetics, Bethesda, MA, USA), was used to take foci images, which were saved in 8-bit BMP gray-scale format. The original images had a size of 1392 × 1040 pixels. One pixel has an equivalent dimension of 0.3502μm, therefore 1μm = 2.8555 pixels.

Foci images were acquired from different *in vitro* cell transformation assays carried out in the same laboratory **(Bussinelli, 2007)**. The criteria of selection of subjects included both normal monolayers (MN) and all typologies of *foci* (untransformed Type I, transformed Type II and Type III). Due to uncontrollable experimental factors such as the efficiency of cell staining or the level of lamp illumination, images were characterized by a certain degree of variability. Therefore, a normalization routine was applied over a 0-255 interval of gray levels, rescaling the minimum value of each image to 0 and the maximum value to 255. Images were then filtered by Gaussian blurring over a 3 x 3 mask to “smoothen”, thus reducing the impact of the white noise over the image.

Each original image was taken trying to frame a whole *focus* in a single shot. This result was rarely achieved due to the mean dimensions of (especially transformed) colonies, far exceeding the observational field. *Foci* images were then cropped in a smaller 1024x1024 sized image

(thereon, “parent image”) and divided in 2 to 4 “tiles” of 512x512 pixel size, so that each tile framed a single, morphological significant detail of a *focus*. Overlapping of tiles was avoided or reduced to a minimum whenever possible. From a total number of 102 parent images, 248 tiles were generated. Each tile inherited the parent image scoring. Expert scoring divided the 102 parent images in 84 images of certain scoring (hereon, MD – Morphologically Defined images) and 20 images of doubtful scoring, including mixed and intermediate types of *foci* (hereon , MU – Morphologically Undefined images).

It is noteworthy that, while the choice of what had to be considered a *significant detail* was done by the trained expert on the basis of classification criteria defined by other Authors **(Landolph 1985)**, no evaluation of the framed features was done, in order to keep the final classifier unbiased and unsupervised. In most cases, regions of interest coincided with border regions; therefore, the tile selection process closely resembled a segmentation process.

3.2 Extraction of Statistical Image Descriptors

Tiles were processed and analyzed with the R 2.8.1 software environment for statistical computing **(R Development Core Team, 2009)**. Routines for the extraction of descriptors and clustering of images were derived from the “cluster” **(Maechler, 2010)**, “biOps” **(Bordese et**

Al., 2007) and “fBasics” **(Wuertz et Al., 2010)** packages for R, all available at **http://cran.r-project.org**.

Nineteen statistical image descriptors were extracted from each tile. Eighteen out of the total were obtained by means of functions included in the “fBasics” package, as summary statistics calculated on the normalized distribution of pixel gray levels: 15 percentiles from the 15^{th} to the 85^{th}, arithmetic mean, skewness and kurtosis. Percentiles are levels of gray below which a certain percent of observations fall, therefore the 15^{th} percentile is the level of gray below which the 15% of the gray pixels of the image fall. Skewness is a measure of asymmetry of a distribution, and kurtosis is a parameter of how much a distribution is sharp around a mean value. Taken together, these parameters resume the most important features of pixel distribution of an image.

The last descriptor, defined by the proportion of edge pixels over the total image, was derived through the use of the Canny edge detectin algorithm. While summary statistics are better suited to capture the macroscopic features of an image, this last descriptor is more closely related to the fine structure of *foci*. In fact, by means of this algorithm, it is possible to evaluate the degree of complexity of the border of a *focus*, by exploiting the gradient of the image in four different directions and identifying the pixels corresponding to the edges of the image itself **(Deriche, 1987)**.

All the nineteen extracted descriptors are orientation independent. Finally, a comprehensive database was built where each of the 248 tiles represented a single entry. The comprehensive archive was divided in two subsets, one for the 196 tiles derived from the 82 MD parent images, and another consisting of the 52 tiles generated from 20 MU parent images.

3.3 Classification Scheme

Clustering of the sets of image tiles into groups was performed exploiting the Partitioning Around Medoids (PAM) algorithm. This algorithm automatically searches for *k* representative objects (medoids) inside a dataset in order to divide it in a *k* user-assigned number of classes. The rest of the entries are then clustered in order to minimize an intra-class dissimilarity measure while maximizing the inter-class dissimilarity **(Kaufman et Al., 1990)**. A graphic representation of the clustering results was achieved by means of principal component analysis (PCA). Through the use of this mathematical model, it was possible to visualize in a two dimensional plot (2 variables) the projections of the higher-dimensional point descriptors (19 variables), while retaining most of the variability of the original dataset **(Jolliffe, 1986)**.

Tile classification was performed in two steps. In the first step (hereon, F1), the classifier was used to divide the dataset in two groups (*k=2*). This separation divided positive/transformed tiles from

negative/untransformed tiles and exploited the first eighteen statistical descriptors. In the second step of classification (hereon, F2), two further runs (each with *k=2*) were applied to further subdivide the F1 generated groups in two for a total number of four classes, thus reflecting the four biological classes structure (MN and the three classes of *foci*). The Canny's edge detection derived descriptor was added during this second step.

Training step was performed both on MD subset and on the comprehensive archive to study the impact of the dimension of a database versus the level of "noise", as introduced by MU images.

Classification step was performed on MU tiles in the space generated from the MD-only *training.*

This approach can thus be summarized as a *two-step hierarchical unsupervised* classification approach. It is *two-step* due to the sequential clustering steps F1 and F2; it is *hierarchical* because the results of the first classification run are used as dataset for the second run; it is *unsupervised* because no information regarding the expert scoring was exploited by the classifier for clustering purposes **(Figure 3.1)**.

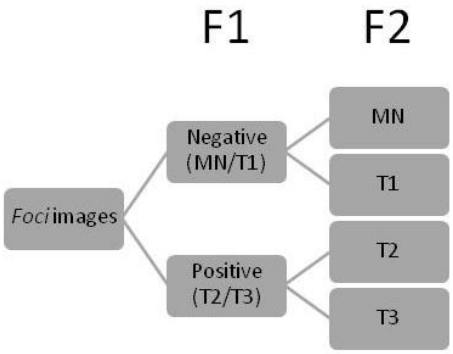


Figure 3.1. *Two-step hierarchical* classification scheme.
A schematic of the Two-step hierarchical classification is shown. A dataset of *foci* images is divided by the algorithm in two classes during F1, corresponding to positive (T2/T3) and negative (MN/T1). During F2, images classified as negative are divided between normal monolayer (MN) and Type I class *foci* (T1). Another classification run divides images classified as positive in Type II (T2) and Type III (T3) classes. The approach is two-step since it's based on two different classification runs and is hierarchical because the results of the first run are used as database for the second run **(Procaccianti et Al., 2011)**.

3.4 Evaluation of Results

After each classification run, a label was given by the expert to the obtained clusters on the basis of the most represented biological class within each group. For example, the label of C2 was given to each F1 group containing a majority of Type II tiles. Classifier performances were evaluated as the agreement between classifier scoring and expert scoring by means of confusion matrices. These are two-way tables built listing the expert scoring by row and the classifier scoring by column. Matrices values indicated the number of tiles labeled in each group, with elements along the diagonal representing the images where expert and classifier agreed upon. Agreement was only evaluated for MD images

and calculated as the ratio number of images classifier and expert agreed upon (thus, the diagonal of the confusion matrix) and the total number of images submitted for the classification run. Other parameters exploited to evaluate the performances were sensitivity, calculated as the ratio between true positives and the sum of true positives and false negatives, and sensibility, as the ratio between true negatives and the sum of true negatives and false positives. These two values summarize the ability of the classifier to recognize real positive and real negative results respectively.

When assigning the parent image class, majority of voting criterion was applied between tiles generated from same parent. Confusion matrices were built as previously described, with the slight modification of having the inferred score instead of the classification score by column.

3.5 Morphologically Undefined Images Classification

After classification step F2 **(Figure 3.1),** the centers of each of the four obtained classes were estimated as the mean value of the nineteen descriptors of the tiles assigned to the same cluster. This process generated a “virtual tile” representing the quintessential characteristics of each classification group: each class center therefore represented the value of the nineteen used descriptors for a perfect image belonging to each given class. The Euclidean distance between each MU tile and the four virtual tiles were calculated and averaged for each parent image,

thus assessing the mean distance of a parent image to the four classification groups. A Quantitative Index of Dissimilarity for a parent image to each of the four canonical classes ($QIoD_c$) was calculated as the ratio between the mean Euclidean distance of the i^{th} MU parent image to the class center ($\overline{D}_{i,c}$ where c is one of the four class centers) and the sum of all the distances of the same image to all the four centers. The Euclidean mean distances were calculated as:

$$\overline{D}_{i,c} = \frac{\sum_{t=1}^{n} D_{i,t,c}}{n}$$

where D is the standardized distance of the t^{th} tile of the i^{th} image to the c class center, and n is the number of generated tiles for the i^{th} parent image

The final formula for the index of dissimilarity is:

$$QIoD_c = \frac{\overline{D}_{i,c}}{\overline{D}_{i,MN} + \overline{D}_{i,T1} + \overline{D}_{i,T2} + \overline{D}_{i,T3}}$$

The total sum of the $QIoD_c$ over the four c classes was always set to 1. This index represents in a quantitative scale how dissimilar a MU parent image is from the representative characteristics of each biological class, and as an inferred result, which class is the most similar.

3.6 Generation of a Whole-*focus* Images Database

It is often impossible to frame a whole *focus* in a single shot. In this context of fragmentation of images, information connected to the complete morphologic structure of transformed colonies might be lost. In

order to move our research towards a whole-*focus* analysis, a new approach in image acquisition was developed. *Foci* images were acquired following a grid pattern, with a 40-50% of intentional overlapping between each shot. Images were merged with Adobe Photoshop CS5 merging routines on a private computer.

Foci dishes used for the generation of this database result from a kind collaboration with Dott.ssa R. Corvi (In vitro Methods Unit/European Centre for the Validation of Alternative Methods, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy).

Chapter 4: *in vitro* results

4.1 C3H Survival is not Impaired by Cadmium Treatment

MTT experiments were performed in order to assess the dose/time schedule of cadmium treatments not inducing cytotoxicity in C3H cells. On day 2 and 3 no obvious difference was found between control and treatment. On day 4, control dishes had a slightly higher level con confluence than treated dishes. Moreover, 1µM $CdCl_2$ treatment presented little signs of toxicity and cell death. No reduction of cell viability was found after spectrophotometer readings. In fact, viability of C3H cells after 72 hours of treatment was over 80% with 1µM $CdCl_2$ and around 90% with 1 µg/ml B[a]P **(Figure 4.1, panels A and B respectively)**.

Viability of HepG2 cells exposed to cadmium has been previously demonstrated in this laboratory as a survival rate of 80-90% after treatment with a highest concentration of 10 µM $CdCl_2$ **(Urani et Al., 2005)**.

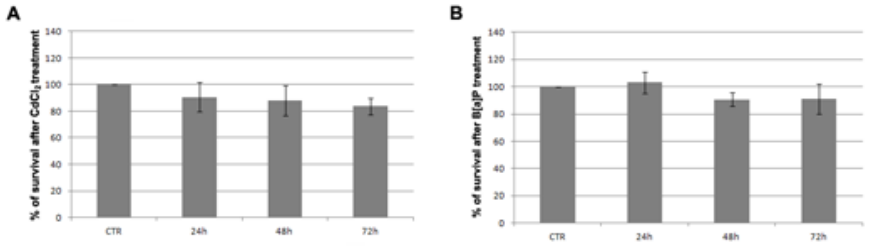


Figure 4.1. Survival of C3H cells after cadmium treatment.
C3H cells were treated with 1 µM $CdCl_2$ for 24, 48 and 72 hours. 1 µg/ml B[a]P was used as positive control for the same durations. Negative control was medium only. At the end of each treatment period, cells were added with MTT and incubated for 4 hours at 37 °C. Formazane crystals were collected in DMSO and samples read at the spectrophotometer. Neither of the samples showed a reduction in survival above 20%.

4.2 Effects of Cadmium on HSP70 and p53 Stress Markers Levels

Effects of cadmium on stress markers were evaluated as a further proof of stress induction. Protein p53 was chosen as a typical genotoxic damage marker, while HSP70 as a more generic stress indicator. Treatment of C3H cells with a wide range of cadmium concentration ranging from 0.1 µM to 2 µM for 24 hours showed no alteration of p53 expression levels **(Figure 4.2, panel A; Figure 4.3 panel A)**. While no change was observed in HSP70 levels at low doses, a strong peak of synthesis has been found at 2 µM. No changes have been found on treatment with DMSO and B[a]P **(Figure 4.2, panel B; Figure 4.3 panel B)**.

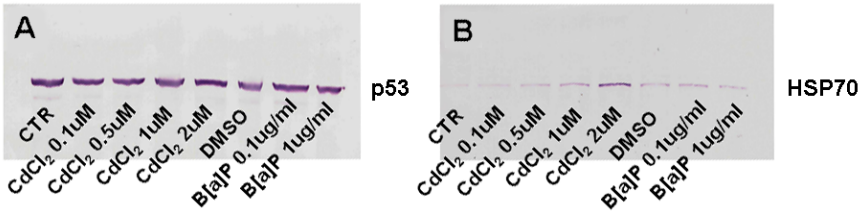


Figure 4.2. Western Blotting of C3H cells lysates after cadmium treatment.
C3H cells were treated with scalar concentrations of cadmium as test compound and positive control was B[a]P in DMSO, all for 24 hours. Negative controls were medium only (CTR) and B[a]P vehicle (DMSO). Panel A: No alteration was found on p53 expression levels. An exemplificative blot is shown. Panel B: Results of HSP70 expression levels are shown in an exemplificative blot. A peak of activity is present at 2 µM.

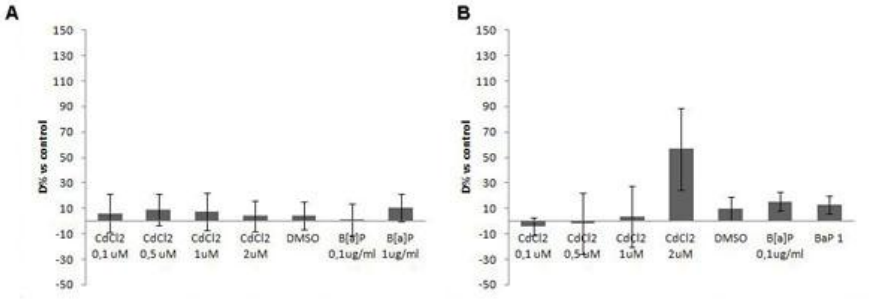


Figure 4.3. Expression levels of stress markers in C3H cells after cadmium treatment.
Cells were treated with $CdCl_2$ at a range of scalar concentration for 24 hours. B[a]P dissolved in DMSO was used as positive control. Medium only (CTR) or B[a]P vehicle (DMSO) were used as negative controls. At the end of each treatment, cells were lysed and blotted for the presence of p53 as a genotoxic marker and for HSP70 as a generic stress marker. Levels of p53 resulted not altered in both cadmium and B[a]P treatment groups (Panel A). Levels of HSP70 showed a peak of concentration at $CdCl_2$ 2 µM more than 50% stronger than control (panel B).

4.3 Cell Transformation Assay

In vitro cell transformation assays were performed in C3H cells treated with low concentration of cadmium, both alone (1-stage) and with the

administration of the inductor TPA (2-stage). Obtained *foci* were used for co-localization studies, in order to assess the correspondence between biochemical markers and morphological features.

4.3.1 Cytotoxicity Assay in 1-stage CTA

Evaluation of cytoxicity is preliminary to cell transformation assays to assess the non-toxic effect of the treatments performed.

Cadmium chloride was used as test compound in 1-stage cytotoxicity assays at concentrations of for 24 hours at a concentration of 1 µM and 2 µM **(Tables 4.1 and 4.3 upper part)** and for 48 hours at a concentration of 1µM **(Table 4.3 upper part**. B[a]P dissolved in DMSO was used as positive control at a concentration of 1µg/ml for 24 **(Table 4.1 and 4.3 upper part)** and 48 hours **(Table 4.3 upper part)**. Negative control dishes received fresh medium instead of treatments. Obtained colonies were circled after fixing and staining, and counted for numerosity (threshold 50 cells/colony) by optical microscopy observation with 25x magnification **(Figure 4.4)**.

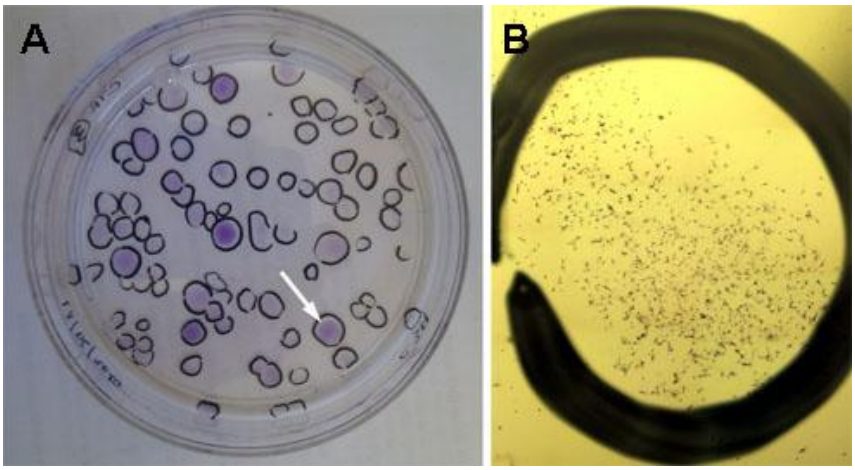


Figure 4.4: Cytotoxicity results and colony counting
A dish obtiained in a cytotoxicity experiment is shown in panel A. Colonies (white arrow) are dark stained against the dish and circled with a permanent marker. The observation at 25x magnification under the optical microscope is shown in panel B. Dark spots are cells and their number has to be over 50 for a colony to be counted towards the plating efficiency and survival evaluation

Plating efficiency was (PE) calculated as

$$PE = \frac{\text{Number of colonies counted}}{\text{Number of cells seeded}}$$

Values were above 28% in all the experiments performed, therefore no serum batch was rejected.

Survival (S) was calculated as the ratio between the number of colonies in treated dishes and the number of colonies in control dishes.

$$S = \frac{\text{Number of colonies counted in treatment groups}}{\text{Number of colonies counted in control groups}}$$

In one case, 1µM $CdCl_2$ survival rates were found to be slightly below 50%. On the other hand, a noteworthy number of colonies with a cell population of 40-45 cells was found in the same treatment group.

Treatment with 2µM $CdCl_2$ resulted in a much lower survival rates than previously reported **(Bussinelli, 2007) (Tables 4.1 and 4.3 upper part)**

Treatment	Mean	SD	PE	S
CTR	76.00	±6.93	38.00	
$CdCl_2$ 1 µM 24h	30.67	±4.73		40.35
$CdCl_2$ 2 µM 24h	15.00	±4.00		19.74
B[a]P 1 µg/ml 24h	55.67	±2.31		73.25

Table 4.1. Results of cytotoxicity assays on C3H cells in 1-stage CTA.
C3H cells were plated at low density and treated with cadmium chloride at 1 µM and 2 µM for 24 hours. Positive control was 1 µg/ml B[a]P and negative control medium only. PE value for negative control is 38%. Treatment with 1 µM $CdCl_2$ survival resulted slightly below 50%, nevertheless numerous colonies of 40-45 cells were counted. Treatment with 2 µM $CdCl_2$ survival value is slightly below 20%. Positive control resulted in more than 70% of survival.

4.3.2 Cytotoxicity Assay in 2-stage CTA

Treatment groups for 2-stage CTAs underwent a slight modification of protocol in comparison with standard 1-stage CTAs. In addition to the negative control (medium only), a TPA negative control group was added **(OECD, 2007)**. The PE value was 29%, and the serum batch approved for following CTA experiments. The survival (S) value for the negative control treated with TPA has been found above 85%, therefore the lack of toxicity of this treatment was confirmed. A comparison between same treatment groups in 1-stage and 2-stage cytotoxicity assays, shows that survival values are only slightly affected by TPA induction. **(Tables 4.2 and 4.3 lower part)**.

Treatment	Mean	SD	PE	S
CTR	58.00	±11.27	29.00	
TPA	50.00	±5.57		86.21
$CdCl_2$ 1 µM 24h + TPA	36.67	±6.03		63.22
$CdCl_2$ 2 µM 24h + TPA	10.33	±0.58		17.82
B[a]P 1 µg/ml 24h + TPA	47.33	±2.08		81.61

Table 4.2. Results of cytotoxicity assays on C3H cells in 2-stage CTA. C3H cells were plated at low density and treated with cadmium chloride at 1 µM and 2 µM for 24 hours. Positive control was 1 µg/ml B[a]P and negative control medium only (CTR) or TPA. Latter group was given medium only for the first 4 days after the beginning of the recovery period, and after the 4th day medium was changed and added with 0.1 µg/ml TPA in DMSO. PE value for negative control is 29% (CTR). Treatment with 1 µM $CdCl_2$ survival resulted slightly above 60%. Treatment with 2 µM $CdCl_2$ survival value is below 20%. Positive control resulted in more than 80% of survival. TPA negative control resulted in over 85% of survival.

.

Treatment	Mean	SD	PE	S
CTR	88.00	±18.33	44.00	
$CdCl_2$ 1 µM 24h	79.33	±11.02		90.15
$CdCl_2$ 1 µM 48h	50.67	±7.02		57.58
B[a]P 1 µg/ml 24h	72.67	±10.69		82.58
B[a]P 1 µg/ml 48h	52.33	±8.08		59.47
TPA	105.33	±3.06		119.70
$CdCl_2$ 1 µM 24h + TPA	72.67	±6.51		82.58
$CdCl_2$ 1 µM 48h + TPA	59.67	±3.51		67.80

Table 4.3. Results of cytotoxicity assays on C3H cells in 1-stage and 2-stage CTA. Upper part: Cytotoxicity results of 1-stage CTA. Cells treated with medium only (CTR) showed a PE level of 44%. Treatment for 24 and 48 hours with $CdCl_2$ and positive control (B[a]P) resulted above 55% in all cases, with a top of 90% in 1 µM $CdCl_2$ treatment for 24 hours. Bottom part: Cytotoxicity results of 2-stage CTA. Negative control (CTR) was common with the upper part. TPA negative control resulted in over 100% of survival. 1 µM $CdCl_2$ survival was enhanced by TPA addition both in 24 and 48 hours treatment groups in comparison with relative 1-stage values (as showed in upper part)

4.3.3 Scoring of Foci

Once fixed and Giemsa stained, petri dishes were observed at the optical microscope with a 25x magnification. Type I, Type II and Type III *foci* were classified based on standard morphological features **(Landolph, 1985)**. Type I *foci* presented only a slight packing of rounded, monolayered cells. Type II *foci* showed a higher degree of packing with a multilayering growth in virtually opaque layers. Cells were moderately polar. Type III *foci* showed pronounced multilayering, deep basophilic staining and a polarized aspect. Invasion spikes were also present, with vortex-like formations at their sides. *Foci* presenting intermediate features between classes were present and relatively common in cadmium treatment groups **(Figure 4.5, panels A-F)**.

Transformation frequencies (TF) were calculated as **(OECD, 2007)**:

$$TF = \frac{\text{Number of transformed } \textit{foci} \text{ counted}}{\text{Number of dishes counted}}$$

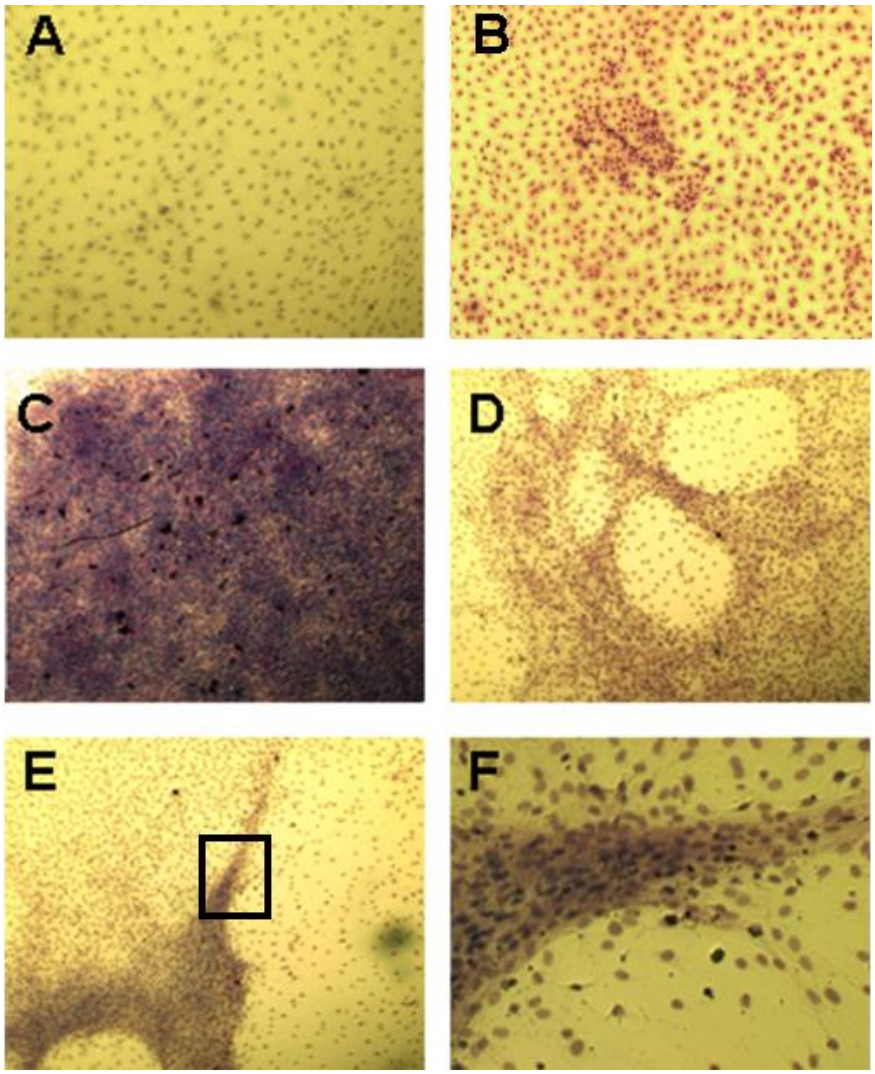


Figure 4.5. Example images of *foci* from cadmium treatment and control monolayer. Panel A shows a normal control monolayer. Type I *foci* (panel B) showed tightly packed cells with no multilayering. Panel C shows a Type III *focus* center, with deep staining, packing of cells and multilayered growth patterns. Panel D shows an intermediate *focus* border. Though invasion and vortex formations are present (typical features of Type III *foci*), cells are still well rounded and multilayering is not pronounced. Panel E shows a detail of a more typical invading *focus*, while panel F shows a detail of the invasion spike. Cells are multilayered, grow all in the same direction and have a fibroblastic shape. Vortexes are present on the sides of the spike. All images were acquired at optical microscope. Panels A-E with 25x magnification, panel F with 100x magnification.

4.3.4 Results of 1-stage Cell Transformation Assay

Cell transformation assays conditions and serum batch were as used in the corresponding preliminary cytotoxicity assay (paragraph 4.3.1).

At the end of the 6 weeks recovery period, dishes were fixed and stained, and observed under optical microscope for *foci* counting. Colonies were divided by type **(Table 4.4, panels A and C upper part)** as per standard classification protocol **(Landolph, 1985)**. Transformation frequencies (TF) were calculated on tallies recorded for Type II and Type III, both separately and united. Mixed and intermediate class *foci* were counted and scored as higher damage class for determining the TFs: Type I/Type II intermediate *foci* were scored as Type II, while Type II/Type III intermediate were scored as Type III **(Table 4.5, panels A and C upper part)**.

Treatment groups were set up both on 100 mm glass petri dishes and 100 mm plastic petri dishes. No difference in growth patterns was found between plastic and glass throughout the weekly observations. Control groups never resulted in transformed colonies formation. Positive control treatment group resulted in the formation of a range of transformed *foci*. Cadmium chloride treatment resulted in the formation of a majority of intermediate *foci*, with less pronounced invasive features than B[a]P treatment.

4.3.5 Results of 2-stage Cell Transformation Assay

Cell transformation assays conditions and serum batch were as used in the corresponding preliminary cytotoxicity assay (paragraph 4.3.2). Negative control groups, both medium only and medium plus TPA induction never resulted in transformed colonies In three cases, TPA treatment resulted in non transformed, Type I *foci* **(Table 4.4, panel B and C bottom part)**. All treatment groups transformation frequencies resulted enhanced by TPA treatment in comparison with the 1-stage assay **(Table 4.5, panel B and C bottom part)**.

Starting from week five, 2µM $CdCl_2$ + TPA treatment group showed signs of toxicity and cell death **(Figure 4.6).**

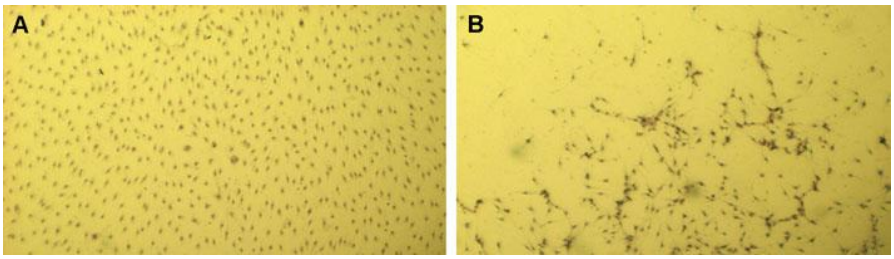


Figure 4.6. Cytotoxic effect of cadmium 2 µM effects upon TPA induction.
An image obtained from a cadmium 2 µM treated dish from a 2-stage CTA is shown in panel B. Normal control monolayer is shown in panel A for comparison purposes. Toxic effects of the cadmium + TPA treatment are evident in terms of cell death and loss of the monolayer.

A

Treatments	N° dishes	Type I	Type I/Type II (MU)	Type II	Type II/Type III (MU)	Type III
CTR	10	16	0	0	0	0
$CdCl_2$ 1 µM 24h	10	27	0	0	0	0
$CdCl_2$ 2 µM 24h	10	24	3	0	0	0
B[a]P 1 µg/ml 24h	10	26	3	2	1	1

B

Treatments	N° dishes (PI + GI)	Type I	Type I/Type II (MU)	Type II	Type II/Type III (MU)	Type III
CTR	2+2	0	0	0	0	0
TPA	2+2	0	0	0	0	0
$CdCl_2$ 1 µM 24h + TPA	5+4	10	8	1	1	1
$CdCl_2$ 2 µM 24h + TPA	5+5	15	7	4	2	0
B[a]P 1 µg/ml 24h + TPA	5+3	3	1	3	1	4

C

Treatments	N° dishes	Type I	Type I/Type II (MU)	Type II	Type II/Type III (MU)	Type III
CTR	3	0	0	0	0	0
$CdCl_2$ 1 µM 24h	10	14	2	3	0	1
$CdCl_2$ 1 µM 48h	10	14	1	2	0	1
B[a]P 1 µg/ml 24h	7	2	1	2	1	1
B[a]P 1 µg/ml 48h	7	3	2	4	3	1
TPA	3	3	0	0	0	0
$CdCl_2$ 1 µM 24h + TPA	9	4	4	4	0	1
$CdCl_2$ 1 µM 48h + TPA	8	8	1	2	1	2

Table 4.4. Counts of *foci* from cell transformation assays performed on C3H cells.
Panel A and Panel C (upper part) show the results of *foci* counting in different 1-stage CTA experiments. Numbers of dishes counted are included (second column). Panel B and Panel C (bottom part) are relative to 2-stage CTA experiments. Panel C upper and bottom part belong to the same experiment but with different protocols (1-stage CTA in the upper part, 2-stage in the bottom part). All cadmium treatment groups started from the same number of dishes (10), but some were lost due to fungine contaminations. No contaminations occurred in control dishes.Tallies are recorded separately for Type I, Type II and Type III canonic *foci*. Mixed and intermediate are recorded separately.

A

Treatments	Type II	TF (Type II)	Type III	TF (Type III)	All	TF (All)
CTR	0	0.00	0	0.00	0	0.00
$CdCl_2$ 1 µM 24h	0	0.00	0	0.00	0	0.00
$CdCl_2$ 2 µM 24h	0 (+2)	0.00 (0.20)	0	0.00	0 (+2)	0.00 (0.20)
B[a]P 1 µg/ml 24h	3 (+2)	0.30 (0.50)	1 (+1)	0.10 (0.20)	4 (+3)	0.40 (0.70)

B

Treatments	Type II	TF (Type II)	Type III	TF (Type III)	All	TF (All)
CTR	0	0.00	0	0.00	0	0.00
TPA	0	0.00	0	0.00	0	0.00
$CdCl_2$ 1 µM 24h + TPA	1 (+6)	0.10 (0.70)	1 (+1)	0.10 (0.20)	2 (+7)	0.2 (0.90)
$CdCl_2$ 2 µM 24h + TPA	2 (+5)	0.22 (0.78)	1 (+2)	0.11 (0.33)	3 (+7)	0.33 (1.11)
B[a]P 1 µg/ml 24h + TPA	2 (+1)	0.25 (0.38)	3 (+1)	0.38 (0.50)	5 (+2)	0.63 (0.88)

C

Treatments	Type II	TF (Type II)	Type III	TF (Type III)	All	TF (All)
CTR	0	0.00	0	0.00	0	0.00
$CdCl_2$ 1 µM 24h	3 (+4)	0.30 (0.70)	1	0.10	4 (+4)	0.40 (0.80)
$CdCl_2$ 1 µM 48h	2 (+1)	0.20 (0.30)	1	0.10	3 (+1)	0.30 (0.40)
B[a]P 1 µg/ml 24h	2 (+1)	0.29 (0.43)	1 (+1)	0.14 (0.29)	3 (+2)	0.43 (0.71)
B[a]P 1 µg/ml 48h	4 (+3)	0.57 (1.00)	1 (+2)	0.14 (0.43)	5 (+5)	0.71 (1.43)
TPA	0	0.00	0	0.00	0	0.00
$CdCl_2$ 1 µM 24h + TPA	3 (+3)	0.33 (0.66)	1	0.11	4 (+3)	0.44 (0.78)
$CdCl_2$ 1 µM 48h + TPA	1 (+1)	0.13 (0.25)	2 (+1)	0.25 (0.38)	3 (+2)	0.38 (0.63)

Table 4.5. Transformation frequencies (TF) obtained from CTA performed on C3H cells. Panel A and Panel C (upper part) show the results of *foci* transformation frequencies in different 1-stage CTA experiments. Panel B and Panel C (bottom part) are relative to 2-stage CTA experiments. Panel C upper and bottom part belong to the same experiment but with different protocols (1-stage CTA in the upper part, 2-stage in the bottom part). TFs are calculated separately for Type II and Type III canonic *foci*. Mixed and intermediate are recorded separately (in brackets). Results are shown for transformed types *foci* both separately and united (All). Type I *foci* were excluded from TFs evaluation due to their non transformed nature.

4.4 Cellular Localization and Functionality of Gap Junctions Upon Cadmium Treatment

Cell transformation assays results confirmed the carcinogenic potential of cadmium in C3H model (paragraph 4.3). Since Cx43 has a strong connection with transformation processes even when genotoxic processes are not involved (paragraph 1.9), the localization and the alteration of functionality of this marker upon cadmium treatment were studied.

4.4.1 Cx43 Localization is Changed by Cadmium Treatment

C3H cells were treated with 1 µM $CdCl_2$ to evaluate the changes of Cx43 distribution. Positive control used was 1 µg/ml B[a]P dissolved in DMSO. Negative control (medium only) show deposits of Cx43 in the peri-nuclear region **(Figure 4.7, panel A)**. After 24 hours of cadmium treatment, Cx43 is completely removed from the deposits, forming small dots with a strong fluorescence both in the cytoplasm and inside the nucleus **(Figure 4.7, panel B)**. Positive control with B[a]P showed after 24 hours a milder effect, with a concurrent presence of both perinuclear deposits and nuclear dots **(Figure 4.7, panel C)**.

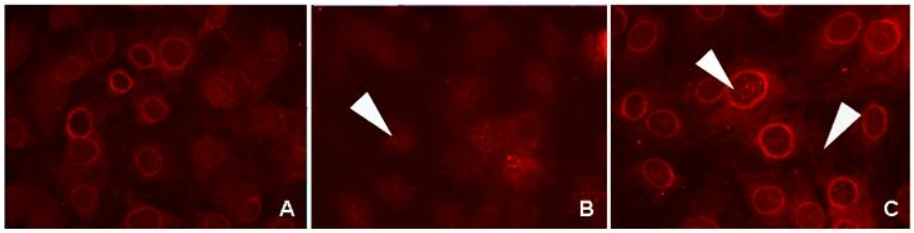


Figure 4.7. Localization of Cx43 in C3H cells upon cadmium treatment.
C3H cells were treated with 1 µM $CdCl_2$ for 24. 1 µg/ml B[a]P was used as positive control. Treatments were carried on for 24 hours. Immunofluorescence staining with anti-Cx43 antibody showed in the control cells (panel A) a localization of Cx43 at the level of a perinucler ring. Panel B: treatment with $CdCl_2$ induces a full displacement of this connexin and a formation of intensely stained dots in the nucleus (arrowheads). Panel C: Treatment with B[a]P shows a milder mobilization, with the same formation of stained dots (arrowheads). Images were taken at 400x magnification and slightly modified for a better visibility.

4.4.2 Cadmium Treatment Reduces GJIC Levels

Since alterations in the localization of Cx43 were found, the level of intercellular communication was studied to ascertain eventual changes in the functionality of gap junctions.

C3H cells were seeded on Day 0 and treated on Day 3, allowing them to grow to subconfluent level. Cells were then treated with 1 µM $CdCl_2$ for 24 hours, negative control received medium only. Images of the cutting line were acquired under normal light illumination at 25x magnification. Fluorescence images were acquired at 400x magnification from both sides of the cut, and the number of fluorescent cells per observation field counted **(Fig 4.8, panels A and B)**. A reduction of 50% in the number of dyed cells was found between control and cadmium treatment **(Fig. 4.8**

panel C). The cutting of the monolayer was proven by observation in normal white light at 25x magnification **(Fig. 4.8 panel D)**.

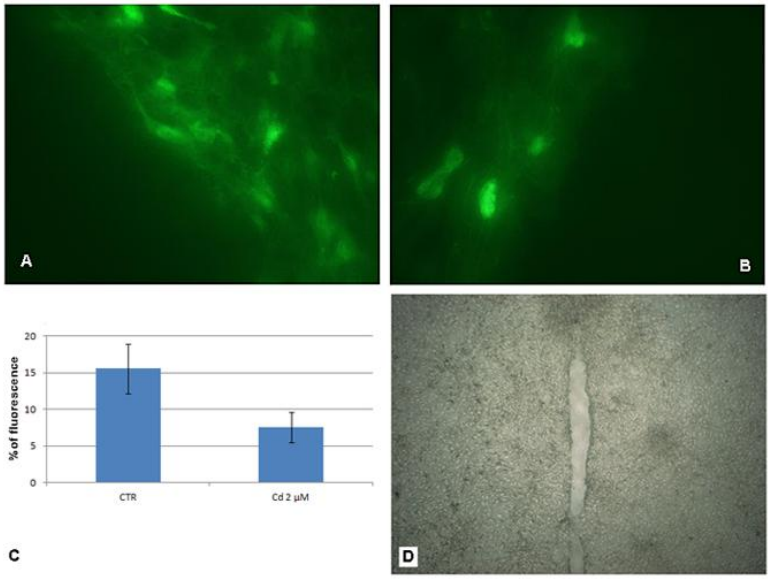


Figure 4.8. GJIC assay results upon cadmium treatment.
C3H cells were treated with $CdCl_2$ 1 µM or vehicle only (control) for 24 hours 3 days after plating. Cell monolayer was cut with a microtome blade in presence of a solution of 0.5 mg/ml of LY. The dye was left in incubation for 6 minutes, then gently removed and cells briefly rinsed in PBS. Images of the cutting edge of control cells are shown in panel A. Treated cells are shown in panel B. The number of dyed cells was counted per observational field. Pictures were taken at 40x magnification for counting purposes. Cadmium treatment resulted in around 50% of reduction in the number of dyed cells (panel C). A normal light image of the cut was taken at 2.5x magnification (panel D).

4.5 *In-focus* Co-localization of Biochemical and Morphological Markers

The experiments described in paragraph 4.4 demonstrated how Cx43 shows alterations in its pattern of localization upon cadmium treatment. In order to determinate if such alterations were also present in cells that

acquired a transformed phenotype upon cadmium treatment at the end of a cell transformation assay, a novel technique for the evaluation of the localization of biochemical markers *in-focus* was developed.

Preliminary experiments assessing the compatibility of indirect immunofluorescence and Giemsa staining were performed, with the aim of determining if the combination of the two procedures could alter cellular morphology. A test antibody for a constitutively expressed protein like β-tubulin was chosen as a test marker. Giemsa staining after the indirect immunofluorescence procedure showed no alteration of cellular morphology **(Figure 4.9)**.

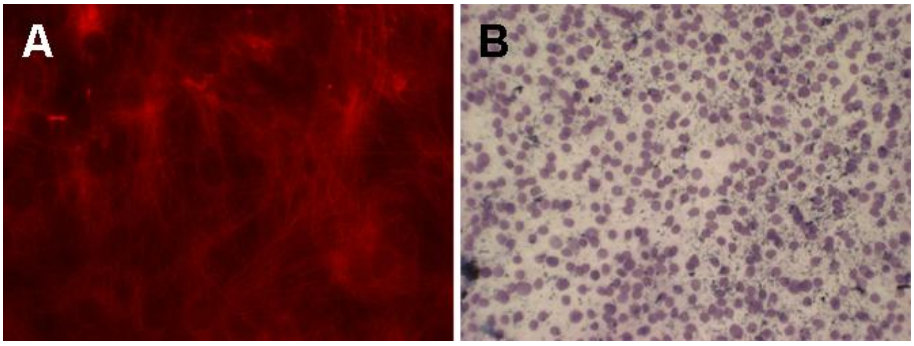


Figure 4.9. Indirect Immunofluorescence and Giemsa Staining of C3H Cells.
C3H cells were seeded and left growing for 3 days, then fixed in methanol. Anti-β tubulin primary monoclonal antibody was used. Images of fluorescence were taken at 400x magnification (panel A). Dishes were then rinsed and Giemsa stained. Giemsa staining did not show alterations in normal cell morphology (panel B. Image taken at 25x magnification).

Therefore, the preliminary results demonstrated how these two protocols were compatible, allowing *in-focus* immunofluorescence experiments to be performed.

CTA dishes were weekly screened at the optical microscope with 25x magnification for *foci* formation, starting from the 3rd week of the assay. At the end of the CTAs, the most promising colonies were selected to be analyzed for co-localization of biochemical and morphological markers. The area surrounding each colony was highlighted with a permanent marker on the base of the dish. Both plastic and glass dishes were used; plastic dishes did not show relevant signs of auto-fluorescence. For each *focus*, another area of normal monolayer of similar dimensions was also selected in the same dish, as an internal negative control. After fluorescence image acquisition was performed, dishes were processed for Giemsa staining as described in detail in paragraph 2.6, and scored as suggested by standard classification protocols **(Landolph 1985)**. Results of classification are showed in paragraph 4.5.

Normal monolayer cells obtained from control dishes (medium only) and hybridized with anti-Cx43 showed faint levels of diffuse fluorescence. Giemsa staining confirmed the lack of morphologic alterations **(Figure 4.10, panels A and B)**. Normal monolayer cells selected from dishes treated with 1 µM $CdCl_2$ showed similar levels of diffuse fluorescence as in negative control cells. Giemsa staining confirmed their normal phenotype **(Figure 4.10, panels C and D)**.

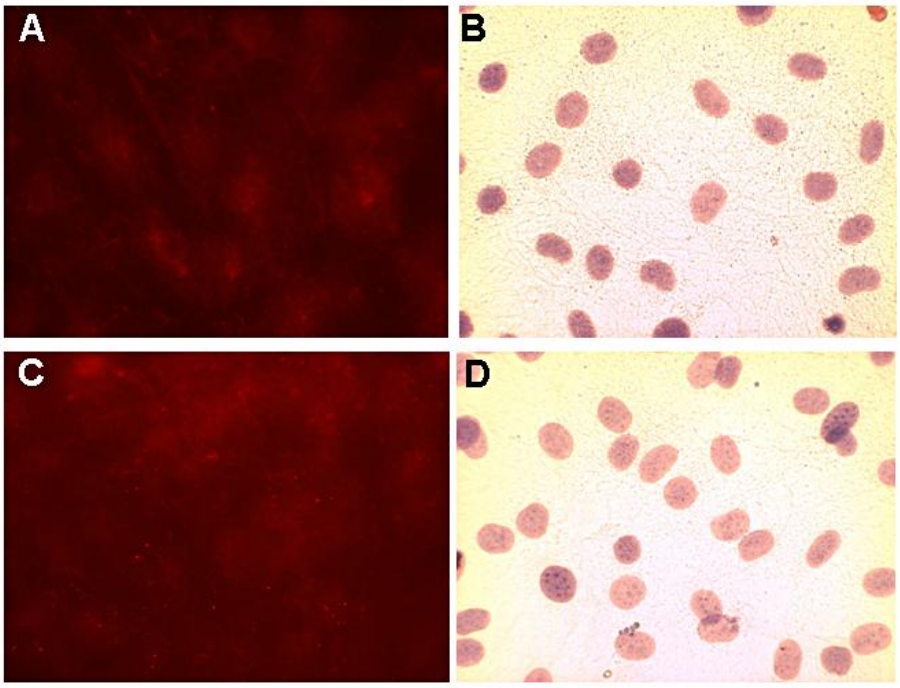


Figure 4.10 Levels of Cx43 fluorescence and morphological features of normal cells. Cell transformation assay was carried out as described in paragraph 2.6 and 2.7. After fixing, monolayer areas were isolated by pap pen tracing and stained for indirect immunofluorescence with an anti-Cx43 primary antibody. Negative control cells (panel A) showed faint diffuse levels of fluorescence. Giemsa staining is shown (panel B). Normal cells from 1 µM cadmium treated dishes showed similar levels of fluorescence in comparison with negative control (panel C). Giemsa staining confirmed the normal morphology. All images were taken at 400x magnification.

The hybridization of 1 µM cadmium transformed *foci* with the anti-Cx43 monoclonal antibody showed a strong change in the distribution of the selected marker in comparison with normal cells. In fact, lined up, intensely stained dots were clearly visible in transformed *foci* **(Figure 4.11, panel A)**. Giemsa staining confirmed the transformed nature of the observed *focus* **(Figure 4.11, panel B)**.

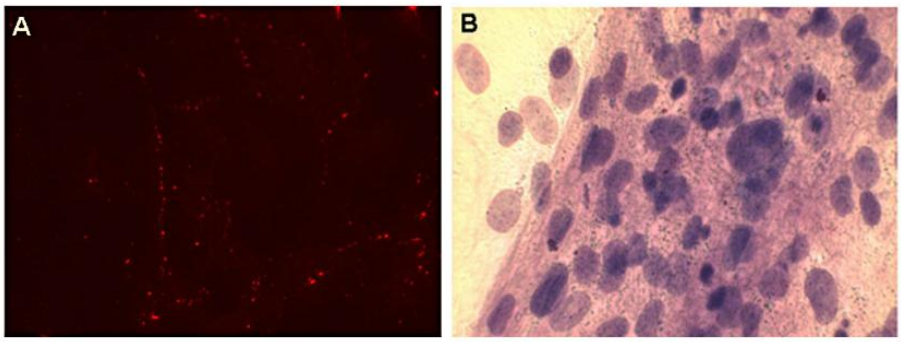


Figure 4.11. Levels of Cx43 fluorescence and morphological features of cadmium transformed cells.
Cell transformation assay was carried out as described in paragraph 2.6 and 2.7. After fixing, *foci* were isolated by pap-pen tracing and stained for indirect immunofluorescence with an anti-Cx43 primary antibody. Images acquired at 400x magnification show lined up, intensely fluorescent dots (panel A). Petri dishes were then rinsed, Giemsa stained and *foci* scored as per standard criteria **(Landolph, 1985)**. A cadmium induced *focus* detail at 400x magnification is shown in panel B.

For clarity's sake, a set of pictures obtained from a 1 μM cadmium transformed *focus* is present in **Figure 4.12**. While classification of *foci* is usually performed at 25x magnification **(Figure 4.12, panel A)**, higher magnification images **(Figure 4.12, panels B and C)** are useful to compare the microscopic features of transformed *foci* with the fluorescence images.

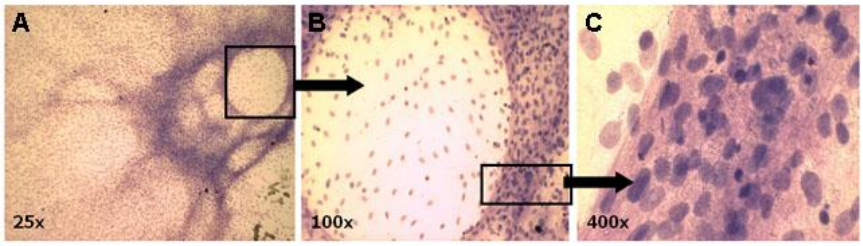


Figure 4.12. Type III *focus* features at higher magnification levels.
A 1 µM cadmium transformed *focus* is shown in panel A at 25x magnification. Presence of multilayering, deep basophilic staining and invasiveness identify the *focus* as a Type III. Magnification levels of 100x (panel B) and 400x (panel C) show the presence of vortexes and spindle shaped morphology, with cells with flattened nuclei in comparison with the normal cells inside the vortex (panel C).

In fact, pictures of the same areas and with the same magnification of fluorescence images were taken and superimposed, in order to study the reciprocal localization of fluorescent markers and morphological features characteristic of transformation. Transformed *foci* showed a correspondence between invasive structures and Cx43 altered localization **(Figure 4.13, panels A-C)**. It is noteworthy that *foci* images used for co-localization purposes were taken at 400x magnification, instead of the usual 25x used for scoring.

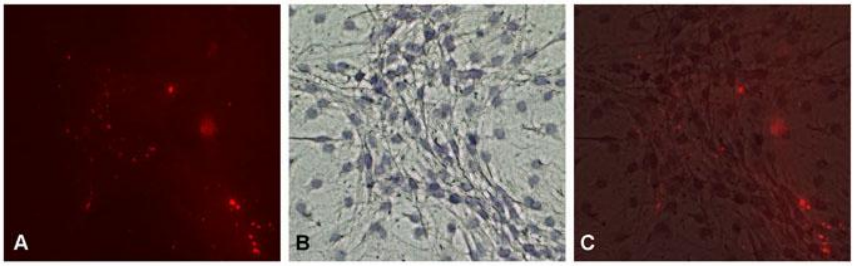


Figure 4.13. Co-localization of fluorescent marker Cx43 and morphological features of transformed cells.
Cell transformation assay was carried out as described in paragraph 2.6 and 2.7. After fixing, *foci* were isolated by pap pen tracing and stained for indirect immunofluorescence with an anti-Cx43 primary antibody. Images were acquired at 400x magnification (panel A). Petri dishes were then rinsed and Giemsa stained. Pictures of the same areas were taken with the same magnification (panel B). Corresponding images were then manually tested for superimposition (panel C). Images were slightly edited for a better visibility.

A comparison of normal cells with cell transformed either with 1 µM cadmium or with 1 µg/ml B[a]P was performed, revealing a difference in the localization of Cx43. In fact, while both transformed foci showed focal intense, B[a]P transformed *foci* lacked the lining up found in cadmium transformed ones **(Figure 4.14, panels A-C)**.

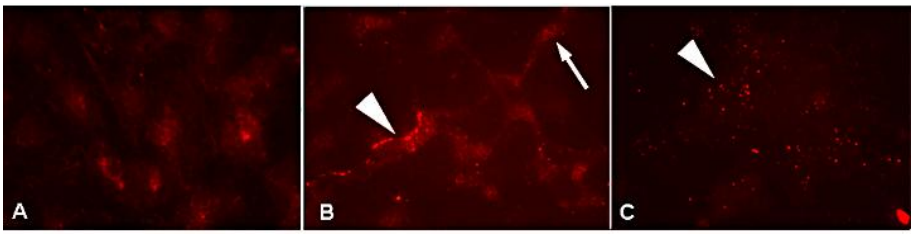


Figure 4.14. *In-focus* immunofluorescence localization of Cx43 in 2-stage CTA
C3H cells were treated with1 µM $CdCl_2$ for 24. Positive control was 1 µg/ml B[a]P. Negative control was medium only. Recovery period lasted for 6 weeks. Four days after the beginning of recovery, the addition of 0.1 µg/ml TPA started and lasted for all the duration of the cell transformation assay. Panel A: control treated cells (normal monolayer) show a faint distribution of Cx43 throughout all the cell. Panel B: 1 µM cadmium treated cells (later recognized as a transformed *focus* after Giemsa staining) show intensely stained dots placed in lines along the cellular membrane (arrowhead). Nuclear dots present in the 24 hours treatment are still present (arrow). Panel C: B[a]P treated cells (later recognized as a transformed *focus* after Giemsa staining) show intensely stained dots (arrowheads) but with no particular association with the cellular membrane. Diffuse fluorescence is also less prominent. All images were acquired at 400x magnification. Images were slightly edited for a better visibility.

Summarizing, no alteration of Cx43 localization was found in normal, monolayer cells, with no changes between medium only and cadmium treated dishes. On the other hand, *in-focus* studies performed on cells transformed with different chemicals (cadmium and B[a]P) showed alterations in Cx43 localization in comparison with normal cells, but with different patterns of localization.

4.6 *Foci*-derived Cell Lines Grow in *Focus-like* Patterns

Cells were scraped from transformed colonies from CTA dishes in order to obtain a stable transformed line (hereon, Re-Seed line) for biochemical and genetic characterization. Control Re-Seed lines were

derived from normal, contact-inhibited cells, both from negative control and from TPA treated group. Treated Re-Seed cell lines were obtained from *foci* obtained from both positive control (B[a]P induced Type III *foci*) and from $CdCl_2$ + TPA treatment group. From this latter group, lines were established both from a fully transformed, invasive Type III *focus* and from intermediate *foci*. For each Re-Seed line obtained from a colony, another control Re-Seed line was established from the normal, contact-inhibited monolayer of the same dish **(Figure 4.12, panels A and B)**.

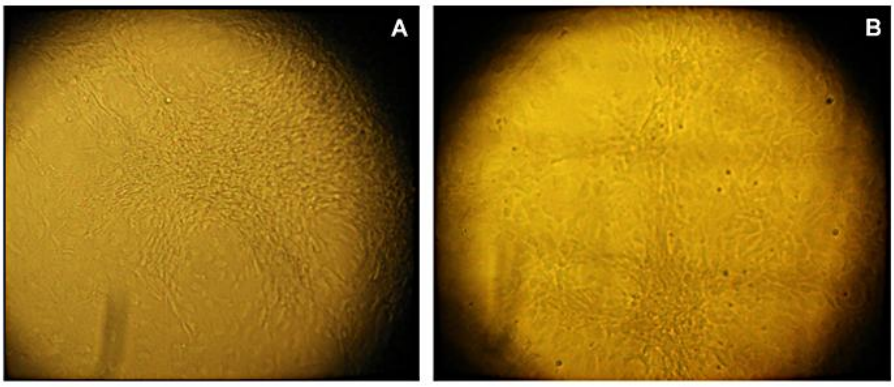


Figure 4.15. Re-Seed cell lines keep the same growth patterns of *foci.*
Cell transformation assay was performed as described in paragraph 2.6. C3H cells treatment with 1 µg/ml B[a]P induced formation of Type III *foci.* Colonies were photographed through the microscope binoculars with a normal compact digital camera before harvesting (panel A). Obtained Re-Seed colony showed the same growth pattern of the originating *focus* when seeded at low density (panel B).

Re-Seed lines obtained from *foci* grew faster than Re-Seed lines obtained by both control and treated monolayer cells, being also faster when obtained from aggressive types than from less invasive *foci* types. When cultured at low density, transformed Re-Seed lines show a *focus*-

like growth pattern characterized by growth in multilayer and invasiveness. Re-Seed lines obtained from normal monolayer cells showed the typical features of normal cells like contact inhibition and growth on monolayer.

4.7 Cadmium Treatment Induces Changes in RNA Profiling in a Hepatoblastoma Cell Line

HepG2 hepatoblastoma cells line was used to investigate the effects of cadmium treatment at a molecular level. These cells still expresses most of their hepatic heritage, thus representing a good model for the assessment of liver-related response.

HepG2 cells were divided in three treatment groups: Negative control (medium only), 2 µM $CdCl_2$ and 10 µM $CdCl_2$. Principal components analysis (PCA) revealed a strong effect in stimulating gene expression after 24 hours of 10 µM $CdCl_2$, while only a small difference was present between 2 µM $CdCl_2$ and negative control RNA expression **(Figure 4.13)**. The first principal component highlighted a possible dose-response effect in cadmium treatment, with a progressive response ranging from 2 to 10 µM.

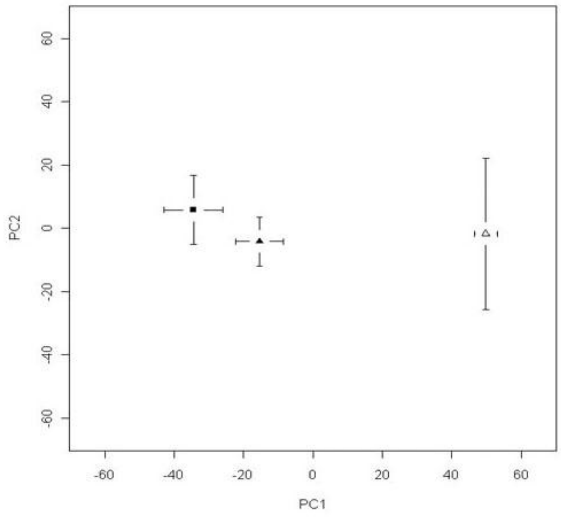


Figure 4.16. Principal component analysis of the HepG2 transcriptome.
HepG2 cells were treated with 2 µM $CdCl_2$, 10 µM $CdCl_2$ or medium only (control) for 24 hours. Cell lysates were analyzed for total RNA transcription trends. PCA resulted in a strong resemblance between control and low cadmium, while a strong difference with high cadmium induction was found. Bars represent the SD within the samples. The first component (PC1) accounts for 33.8% and the second (PC2) for the 25.2% of the total variance. Black triangle represents 2 µM $CdCl_2$, white triangle 10 µM $CdCl_2$ and black square control.

Low cadmium concentration induced mostly genes belonging to the metallothionein family. High cadmium concentration strongly inhibited genes involved in pathways involved in the typical of the liver activities, fatty acid metabolism and complement cascade, while it enhanced genes involved in the pathways of cancer (most notably, focal adhesion) and inflammation. The alteration of the expression levels of a subset of the regulated genes was confirmed by RT-PCR.

The involvement of miRNAs in the regulation of many vital biological processes including differentiation, proliferation and apoptosis is supported by a growing body of experimental evidences. miRNAs are small, non coding regulatory RNA molecules that bind target mRNA, thus suppressing their translation into proteins on inducing its degradation. Alterations in miRNAs function and regulation have emerged as key factors in cancer. Expression patterns of miRNA were analyzed to better understand the molecular mechanisms of cadmium effects. HepG2 cells were treated with 10µM $CdCl_2$ for 24 hours. Negative control was represented by cells grown in culture medium only. Most of the miRNA whose levels were resulted regulated from cadmium treatment belonged to the *let-7* family, which is known to have a role as tumor suppressor **(Fabbri et Al., submitted)**.

Chapter 5: *in silico* results

5.1 Morphologically Defined Data Subset is Efficient in Training in a Hierarchical Classification Approach

An algorithm for automatic classification with a hierarchical approach and of the unsupervised type was developed to divide images belonging to transformed and non transformed *foci* types in order to give decisional support to the operator in borderline cases, thus minimizing the possible subjectivity driven under or over estimation of results.

For clarity's sake, **Figure 3.1** will be reported as a reminder of the hierarchical architecture.

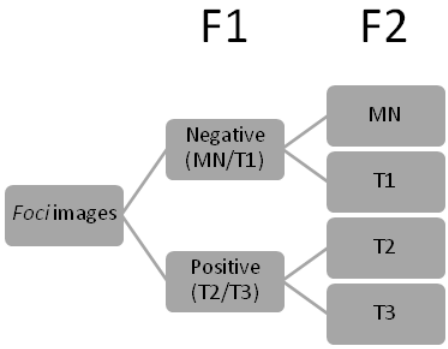


Figure 3.1 (repeated). *Two-step hierarchical* classification scheme.
A schematic of the Two-step hierarchical classification is shown. A dataset of *foci* images is divided by the algorithm in two classes during F1, corresponding to positive (T2/T3) and negative (MN/T1). During F2, images classified as negative are divided between normal monolayer (MN) and Type I class *foci* (T1). Another classification run divides images classified as positive in Type II (T2) and Type III (T3) classes. The approach is two-step since it's based on two different classification runs and is hierarchical because the results of the first run are used as database for the second run.

A subset of tiles composed of 196 tiles obtained from 82 parent images of defined morphology (hereon MD) was clustered by PAM algorithm with a *k=2* parameter set. The *k* parameter is used by the algorithm to know the number of classes in which divide a data set. F1 step of classification resulted in classes C1 and C2, corresponding respectively to non transformed and transformed biological classes (negative and positive, **Figure 3.1, F1**), and class attribution was scored against non-transformed/transformed biological classes. The classifier-expert level of agreement achieved in this step was of 93.37%, with a sensitivity of 95.70% and a specificity of 91.26% **(Table 5.1, panel A)**.

C1 and C2 were clustered independently, each with another classification run, in two classes. Therefore with a *k=2* parameter was used during F2 step of classification, to a final number of four clusters. C1 (negative tiles) were divided in C'1 and C'2 (MN and Type I respectively, both non transformed), C2 (positive tiles) in C'3 and C'4 (Type II and Type III, both transformed) **(Figure 3.1, F2)**. Tables were unified in a comprehensive confusion matrix resulting in 80.61% of agreement, 80.71% of sensitivity, and 93.54% of specificity **(Table 5.1, panel B)**.

Majority of voting criterion was applied and the classification scores of parent images inferred for each step of classification. For example, if 4 tiles were obtained from a single parent images, and 3 were classified as

Type III and 1 as Type II, the assigned class for the parent image would have been Type III. F1 resulted in agreement 85.37%, sensitivity 78.38% and specificity 91.11% **(Table 5.1, panel C)**; while F2 resulted in agreement 80.49%, sensitivity 83.51% and specificity 93.98% **(Table 5.1, panel D)**. It is noteworthy that the 2 × 2 confusion matrix produced in F1 for parent images shows that only one Type III derived image tile had been scored as negative. Moreover, all of the Type II misclassified images were correctly classified during step F2, and no Type III image was classified as NM or T1 **(Table 5.1, panel D)**.

A

F1 : 512 x 512		
Exp Class	Aut Class	
	MN/T1	T2/T3
MN/T1	89	4
T2/T3	9	94

B

F2 : 512 X 512				
Exp Class	Aut Class			
	MN	T1	T2	T3
MN	50	4	0	0
T1	1	34	4	0
T2	1	7	42	4
T3	0	1	16	23

C

F1 : 1024 x 1024		
Exp Class	Aut Class	
	MN/T1	T2/T3
MN/T1	29	8
T2/T3	4	41

D

F2 : 1024 X 1024				
Exp Class	Aut Class			
	MN	T1	T2	T3
MN	16	0	0	0
T1	1	16	0	0
T2	0	7	15	8
T3	0	0	0	19

Table 5.1. Confusion matrices for morphologically defined *foci* (MD) classification with hierarchical approach (*k*=2).

Classification of the MD dataset (192 tiles from 82 parent images) was carried out in two independent classification runs, each dividing the dataset in 2 clusters (therefore, each step with *k=2*).

Panel A: F1 step of tiles classification. Panel B: F2 step of tiles classification. Panel C: F1 step of parent images classification. Panel D: F2 step of parent images classification. See text for details.

5.2 Morphologically Defined Data Subset is Less Efficient in a Single Step Classification Approach Rather Than in Hierarchical Approach for Training Step

In order to test the algorithm on a different architecture, a single step approach was performed. The MD subset was divided directly in 4 classes, corresponding to the four biological classes of normal monolayer and Type I, Type II and Type III *foci*. Therefore, the parameter set for this single classification run was *k=4.* The classifier was kept of the unsupervised type. The classification was carried out generating a 4 x 4 matrix, equivalent to the F2 comprehensive matrix **(Table 5.2, panel A)**. An aggregated 2 x 2 matrix was generated from the 4 x 4 by uniting C'1 and C'2 classes into C1 (MN and Type I into negatives), while C'3 and C'4 were united into C2 (Type II and Type III into positives), as by F1 obtained matrices **(Table 5.2, panel B)**.

Levels of achieved agreement are: 93.37% for the 2 x 2 aggregated matrix and 75.51% for the 4 x 4 matrix.

A

Confusion Matrix				
Exp Class	Aut Class			
	MN	T1	T2	T3
MN	46	8	0	0
T1	6	23	2	8
T2	0	2	35	17
T3	1	0	4	44

B

Aggregated Matrix		
Exp Class	Aut Class	
	MN/T1	T2/T3
MN/T1	83	10
T2/T3	3	100

Table 5.2. Confusion matrix for MD classification with one-step approach (*k*=1).
Classification of the MD dataset (192 tiles from 82 parent images) was carried out in a single classification run, dividing the dataset in 4 clusters (therefore, a single step with *k=4*).
Panel A: confusion matrix for tiles classification. Panel B: aggregated matrix obtained by joining MN+T1 and T2+T3 classification results.

5.3 Effects of Database Noise and Dimension of Dataset on Classification

Robustness of the method was tested on the MD+MU dataset: this dataset was generated by joining the MD subset, composed of 196 images derived by 82 parent images, and the MU subset, consisting in 52 tiles derived from 20 parent images. This experiment was performed in order to assess the impact of a less refined training database on the quality of classification. This step is important in the perspective of unburdening the operator with the selection of MD and MU images. Agreement was evaluated only on MD images in order to avoid that the uncertainty in scoring the border line *foci* might result in a faulty

assessment of the classifier performances. Results showed a slight improvement in comparison with the results obtained in paragraph 5.1. In detail, a hierarchical approach was exploited, with two classification runs, both with *k*=2 parameter, as detailed in paragraph 5.1. Classification step F1 **(Figure 3.1, F1)** resulted in agreement 88.78%, sensitivity 97.85% and specificity 80.58% **(Table 5.3, panel A)**. Results achieved in F2 **(Figure 3.1, F2)** are: agreement 81.63%, sensitivity 82.42% and specificity 94.01% **(Table 5.3, panel B)**. The overall increment of classifier performances indicate that a larger dataset is more efficient than an smaller but less noisy one. Graphical representation of classification steps were made by exploiting a PCA approach **(**an example is showed in **Figure 5.1)**.

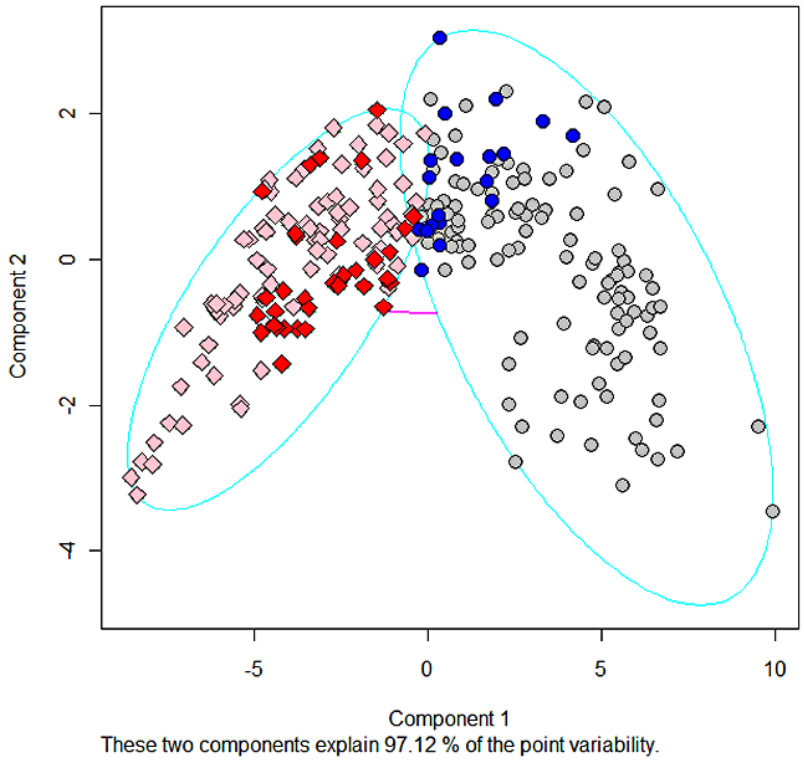


Figure 5.1. Statistical clustering of *foci* tiles by means of the PAM algorithm.
PCA graphic showing F1 step of classification of 512 x 512 tiles for MD+MU dataset. In this step of classification, positive (transformed) and negative (non-transformed) images are divided (*k=2*). The first two PCs were used for graphical representation, accounting for 97.12% of the point variability). Filled circles represent negative/untransformed images, while diamonds represent positive/transformed. Light colors stand for tiles belonging to the MD dataset. Deep colors correspond to images from the MU set.

Classification of parent images was performed after both F1 and F2 steps exploiting the majority of voting criterion. During F1, tiles derived from eight different parent images were evenly split between C1 and C2. Score of the parent image was assigned following a non-conservative approach: parent images with 2 tiles scored as negative (C1) and 2 tiles as positive (C2) were considered positive (transformed). Agreement

reached a level of 91.46%, sensitivity of 94.59% and specificity of 88.89% **(Table 5.3, panel C)**. In F2, 14 tiles derived from different parent images were divided between different classes. The same higher damage class criterion was applied for assessment of parent image scores. Thus, if a parent image resulted in one tiles of Type II, one tile of Type I and one tile of MN, the final score would have been Type II. Obtained values for step F2 were: agreement 86.58%, sensitivity 87.72% and specificity 95.48% **(Table 5.3, panel D)**.

A

F1 : 512 x 512		
Exp Class	Aut Class	
	MN/T1	T2/T3
MN/T1	91	2
T2/T3	20	83
MU	19	33

B

F2 : 512 X 512				
Exp Class	Aut Class			
	MN	T1	T2	T3
MN	52	2	0	0
T1	1	36	1	1
T2	2	14	31	7
T3	2	2	4	41
MU	2	17	31	2

C

F1 : 1024 x 1024		
Exp Class	Aut Class	
	MN/T1	T2/T3
MN/T1	32	2
T2/T3	5	40

D

F2 : 1024 X 1024				
Exp Class	Aut Class			
	MN	T1	T2	T3
MN	17	0	0	0
T1	0	18	1	1
T2	0	5	14	4
T3	0	0	0	22

Table 5.3. Confusion matrices for MD+MU classification with hierarchical approach (*k*=2).

Classification of the MD dataset (192 tiles from 82 parent images) plus the MU dataset (52 tiles from 20 parent images) was carried out in two independent classification runs, each dividing the dataset in 2 clusters (therefore, each step with *k=2*). Agreement of classification was evaluated only for MD tiles.

Panel A: F1 step of tiles classification. Lower part reports MU tiles scorings. Panel B: F2 step of tiles classification. Lower part reports MU tiles scorings. Panel C: F1 step of parent images classification. Panel D: F2 step of parent images classification.

5.4 Exploiting the QIoD for the Classification of Intermediate and Mixed *Foci*

A quantitative method for classification of the uncertain MU images was inferred from the MD images classification. Centers of the four classes obtained in paragraph 5.3 were calculated as reference mean tiles from the values of the descriptors of actual tiles. Distances between the 52 MU tiles were calculated and a mean value of the distances over the derived tiles for each parent image was calculated. The sum of the distances of each parent image to all the four centers was set to 1, by definition. The obtained index is a quantitative measurement related to the distance of a mixed or intermediate *focus* from the four considered classes (normal monolayer and the three *foci* classes).

MU parent images were then indexed from 1 to 20 and a comprehensive bar-chart plotted. All indexes obtained have shown an maximum level of dissimilarity (above 0.5) towards normal monolayer **(Figure 5.2)**.

The original implementation of the whole algorithm was developed in the S-language for the R environment, and an example of its application on the MD+MU dataset is available in **Annex I**.

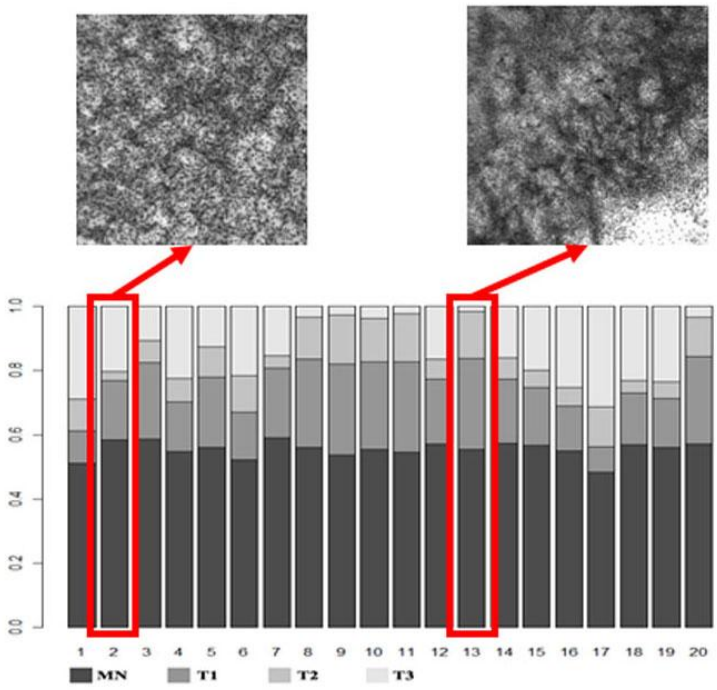


Figure 5.2. Bar-chart graphic of indices of dissimilarity.
The indexes of dissimilarity were calculated on the mean distances over generated tiles for each of the 20 parent images belonging to the MU dataset. All bars are scaled to sum up to 1. Large rectangles corresponds to a high dissimilarity to quintessential images of a given class, small rectangles to low dissimilarity. Legend shows the correspondence between gray levels and biological classes. Two MU images are shown in the top half, each representing the parent image with the smallest dissimilarity to a class. Noteworthy, all 20 images showed the maximum dissimilarity to normal monolayer (MN) **(Procaccianti et Al., 2011)**.

5.5 Generation of a *Foci* Images Database

Parent images were acquired as described (paragraph 3.1). The acquisition routine was modified and tested on different combination of time of exposure and illumination. An exposure of 0.593 milliseconds with the minimum possible illumination were found to be the best pair for

homogeneity of illumination between images. Though 0.017 milliseconds with an illumination of 3 on the microscope knob gave a better background color, the illumination level between tiles was too unstable due to the low time exposure (data not shown).

Images were always taken in the same order. An imaginary square circling the *focus* of interest was traced. Acquisition started from top-left corner of the square, moving down by row first, then by column. Images were shot with a partial overlapping of about 40-50%. Colonies growing close to the border of dishes were excluded. The majority of colonies were fully covered with a number of shots between 9 and 25. Extraordinarily big transformed *foci* (usually Type III) needed more than 70 shots to be fully covered. Images were processed with the Photomerge routine in Adobe Photoshop CS5, in batches of maximum 20 images in order not to overload the program. In case more batches were needed for a single *focus*, partial merging images were saved and then used for a hierarchical merging approach **(Figure 5.3)**.

Eighty-eight petri dishes were acquired, corresponding to more than 300 *foci* super-images and over 5’000 parent images

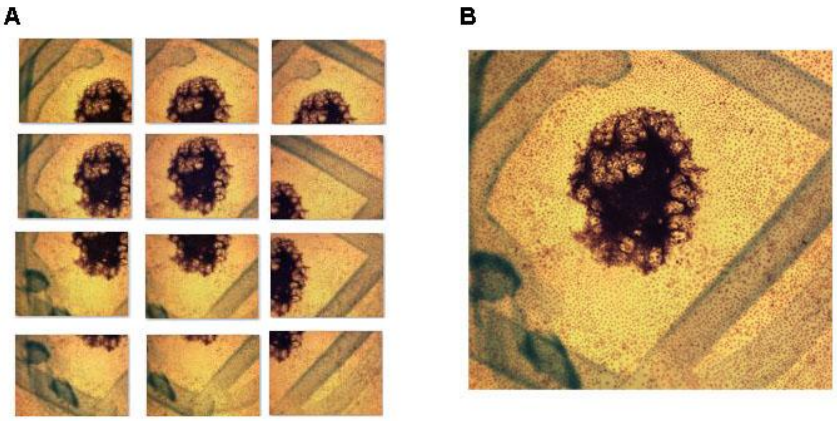


Figure 5.3. Generation of *focus* super-image.
1239x1040 images were taken following a grid-like pattern until the whole area of the *focus* was covered (panel A, each thumbnail representing a single parent image). Overlapping was intentional. Merging of pictures resulted in a super-image including the whole *focus* with the same resolution of details of the single parent images.

Chapter 6: Discussion

Carcinogenesis is a genetic disease with a complex background. Various independent genetic events must take place for a cell to undergo tumoral transformation, leading to the inactivation of genes involved in the regulation of the cell cycle. Initiated cells will then proliferate uncontrollably and accumulate mutations, further progressing in the neoplastic transformation. Other hallmarks of carcinogenesis than genes involved in cell survival, are the inactivation of the DNA damage repair system and the inhibition of the pathways involved in cell cycle block and apoptosis pathways **(Voegelstein et Al., 2002)**.

Chemical induction of carcinogenesis is referred to as the potential of a substance to induce the formation of tumors or to increase their frequency of appearance or malignancy **(Maurici, 2005)**. Mechanisms involved in this process are classified as genotoxic, when direct DNA damage is involved, or non-genotoxic, when the substance does not show a potential for direct interaction but rather acts by means of epigenetic mechanisms such as reactive oxygen species (ROS) generation or by interfering with damage repair systems and regulation of transcription. At the state of the art, the only assay validated for the evaluation of carcinogenicity at regulatory level is the two year bioassay to be performed in rodents **(OECD, 2007)**. This test is heavily demanding both in terms of monetary and time related costs: it has been estimated that by 1998 only the 2.7% of chemicals were tested for their

carcinogenetic potential, summing up to millions of dollars of expanses and millions of animal lives **(Cohen, 2001; Goodman 2001; Knight et Al., 2006).** Though the *in vivo* bioassay is the only test that reached an international level of validation, it bears some characteristics that makes it differ from an ideal assay. First, there is a lot of scientific debate on the reliability of results obtained in rodent models when transported to human pathogenesis. In fact, studies have demonstrated a low efficiency in spotting human non-carcinogenic substances **(Knight et Al., 2006)**. In some cases, substances spotted as carcinogens in murine models lost their potential when transferred to human findings **(IARC 1972; Ennever et Al., 1987; Gold et Al., 1998; Haseman 2000)**. Moreover, due to all the stress induced by handling and testing, often painful and maybe ethically questionable, animals might experience changes in their basal hormonal levels, that might in turn alter the response of the animal on the chemical. Based on these evidences, IARC took a policy of not accepting as definite human carcinogens substances that only offer experimental data acquired in animals **(Knight et Al., 2006)**.

Due to the ethic and scientific issues regarding the *in vivo* testing, the scientific community has developed a strong interest around *in vitro* techniques. The philosophy beyond this research can be resumed with the “3R” acronym, which stand for *replacement, reduction* and *refinement* **(Russel et Al., 1959)**. Though no *in vitro* assay is available for a complete replacement of the animal models in carcinogenesis

studies, reduction and refinement are possible on the basis of careful planning, of *in vitro* and *in silico* preliminary tests and by the use of transgenic models **(Knight et Al., 2006)**

The promulgation of the REACH directive by European Commission in 2006 has set a strong need for large scale methods for assessing the carcinogenic potential of compounds **(REACH EC 1907/2006)**. Upon the enforcement of this directive, in fact, all the chemicals available and all the new ones need to be tested for their carcinogenic potential prior commercialization. Due to its costs, the *in vivo* bioassay is not suited for the screening of such a huge amount of substances. The REACH directive also prescribes as mandatory the *in vitro* assays for screening purposes and, only if preliminary results are positive, the use of *in vivo* assays is allowed. Moreover, the European directive strongly emphasizes the importance of developing and perfecting *in vitro* techniques.

The cell transformation assay (CTA) is a fast and reliable tool for the assessment of the carcinogenic potential of chemicals. The CTA has been proposed as a model for studying human carcinogenicity and is currently used by both cosmetic and pharmaceutical industries for screening purposes, by regional agencies for environmental protection (ARPAs) and for academic purposes (**Combes et Al., 1999)**. Various non-animal tests are validated for the study of genotoxic potential of chemicals, both in prokaryotic and eukaryotic (e.g. Ames test,

micronuclei). On the other hand, no *in vitro* assay has so far been validated for the study of non-genotoxic carcinogenic potential. The CTA is in the most advanced state of validation showing a potential for the assessment of epigenetic induction of carcinogenesis.

The *in vitro* transformation process is closely related to *in vivo* transformation and the insurgence of tumors in the clinical practice, all of them sharing a multi-step nature. Initiation, progression and metastasis clinical phases have counterparts both in the *in vivo* (initiation by chemicals with the acquisition of growth advantages, clonal expansion, mass formation and progression induced by non-mutagen TPA, fixation of mutations, formation of invasive tumors) and in the *in vitro* tests (morphological alterations, growth on multilayer, invasiveness) **(Kehsava, 2000; Yokota et Al., 1993)**.

Cell transformation assay relies on the use of particular cell lines, characterized with a strong sensibility to carcinogenic compounds and a low rate of spontaneous mutations. The CTAs in the most advanced state of validation are the SHE, the BALB/c and the C3H based assays **(Knight et Al., 2006; OECD, 2007)**.

While the SHE based is the most advanced between the three, it is best suited for studying early steps of transformation. BALB/c and C3H transformation assay are instead suited for studying the late carcinogenesis. CTA protocols prescribes a low-density seeding of cells followed by the treatment with the chemical to be analyzed for a short

period (24-48 hours). The endpoint of the assay is the formation of colonies (*foci*) against the normal monolayer. As for the study of promoting/protective activity of non-genotoxic compounds, it exploits the comparison with a known promoter (TPA) on cells primed with sub-toxic concentration of a known carcinogen (MCA) **(OECD, 2007)**. The scoring of colonies obtained in a CTA is based on morphological features. While SHE and BALB/c based assays only have outcomes in terms of positive or negative, the C3H based features a more complex evaluation of results. Obtained colonies are scored not only as transformed or non transformed, but also divided in damage classes: Type I, Type II and Type III *foci* are in fact distinguished on the basis of growth pattern, morphology of the cells, and invasion of the surrounding monolayer **(Landolph, 1985)**. The transformed status of Type II and Type III *foci* was confirmed by growth on soft agar and xenograft formation in a percentage of cases between 50 and 85%, with the generation of 100% sacromas **(Keshava, 2000)**. While the canonical criteria for classification prescribe the assignment of doubtful *foci* to the higher or lower damage class depending on the cases **(Landolph, 1985)**.

The international validation of the CTA has been slowed by some critical aspects, namely the subjectivity in the evaluation of results, especially in border line cases, and the reproducibility of results. Cross-laboratories studies for international reproducibility have been promoted by ECVAM and OECD **(Combes et Al., 1999)**. The development of a novel method

to provide a quantitative, objective tool for decisional support to the operator in order to reduce to a minimum the possible subjectivity driven error in the evaluation of morphologic structures was object of this work. Given the strong interest of both the academic community and industries as well after the promulgation of the REACH directive, a system that can help to standardize the interpretation of results is of high priority. To achieve this result, two different approaches were exploited: the first based on a biochemical approach in order to study the co-localization of biological markers of transformation with the morphological structures needed for the scoring; the second exploiting elements of image analysis in order to extract quantitative information from qualitative descriptors. The biological approach used as experimental model C3H cells and as test chemical cadmium. Cadmium is a heavy metal and a widespread environmental pollutant. It has a good bioavailability and a long half-life (20-30 years). It forms deposits in various organs, especially kidneys and liver. It has been declared as human carcinogen with effect on kidney, lungs, liver, bladder, prostate and breasts by IARC, but its mechanisms are still not fully understood yet **(IARC, 1993)**. It shows no direct potential of DNA damage but it seems to act through various epigenetic mechanisms (generation of ROS by inhibition of scavenging pathways, inhibition of DNA damage repair systems, alteration of cell cycle controls). Part of its effects might be due to its affinity with zinc and the derived possibility of interaction with zinc-finger proteins. Cadmium in fact shows

high affinity towards thiolic groups present in the cystein residues of zinc finger motives. The *gatekeeper* p53 DNA binding domain is rich in cystein residues: the binding of Cd2+ instead of Zn2+ induces a conformational change that renders the wild type protein inactive **(Méplan et Al., 1999)**. Cadmium induces the synthesis of metallothionens (MT) through the direct interaction with the MTF1 factor. Low concentration of this metal, therefore, create the conditions for the tolerance due to MT production and, due to the long half-life, for the accumulation and the reaching of toxic levels **(Hartwig, 2010)**

The concentrations of the metal used in the experiments were set on the basis of previous finding of the working group **(Bussinelli, 2007)** and confirmed with MTT assay. In fact, results from the survival experiments **(Figure 4.1, panel A)** shown a good viability rates of cells upon treatment with cadmium salts at a concentration of 1 µM, well below the EC_{50} levels, coherently with the previous findings. While these findings may seem in contrast with the results obtained in the cytotoxicity assays preliminary to CTA, it is noteworthy that only in one case cadmium treatment resulted in less than 50% of survival **(Table 4.1)**. Moreover, dishes belonging to that treatment group showed a fair number of colonies between 40 and 45 cells each (data not shown). Cell transformation assay heavily relies on the quality of serum batches, so much that the plating efficiency is the first parameter evaluated in order to accept or refuse a stock. In all experiments, plating efficiency was

above the threshold levels. It is important to note how this dependence might result in a strong limitation for the reproducibility of results, especially on an international, inter-laboratory level. The development of a synthetic analogue of serum may therefore take a role in the reproducibility studies of results **(IARC, 1985).**

While the results of MTT assay show that treatment with low concentration of cadmium for 48 hours resulted in more than 80% of survival, cytotoxicity results show levels only barely above 50%, and are only slightly higher in TPA treated cells **(Table 4.3)**. In contrast with previous findings **(Bussinelli, 2007)**, 2 µM $CdCl_2$ treatment always resulted in survival rates always below 50% **(Table 4.1 and 4.2)**. The toxic effects of 2 µM $CdCl_2$ **(Figure 4.6)** might on the other hand relate with the peak of HSP70 found in western blotting experiments at the same concentration **(Figure 4.2, panel B)**. Results of the *in vitro* transformation confirm the carcinogenic potential of cadmium. Its frequency of transformation seems slightly improved by TPA induction. On the other hand, long term cytotoxic effects of sub-toxic TPA were strongly enhanced by 2 µM $CdCl_2$ treatment. **(Tables 4.5 and 4.6)**. Though rules for mixed and intermediate *foci* scoring are set, following an approach both conservative and non-conservative depending on the cases, in this work a stricter non-conservative criterion was applied. In fact, in all doubtful cases, higher class of damage was chosen. Given the alarming increase of tumor incidence during the last decades, a stricter

policy of public health risk assessment might be advisable **(Procaccianti et Al., 2011)**.

Gap junctions have been reported to exhibit both tumor suppressive and tumor promoting effects. Such complex behavior seems to be due to different mechanisms, some related and some other unrelated to the actual intercellular communication levels. Inhibition of gap junction intercellular communication assay (GJIC) is common in transformed tissues and gap junctions are often found altered in human tumors. Since no direct relationship has been found between GJIC and genotoxicity but only between GJIC and transformation, these proteins seem a fitting model to study non-genotoxic compounds. The connexin 43 is the most ubiquitary expressed member of the connexin family, and its association with cancer, though with a complex role of both suppressant of transformation and promoter of metastasis in late stadies, is supported by a large body of experimental evidences **(Langlois et Al., 2010; Naus et Al., 2010)**. Localization of Cx43 was herewith selected as a test biological marker linked to transformation processes for the *in-focus* co-localization approach. The effectiveness of this marker in this experimental model has been beforehand tested by the results obtained after 24 hours of administration of cadmium salts at sub-toxic concentrations. In fact, cells showed a marked reduction in the levels of intercellular communication **(Figure 4.7)**, as well as a change in the cellular localization of the proteins **(Figure 4.8)**. Analysis of the variation

levels of Cx43 expression was performed with complex and contradictory results between replicas (data not shown). Though these findings might appear contradictory with the role of tumor suppressant of Cx43, survival pathways activated by gap junctions depending not on the function but on the association of Cx43 with other proteins have been reported in literature **(Langlois et Al., 2010).** Fluorescence findings pointed out marked differences in localization of Cx43 and an association with morphological features of transformed *foci*, thus confirming the validity of this marker for co-localization studies.

The standard protocol for the evaluation of *foci* is based on morphological criteria **(Landolph, 1985)**. The developed technique allows for a double study of both morphological and biochemical markers. This technique has a strong value for its flexibility. In fact, by simply replacing the antibody used for Cx43 with any other antibody, it is possible to study a virtually infinite number of markers, as long as the fixing protocol for the biochemical marker is compatible with the fixing protocol for *foci* scoring. Moreover, biochemical signals are much more specific and less subject to any kind of personal interpretation than morphological markers. The individuation of alterations in the presence or localization of proteins with a leading role in transformation processes might add a new, objective criterion for the assessment of *in vitro* carcinogenesis. This technique can be also used to study the molecular mechanisms of transformation. In fact, the possibility of studying not only

the short term alteration induced by chemicals but also the effects on the long run can lead to new insights on the molecular mechanisms underlying chemical induced transformation. The association of a biochemical marker with specific morphological structures of transformed colonies (e.g. the invading processes of Type III *foci*) might help in shedding some light on the mechanisms of invasion of tumors. In order to do so, though, it is necessary to develop a sharp system of coordinates applicable to petri dishes that could ensure the same placing between the fluorescence and the Giemsa stained acquisitions.

A good source of information might also derive from the isolation of cells lines derived from *foci* (Re-Seed). The fact that such lines exhibit *focus-like* growth pattern characterized with piling, criss-crossing and invasion of the surrounding space can be a proof of their *focus* heritage and indicative of a conserved genetic background. These clones could therefore be used for screening biological markers before testing on the more precious and rare samples of *foci*. In fact, having a number of *foci* suitable for testing markers could take years, due both to the low rate of induction of transformation by many chemicals and to time lapse of the transformation protocol itself.

Due to these characteristics, Re-seed lines may also harbor interesting results in toxicogenomic approaches. Tracing a molecular fingerprint of the action of a chemical on the whole genome produces a high throughput of information with a relatively affordable cost in term of

money and time spent on the analysis of data **(Dakeshita et Al., 2009; Ao et Al., 2010)**. Toxicogenomics studies can also be used as a candidate gene approach for the identification of biochemical markers to be used in co-localization studies. Microarray screening allows the study of numerous genes in the same moment, thus investigating the effects of carcinogens on whole expression profiles. Microarrays can also be used to assess the potential of non-genotoxic carcinogens, and its use, for how recent, seems promising **(Goodman, 2001; Kramer et Al., 2004; Ao et Al., 2010).** Anyway, the real challenge with such a massive amount of generated data is to analyze the results, converting the raw output data, composed of a huge flow of numbers, in an easily interpretable and form, searching for recognizable patterns **(Docterman et Al., 2002)**.

In this perspective, the effects of cadmium treatment were studied using as a model the HepG2 cell line. These are human low deviation hepatoblastoma cells and therefore an inherently interesting case due to the hepatotoxic effects of this metal. HepG2 cells are, in fact, highly differentiated and still maintain most of the function inherited from their hepatic origins. The results obtained can with a certain degree of approximation be translated onto C3H cells, taking into account the differences between the two models: while HepG2 are human cells of hepatic origin and with a transformed status, C3H cells are fibroblastic murine normal cells. Results obtained on the whole transcriptome upon

cadmium treatment show that low concentration differ from the basal status (negative control) mostly just by the induction of the MT genes **(Fabbri et Al., submitted)**. Metallothioneins are in fact known for being induced by cadmium and other metals and have the biological function of sequestering them **(Klaassen et Al., 2009; Urani et Al., 2007)**. High concentrations, on the other hand, differ for the expression profile of many genes, most of which are involved in the fatty acid transport, cholesterol metabolism, fatty acid oxidation as well as many substrates of fatty acid synthesis and biotransformation. Strongly affected by cadmium treatment resulted also the complement cascade, both of the classic and the intrinsic way. Enhanced levels of cancer related pathways (focal adhesion) were found, as well as stress level markers such as HSP70, coherently with the biochemical findings **(Figure 4.2 and 4.3) (Fabbri et Al., submitted)**.

Recent literature shows with a growing body of evidences the involvement of miRNAs in the regulation of many vital biological processes including differentiation, proliferation, apoptosis and cancer development. These small, non coding regulatory RNA molecules target mRNA, suppressing their translation or inducing their degradation.

The *let-7* miRNA family has been found altered in cadmium treated cells, and it has reported to be involved in antitumoral pathways **(Boyerians et Al., 2010)**. The strong effects of cadmium treatment on this genetic

product is a further proof of the carginogenic potential of this metal **(Fabbri et Al., submitted)**.

Another different approach to account for the subjectivity in *foci* classification exploited the use of image analysis techniques. By extracting quantitative information (image descriptors) from qualitative features (morphology), it might in fact be possible to develop a standard method offering decisional support to the expert. This is particularly important in the perspective of scoring the border line, morphologically undefined *foci*, where most of the scientific debate is focused. Disagreement between experts is a central issue hampering the process of international validation **(Combes et Al., 1999)**.

Classification algorithm are widely used in the biological sciences, with varying applications **(Ridder et Al., 1997; Crosta et Al., 2006; Urani et Al., 2009; Urani et Al., 2010).** Classifiers are used in order to score new, unknown samples on the basis of a well-consolidated database of reference. The realization of a standard database common for all the CTA users to be used as training material could be of great importance in this step, as it has been pointed out also by other Authors **(Poth, 2009)**.

In order to even only partially account for the problem of subjectivity in the scoring of *foci*, this work focused on the realization of an *unsupervised* classifier: thus, since no information generated by the expert is submitted to the algorithm, no operator-related bias is

introduced in the classification process. It is important to emphasize that the expert scoring was only used in order to evaluate the performances of the classifier, and never for any classification step. In fact, the process of selection of tiles is based on standard features and no evaluation of them is done, thus resembling a segmentation process. Other approaches for clustering were also tested, with inconclusive results. Co-correlation studies searching for recurring patterns in gray levels between neighborhood pixels were performed. *Foci* images tested with such method seem to correlate only on the basis of the illumination level (data not shown).

Due to the possible bias included even in the expert decision of which images fall in the canonic classes and which not, we tested both the "clean" dataset composed only of preselected Morphologically Defined images (MD) and the "noisy" dataset, including both the Morphologically Defined and Undefined images (MD+MU), and evaluated their results for training processes. Interestingly, the balance between the signal-to-noise ratio and the dimension of the sample leaned toward the latter, also proving how the presence of border line *foci* in the training step is well tolerated by the developed system **(Tables 5.1 and 5.3)**. The use of a reasonably large dataset of at least 100 images is advised, though a bigger dataset, as our result show, harbors an improvement in the performances, emphasizing again the importance of a network between

laboratories in order to create a common, universal database of *foci* images.

Another quantitative classifier based on C3H-derived *foci* images was presented in previous works **(Urani et Al., 2009)**, though bearing some major differences **(Urani et Al., 2010)**. First, the classifier was based on a supervised training step. Though it featured a very good rate of agreement between expert and classifier, the fact that it relied on expert provided information made the classifier subject to possible bias, thus not optimal for mixed and intermediate *foci* types. Second, it exploited the use of complex descriptors to be extracted by an equally complex algorithm. While this feature is of marginal relevance for image analysis experts, it made the classifier not suited for being used by personnel not extensively trained in computer sciences. Moreover, all the descriptors utilized in this work are independent from the image orientation. In the perspective of making this classifier accessible to all biologists working with *in vitro* methods, the simplicity of use is a must. Finally, much as both shared a training/validation architecture, only the classifier developed in this work has proved to work without any pre-selection of images.

Though other Authors proposed other systems for automatic classification **(Ridder et Al., 1997),** only this classifier approaches the problem of border line *foci* classification. The Index of Dissimilarity proposed by our group is a quantitative measure of the closeness of a

non-canonical *focus* and the canonical classes (Type I, Type II and Type III) **(Procaccianti et Al., 2011)**.

One of the most important critical points of this work lies in the fact that the morphological features used for *foci* classification are intended for whole *foci* evaluation, and not for tiles scoring **(Reznikoff et Al., 1973)**. Thus, the actual comparison of results is not entirely rigorous. The last part of this work was focused on the development of a protocol to acquire high resolution, whole *foci* super-images. In fact, while tile cropping of parent images is optimal for lessening the computational burden and for extracting microscopic descriptors, information regarding macroscopic structures might result crippled or lost in the process. The selection and extraction of image descriptors processes capable of capturing the macroscopic information still need study. The joining of the microscopic and macroscopic descriptors, together with the biological and molecular markers exploitation might be used as a basis for a multilevel, high-content analysis of *foci*, possibly depriving the set standard in use of its inherent subjectivity and perhaps in a future help setting a new, objective gold standard for classification.

Bibliography

Aaronson S.A., Todaro G.J. Development of 3T3-like lines from Balb-c mouse embryo cultures: transformation susceptibility to SV40**,** Journal of Cellular Physiology 72, 141-148 (1968).

Ashby J., Purchase I.F. Will all chemicals be carcinogenic to rodents when adequately evaluated?, Carcinogenesis 8:489–495 (1993).

Ao L., Liu J.I., Liu W.B., Gao L.H., Hu R., Fang Z.I., Zhen Z.X., Huang M.H., Yang M.S., Cao J. Comparison of gene expression profiles in Balb/c3T3 transformed foci exposed to tumor promoting agents, Toxicology In Vitro 24(2):430-438 (2010).

Bordese M., Alini W. Image processing and analysis (biOps Package) Version: 0.2.1.1, Repository: CRAN (2007).

Boyerinas B., Park S.M., Hau A., Murmann A.E., Peter M.E. The role of let-7 in cell differentiation and cancer, Endocrine Related Cancer 17(1):F19-F36 (2010).

Bussinelli L. Saggio di trasformazione neoplastica per studi di cancerogenesi: sviluppo di un classificatore morfologico, Bachelor degree thesis dissertation, University of Milano Bicocca (2007).

Cohen S.M. Alternative models for carcinogenicity testing: weight of evidence evaluations across models, Toxicologic Pathology 29(1):183–190 (2001).

Comber M.H., Walker J.D., Watts C., Hermens J. Quantitative structure-activity relationships for predicting potential ecological hazard

of organic chemicals for use in regulatory risk assessments, Environmental Toxicology and Chemistry 22:1822-1828 (2003).

Combes R., Balls M., Curren R., Fishbach M., Fusering N., Kirkland D., Lasne C., Landolph J., LeBoeuf R., Marquard H., McCormick J., Müller L., Rivedal E., Sabbioni E., Tanaka N., Vasseur P., Yamasaki H. Cell transformation assays as predictors of human carcinogenicity, Alternatives to Laboratory Animals 27:745-767 (1999).

Corvi R., Aardema M., Hayashi M., Hoffman S., Thomas B.C., Schechtman L., Vanparys P. Cell transformation assays: current status, 7th World Congress on Alternatives & Animal Use in the Life Science Rome 26:249 (2009).

Crosta G.F., Fernandes G.E. Feature extraction and classification of wide angle optical scattering patterns from single aerosol particles, In Progress in Electromagnetics Research Symposium Abstracts: 458-459. Cambridge, MA: The Electromagnetics Academy, 777 Concord Avenue, Suite 207 (2008).

Crosta G.F., Urani C., Fumarola L. Classifying structural alterations of the cytoskeleton by spectrum enhancement and descriptor fusion, Journal of Biomedical Optics 11: 024020_1-024020_18 (2006).

Dakeshita S., Kawai T., Uemura H., Hiyoshi M., Oguma E., Horiguchi H., Kayama F., Aoshima K., Shirahama S., Rokutan K., Arisawa K. Gene expression signatures in peripheral blood cells from Japanese

women exposed to environmental cadmium, Toxicology 257:25-32 (2009).

Deriche R. Using Canny's criteria to derive a recursively implemented optimal edge detector, The International Journal of Computer Vision, 1(2):167-187 (1987).

Docterman K.E., Smith S.M. Of mice and men: lessons from a microarray study of teratogen action, Teratology 66:217–223 (2002).

EFSA Panel on Contaminants in the Food Chain (CONTAM); Scientific Opinion on tolerable weekly intake for cadmium, EFSA Journal 9(2):1975. [19 pp.] (2011).

El-Fouly M.H., Trosko J.E., Chang C.C. Scrape-loading and dye transfer. A rapid and simple technique to study gap junctional intercellular communication, Experimental Cell Research 168(2):422-430 (1987).

Ennever F.K., Noonan T.J., Rosenkranz H.S. The predictivity of animal bioassays and short-term genotoxicity tests for carcinogenicity and non-carcinogenicity to humans, Mutagenesis 2(2):73–78 (1987).

Fabbri M., Urani C., Sacco M.G., Procaccianti C., Gribaldo L. Whole genome analysis and microRNAs regulation in HepG2 cells exposed to cadmium (Submitted)

Fitzgerald D.J., Murray A.W. A new intercellular communication assay: its use in studies on the mechanism of tumour promotion, Cell Biology International Reports 6(3):235-42 (1982).

Fumarola L., Urani C., Crosta G.F. Quantitative kinetics of damage and recovery of cytoskeletal structure by means of image analysis, Toxicology in Vitro 19(7): 935-941 (2005).

Gold L.S., Slone T.H., Ames B.N. What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database, Drug Metabolism Reviews 30(2):359–404 (1998).

Goodman J.I. A perspective on current and future uses of alternative models for carcinogenicity testing, Toxicologic Pathology 29(1):173–176 (2001).

Hartwig A. Mechanisms in cadmium-induced carcinogenicity: recent insights, Biometals 23(5):951-960 (2010).

Haseman K. Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk, Drug Metabolism Reviews 32(2):169–186 (2000).

Hirschi K.K., Rohovsky S.A., D'Amore P.A. PDGF, TGF-β, and heterotypic cell-cell interactions mediate endothelial cell-induced recruitment of 10t1/2 cells and their differentiation to a smooth muscle fate, The Journal of Cell Biology 141(3):805-814 (1998).

Hoffman E.P., Brown R.H. Jr, Kunkel L.M. Dystrophin: the protein product of the Duchenne muscular dystrophy locus, Cell. 51(6):919-928 (1987).

International Agency for Research on Cancer IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. IARC Scientific

Publications (1–55). Lyon, France: International Agency for Research on Cancer (1972–1992).

International Agency for Research on Cancer IARC Monographs on the Evaluation of Carcinogenic Risk to Humans: Berillium, Cadmium, Mercury and Exposures I the Glass Industry, IARC Scientific Publications (58):119–238 (1993)

IARC/NCI/EPA Working Group Cellular and molecular mechanisms of cell transformation and standardization of transformation assays of established cell lines for the prediction of carcinogenic chemicals: overview and recommended protocols, Cancer Research (45):2395-2399 (1985).

Jain A.K. Fundamentals of image processing, 569 pp. Prentice Hall Inc., New Jersey, USA (1989).

Jolliffe I.T. Principal Component Analysis, 487 pp. Springer-Verlang, New-York, NY, USA (1986).

Kaufman L., Rousseeuw P.J. Finding Groups in Data: An Introduction to Cluster Analysis, 368 pp. J. Wiley & Sons, New York, NY, USA (1990).

Keshava N. Tumorigenicity of morphologically distinct transformed foci induced by 3-methylcholanthrene in BALB/c-3T3 cells, Mutational Research 447(2):281-286 (2000).

Klaassen C.D., Liu J., Bhalchandra A.D. Metallothionein protection of cadmium toxicity, Toxicology Applied Pharmacology 238(3):215-220 (2009).

Knight A., Bailey J., Balcombe J. Animal Carcinogenicity Studies: 1. Poor Human Predictivity, Alternatives to Laboratory Animals 34(1):19–27 (2006).

Knight A., Bailey J., Balcombe J. Animal Carcinogenicity Studies: 2. Obstacles to Extrapolation of Data to Humans, Alternatives to Laboratory Animals 34(1):29-38 (2006).

Knight A., Bailey J., Balcombe J. Animal Carcinogenicity Studies: 3. Alternatives to the Bioassay, Alternatives to Laboratory Animals 34(1):39–48 (2006).

Knight A., Bailey J., Balcombe J. Animal Carcinogenicity Studies: Implications for the REACH System, Alternatives to Laboratory Animals 34:139–147 (2006).

Knudson A.G. Jr. Mutation and cancer: statistical study of retinoblastoma, Proceedings of the National Academy of Science of the U.S.A. 68(4):820–823 (1971).

Kramer J.A., Curtiss S.W., Kolaja K.L., Alden C.L., Blomme E.A., Curtiss W.C., Davila J.C., Jackson C.J., Bunch R.T. Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling, Chemical Research in Toxicology 17(4):463–470 (2004).

Laird D. Life cycle of connexins in health and disease, Biochemical Journal 394:527–543 (2006).

Landolph J.R. Chemical transformation in C3H10T1/2Cl8 mouse embryo fibroblasts: historical background, assessment of the

transformation assay, and evolution and optimisation of the transformation assay protocol. Transformation assay of established cell lines: mechanisms and applications, International Agency for Research on Cancer (IARC) Scientific Publications 67:185-201 (1985).

Langlois S., Cowan K.N., Shao Q., Cowan B.J., Laird D.W. The tumor-suppressive function of Connexin43 in keratinocytes is mediated in part v ia interaction with caveolin-1, Cancer Research 70(10):4222-4232 (2010).

LeBoeuf R.A., Kerckaert G.A., Aardema M.J., Gibson D.P. Multistage neoplastic transformation of Syrian hamster embryo cells cultured at pH 6.70, Cancer Research 50:3722-3729 (1990).

Levine A.J., Tomasini R., McKeon F.D., Mak T.W., Melino G. The p53 family: guardians of maternal reproduction, Nature Reviews Molecular Cell Biology 12(4):259-265 (2011).

Maechler M., based on S original by Peter Rousseeuw Cluster Analysis Extended Rousseeuw et al. (cluster Package) Version: 1.13.2 Repository: CRAN (2010)

Martelli A., Rousselet E., Dyke C., Bouron A., Moulis J.M. Cadmium toxicity in animal cells by interference with essential metals. Biochimie 88:1807-1814 (2006).

Mascolo M.G., Perdichizzi S., Rotondo F., Morandi E., Guerrini A., Silingardi P., Vaccari M., Grillic S., Colacci A. BALB/c 3T3 cell transformation assay for the prediction of carcinogenic potential of

chemicals and environmental mixtures, Toxicology in Vitro 24(4):1292–1300 (2010).

Matthews, E.J., Contrera, J.F. A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodents using enhanced MCASE QSAR-ES software, Regulatory Toxicology and Pharmacology 28(3):242–264 (1998).

Maurici D. Carcinogenicity, Alternatives to Laboratory Animals 33:177-182 (2005).

Méplan C., Mann K., Hainaut P. Cadmium induces conformational modifications of wild-type p53 and suppresses p53 response to DNA damage in cultured cells, The Journal of Biological Chemistry 274(44):31663-3170 (1999).

Mondal S., Heidelberger C. Inhibition of induced differentiation of C3H/10T1/2 clone 8 mouse embryo cells by tumor promoters, Cancer Research, 40, 334-338 (1980).

Naus C.C., Laird D.W. Implications and challenges of connexin connections to cancer, Nature Reviews Cancer. 10(6):435-41 (2010).

OECD Carcinogenicity Studies TG451. Available at www.oecd.org/dataoecd/ (1981, reviewed 2009)

OECD Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens, Series on testing and assessment 31:18 (2007).

Ozaki T., Nakagawara A. Role of p53 in Cell Death and Human Cancers, Cancers 3(1):994-1013 (2011).

Poth A. Cell transformation assay – past – present – future, 7th World Congress on Alternatives & Animal Use in the Life Science Rome 26:318 (2009).

Procaccianti C., Stefanini F.M., Urani C. The cell transformation assay: toward a statistical classification of mixed and intermediate foci images, Alternatives to Laboratory Animals 39(1):23-36 (2011).

Procaccianti C., Urani C., Stefanini F.M. The assessment of mixed and intermediate foci images using the R software environment, Working Papers del Dipartimento di Statistica "G. Parenti", Università degli Studi di Firenze, available at http://www.ds.unifi.it/ricerca/pubblicazioni/working_papers/ls_working_paper.php#2011 (2011).

R Development Core Team R: A language and environment for statistical computing. R Foundation for statistical Computing, Available at: http://www.R-project.org. Vienna, Austria (2009).

REACH EC 1907/2006 available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006R1907:EN:NOT (2006).

Reznikoff C.A., Bertram J.S., Brankow D.W., Heidelberger C. Quantitative and qualitative studies on chemical transformation of cloned

C3H mouse embryo cells, sensitive to postconfluence inhibition of cell division, Cancer Research 33:3239–3249 (1973).

Ridder G.M., Stuard S.B, Kerckaert G.A., Cody D.B., LeBoeuf R.A., Isfort R.J. Computerized image analysis of morphologically transformed and non transformed Syrian hamster embryo (SHE) cell colonies: application to objective SHE cell transformation assay scoring, Carcinogenesis 18:1965–1972 (1997).

Russell W.M.S., Burch R.L. The Principles of Humane Experimental Technique, 238 pp. Methuen, London, UK (1959).

Taylor S.M., Jones P.A. Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine, Cell. 17(4):771-9 (1979).

Urani C., Melchioretto P., Canevali C., Crosta G.F. Cytotoxicity and induction of protective mechanisms in HepG2 cells exposed to cadmium, Toxicology In Vitro 19(7):887-892 (2005).

Urani C., Melchioretto P., Canevali C., Morazzoni F., Gribaldo L. Metallothionein and hsp70 expression in HepG2 cells after prolonged cadmium exposure, Toxicology in Vitro 21(2):314–319 (2007).

Urani C., Stefanini F.M., Bussinelli L., Melchioretto P., Crosta G.F Image analysis and automatic classification of transformed foci, Journal of Microscopy 234(3):269-279 (2009).

Urani C., Crosta G.F., Procaccianti C., Melchioretto P., Stefanini F.M. Image classiers for the cell transformation assay: a progress report. Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues

VIII, Ed. by Farkas D. L., Nicolau D. V., Leif R. C. Proceedings of the SPIE 7568: 75681F-75681F-11 (2010).

Valko M., Rhodes C.J., Moncol J., Izakovic M., Mazur M. Free radicals, metals and antioxidants in oxidative stress-induced cancer, Chemico-Biological Interactions 160(1):1-40 (2006).

Voegelstein B., Kinzler K.W. The genetic basis of human cancer, 821 pp. Ed. McGraw-Hill Professional (2002).

Wuertz D., Rmetrics core team members Rmetrics - Markets and Basic Statistics (fBasics Package) Version: 2110.79 Repository: CRAN (2010).

Yokota J., Sugimura T. Multiple steps in carcinogenesis involving alterations of multiple tumor suppressor genes, The Journal of the Federation of American Society for Experimental Biology 7(10):920-925 (1993).

Zhang H., Borman H.D., Myhr B.C. Enhancement of the morphological transformation of Syrian hamster embryo (SHE) cells by reducing incubation time of the target cells. Mutation Research 548:1–7 (2004).

Index of Figures

Index of Tables

Annex I: Excerpt from "The assessment of mixed and intermediate *foci* images using the R software environment."

Procaccianti C.[1], Urani C.[1], Stefanini F.M.[2]

[1]*Dipartimento di Scienze dell'Ambiente e del Territorio (DISAT)*
Università degli Studi di Milano-Bicocca, Milano, Italia

[2]*Dipartimento di Statistica "G.Parenti"*
Università degli Studi di Firenze, Firenze, Italia

December 27, 2011

Abstract

Stand-alone, complete subset of the developed R code developed to calculate a quantitative dissimilarity index and classify mixed or intermediate *foci* by exploiting the quantitative information provided by digital images of *foci* colonies. Full paper is available at texttthttp://www.ds.unifi.it/ricerca/pubblicazioni/working_papers/2011/wp2011_02.pdf

Contents

1 Setting up the R environment

The packages needed for the descriptors extraction and images classification have to be loaded into the R environment. A `Startup` function that will load the "`biOps`", "`fBasics`" and "`cluster`" libraries will be built and then ran.

```
> Startup <- function() {
+     library(biOps)
+     library(fBasics)
+     library(cluster)
+ }

> Startup()

 [1] "cluster"    "fBasics"    "timeSeries" "timeDate"
 [5] "MASS"       "biOps"      "stats"      "graphics"
 [9] "grDevices"  "utils"      "datasets"   "methods"
[13] "base"
```

2 Preprocessing

In this section, the function needed to extract the image descriptors from an image's gray levels distributions will be described. An example on how to apply this function to an image will also be provided.

2.1 Step 1. Building the function

The `DescrExtr` function is compiled to extract from an image the following descriptors:

- mean
- skewness
- kurtosis
- the 16 central vigintiles (quantiles of m/20 order, spanning from Q15 to Q85)
- the Canny's edge enhancement derived index, calculated as the number of enhanced (edgy) pixels over the image

```
> DescrExtr <- function(immagine) {
+     myQ <- seq(0.15, 0.85, by = 0.05)
+     x <- readTiff(immagine)
+     x <- imgRGB2Grey(x)
+     x <- imgNormalize(x)
+     x <- imgBlur(x)
+     meanX <- mean(c(x))
+     skX <- skewness(c(x))
+     kuX <- kurtosis(c(x))
+     quX <- quantile(x, probs = myQ)
+     canX <- imgCanny(x, 0.5)
+     cnX <- length(which(canX == 0))
+     tmpX <- c(meanX, skX, kuX, quX, cnX)
+     names(tmpX) <- c("Mean", "Skewness", "Kurtosis",
+         "Q15", "Q20", "Q25", "Q30", "Q35", "Q40",
+         "Q45", "Q50", "Q55", "Q60", "Q65", "Q70",
+         "Q75", "Q80", "Q85", "CannyDerived")
+     return(tmpX)
+ }
```

2.1.1 Exemple of descriptors extraction

In this part, the `DescrExtr` function will be applied to a given image *tile* (*Figure 1*).

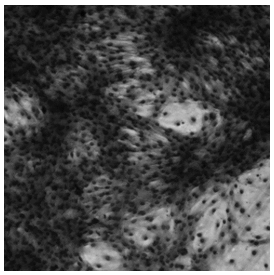


Figure 1: 512x512 *tile* (resized)

The 512x512 image shown in *Figure 1* is a *tile* generated from a 1024x1024 image, hereon referred to as parent image (*Figure 2*). We will work on the assumption that maximum 4 tiles will be generated for a single parent image. In fact, more than 4 512x512 tiles will only represent an over-sampling of the 1024x1024 original image.

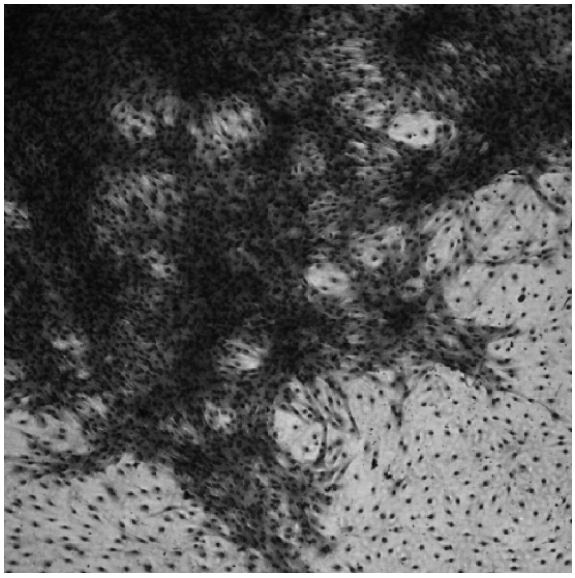


Figure 2: 1024x1024 parent image (resized)

The `DescrExtr` function is used to extract the image's descriptors from the processed *tile* the gray levels distribution (Figure 3).

Finally, the function will build a vector out of the extracted descriptors.

```
> Esempio <- DescrExtr("7[c]05-piastra 3 tipo III luce 7 p_A.tif")

        Mean     Skewness     Kurtosis          Q15
7.111242e+01 1.081049e+00 2.917967e-01 2.200000e+01
         Q20          Q25          Q30          Q35
```

```
2.700000e+01 3.100000e+01 3.500000e+01 3.900000e+01
         Q40          Q45          Q50          Q55
4.400000e+01 4.900000e+01 5.500000e+01 6.100000e+01
         Q60          Q65          Q70          Q75
6.800000e+01 7.600000e+01 8.500000e+01 9.700000e+01
         Q80          Q85 CannyDerived
1.120000e+02 1.320000e+02 2.131200e+04
```

2.2 Step 2. Creating the database

Following the described procedure, we applied the `DescrExtr` function to the image database. All the *tiles* were collected in the project folder. The used database featured:

- 82 parent images of canonical morphology (MD class)
- 20 parent images of unconventional morphology (MU class)
- 196 *tiles* generated from the MD parent images (from number 1 to number 196)
- 52 *tiles* generated from the MU parent images (from number 197 to number 248)

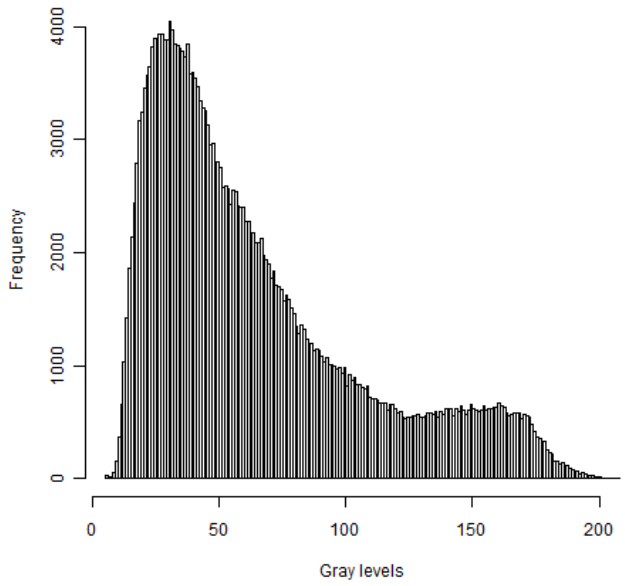


Figure 3: Graphic of the gray levels distribution for the 512x512 *tile*

```
> Images <- dir()
> DescrTable <- DescrExtr(Images)
```

A `parent` object was created and the names of the corresponding parent image for each *tile* were stored. This vector was then added to the `DescrTable` table. This step is needed in order to keep the linkage between *tiles* and generating parent images.

```
> DescrTable <- cbind(DescrTable, parent)
```

The first 3 records are shown for exemplificative purposes. Row names of the generated table represent the *tiles* names.

```
                     Mean  Skewness Kurtosis Q15 Q20 Q25
1[c]01-MN1_A.tif 193.2694 -1.458412 2.995726 167 175 181
1[c]01-MN1_B.tif 191.8352 -1.665833 3.624617 163 173 180
1[c]01-MN1_C.tif 201.0577 -1.737660 3.632982 172 183 190
                 Q30 Q35 Q40 Q45 Q50 Q55 Q60 Q65 Q70 Q75
1[c]01-MN1_A.tif 186 190 194 197 199 202 205 208 210 214
1[c]01-MN1_B.tif 185 190 193 197 199 202 205 208 211 214
1[c]01-MN1_C.tif 196 200 204 207 210 213 214 217 220 223
                 Q80 Q85 Canny     Parent
1[c]01-MN1_A.tif 216 221 12965 1[c]01-MN1
1[c]01-MN1_B.tif 215 220 15092 1[c]01-MN1
1[c]01-MN1_C.tif 226 229 14170 1[c]01-MN1
```

Finally, all the expert scoring for the parent images were included in the `ExpClass` object. This vector will **exclusively** be used for evaluation of agreement purposes. Therefore, the *unsupervised* classification is ensured. Expert classification scores are present for MD images only.

```
> ExpClass
```

```
  [1] MN MN MN MN MN MN MN MN MN MN MN MN MN MN MN MN MN T2
 [19] T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 MN T1
 [37] T1 T1 T1 T1 T1 T1 T1 T2 T2 T2 T2 T2 T3 T3 T3 T3 T3 T3
 [55] T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 MN MN T1 T1 MN
 [73] MN MN MN MN MN MN T1 T1 T1 T2 T2 T2 T2 T2 T2 T2 T2 T3 T3
 [91] T3 T3 T3 T3 T3 T3 T3 T3 T1 T1 T1 T1 T1 T1 MN MN MN MN
[109] T1 T1 T1 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2
[127] T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T3 T3 T3 T3 T3 T3 T3 T3
[145] T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 T3 MN MN MN MN T1 T1
[163] T1 T1 T1 T1 T1 T1 MN MN MN MN MN MN MN MN MN MN MN MN MN
[181] MN MN MN MN MN MN MN T1 T1 T1 T1 T1 T1 T1 T1 T1 T1
Levels: MN T1 T2 T3
```

3 Clustering

3.1 Clustering in F1

The classification scheme applied in this work is based on a hierarchical classification approach. In this two-steps classification model:

- in the first step of classification (hereon, F1) we will divide the *tiles* images in two classes, corresponding to the transformed/untransformed biological division.

- in the second step of classification (F2) we will further divide the two F1 obtained classes in two sub-classes each, corresponding to the canonical *foci* classification.

Both F1 and F2 steps of classification are based on the PAM clustering algorithm.

3.1.1 F1 clustering

The PAM algorithm was applied to the `DescrTable` table generated in the Preprocessing step (**section 3**). In this step the dataset will be divided in two classes using a PAM `k` paramenter of 2.

- The C1 (score of 1) class corresponding to normal monolayer or untransformed *foci*(noted as "NT")
- The C2 (score of 2) class corresponding to transformed *foci* (noted as "T")

The Canny's derived index (contained in the column 19) was not exploited in this step of classification. Moreover, the parent images scoring (column 20) was excluded to ensure the unsupervisedness of the classification process.

```
> clus512_F1_k2 <- pam(DescrTable[, -c(19, 20)],
+     k = 2)
```

The obtained `clus512_F1_k2` object contains the classification vector for F1:

```
> as.numeric(clus512_F1_k2$cluster)
```

```
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2
 [28] 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 2 2 1 2 2 2 2 2 2
 [55] 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [82] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1
[109] 1 1 1 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 1 2 2 2 2 2 1 2 1
[136] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 1 1 1 2 1
[163] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
[190] 1 1 1 1 1 1 1 2 1 2 2 2 2 2 2 2 1 1 2 1 1 2 2 1 1 2 2
[217] 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 1 1 1 2 1 1 1 1 1 2 1
[244] 2 1 2 2 2
```

3.1.2 F1 confusion matrix

In order to evaluate the performances of the classifier in terms of agreement of judgment with the expert and of sensitivity and specificity, a confusion matrix was built for each step of classification. The matrices featured:

- by **rows**: the expert scoring
- by **columns**: the algorithm scoring
- by **diagonal**: images (*tiles*) where the expert and the algorithm agree on.

The confusion matrix will be built around the images that are of the MD class (*tiles* 1 to 196), while images of mixed and intermediate nature (MU) will be excluded (*tiles* 197 to 248).

```
> confusion512_F1_k2 <- table(ExpClass, clus512_F1_k2$clustering[1:196])
```

```
ExpClass  1  2
      MN 54  0
      T1 37  2
      T2 16 38
      T3  4 45
```

The total number of *tiles* of the MD class scored as C1 is 111, while *tiles* scored as C2 are a total of 85. Moreover, while the automatic classifier is set to divide the images in 2 classes (C1 and C2), the expert scoring is on based on 4 classes (MN and the three canonical *foci* classes). For the sake of comparison of results, we will transform the confusion matrix from a 4x2 matrix in a 2x2, by modifing the `ExpClass` object transforming all the "MN" and "T1" instances in "NT" (untransformed), and the "T2" and "T3" in "T" (transformed).
The obtained `ExpClass_F1` vector is shown.

```
> ExpClass_F1
```

```
  [1] NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT T
 [19] T  T  T  T  T  T  T  T  T  T  T  T  T  T  T  T  NT NT
 [37] NT NT NT NT NT NT NT T  T  T  T  T  T  T  T  T  T  T
 [55] T  T  T  T  T  T  T  T  T  T  T  T  T  NT NT NT NT NT
 [73] NT NT NT NT NT NT NT NT NT T  T  T  T  T  T  T  T  T
 [91] T  T  T  T  T  T  T  T  NT NT NT NT NT NT NT NT NT NT
[109] NT NT NT T  T  T  T  T  T  T  T  T  T  T  T  T  T  T
[127] T  T  T  T  T  T  T  T  T  T  T  T  T  T  T  T  T  T
[145] T  T  T  T  T  T  T  T  T  T  T  T  NT NT NT NT NT NT
[163] NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT
[181] NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT NT
Levels: NT T
```

The modified confusion matrix will be as follows:

```
> confusions512_F1_k2_2x2 <- table(ExpClass_F1,
+     clus512_F1_k2$clustering[1:196])

ExpClass_F1  1  2
         NT 91  2
         T  20 83
```

Agreement, sensitivity and specificity values calculated on the 2x2 matrix are reported.

Summary	Value
Agreement	88.78%
Sensitivity	97.85%
Specificity	80.58%

3.2 Clustering in F2

During the F2 step of classification, the two F1 clusters will further be divided in two classes each, so to mimick the division of the untransformed images in normal monolayer and Type I, and of the transformed images in Type II and Type III.

3.2.1 Dividing the database

The classification vector obtained in F1 (`clus512_F1_k2$clustering`) was used to divide the database in two blocks:

- `DescrTable_C1` containing the C1 scored *tiles* (111 MD, 19 MU)
- `DescrTable_C2` containing the C2 scored *tiles* (85 MD, 33 MU)

```
> estrC1 <- which(clus512_F1_k2$cluster == 1)
> DescrTable_C1 <- DescrTable[estrC1, ]
> DescrTable_C2 <- DescrTable[-estrC1, ]
```

3.2.2 MN and T1 clustering

The PAM algorithm was applied to each of the two blocks separately, each time with `k`=2. First, the `DescrTable_C1` database was processed.

```
> clus512_F2_k2_MNT1 <- pam(DescrTable_C1[, -20],
+     2)
```

The expert scoring for the MD images classified as C1 during F1 was extracted from the `ExpClass` object. The obtained vector was used to build the partial 4x2 confusion matrix on the 111 MD *tiles*. Column names were set as "C'1" and "C'2".

```
> confusioni512_k2_F2_MNT1 <- table(ExpClass_C1,
+     clus512_F2_k2_MNT1$clustering[1:111])
> colnames(confusioni512_k2_F2_MNT1) <- c("C'1",
+     "C'2")

ExpClass_C1 C'1 C'2
         MN  52   2
         T1   1  36
         T2   2  14
         T3   2   2
```

3.2.3 T2 and T3 clustering

The approach used for MN and T1 classification was repeated to cluster the *tiles* contained in the `DescrTable_C2` table.

```
> DescrTable_C2 <- DescrTable[-estrC1, ]
> clus512_F2_k2_T2T3 <- pam(DescrTable_C2[, -20],
+     2)
```

The expert scoring for the C2 images was extracted from the `ExpClass` object, and the classification matrix built. Column names were set as "C'3" and "C'4".

```
> confusioni512_k2_F2_T2T3 <- table(ExpClass_C2,
+     clus512_F2_k2_T2T3$clustering[1:85])
> colnames(confusioni512_k2_F2_T2T3) <- c("C'3",
+     "C'4")

ExpClass_C2 C'3 C'4
         MN   0   0
         T1   1   1
         T2  31   7
         T3   4  41
```

3.2.4 Obtaining the 4x4 confusion matrix

Finally, the two partial 4x2 matrices were joined to obtain the complete F2 4x4 confusion matrix, on which agreement, sensitivity and specificity were calculated.

```
   C'1 C'2 C'3 C'4
MN  52   2   0   0
T1   1  36   1   1
T2   2  14  31   7
T3   2   2   4  41
```

Summary	Value
Agreement	81.63%
Sensitivity	82.42%
Specificity	94.01%

4 Merging *tiles* into parent images

4.1 Merging in F1

This section of code will approach the problem of scoring the parent images. In fact, while the scores we obtained in both F1 and F2 are relative to 512x512 *tiles*, the expert classification we used for comparison is relative to the parent images. Therefore, a majority of voting criterion was applied to assess a parent image's score based on the scores of its derived *tiles*. In case of draw, higher class was chosen.

4.1.1 Grouping the classification scores in F1

A `Voti` function was built to divide the classification *tiles* scores contained in a generic `y` table presenting in column 1 the parent images names and in column 2 the actual scores, grouping them on the basis of the generating parent images number (n = `x`)

```
> Voti <- function(x, y) {
+     matVoti <- matrix(0, x, 4)
+     rownames(matVoti) <- rep(0, x)
+     auxI <- 1
+     auxII <- 1
+     for (aux in 1:nrow(y)) {
+         if (aux < nrow(y)) {
+             if (y[aux, 1] == y[aux + 1, 1]) {
+                 matVoti[auxI, auxII] <- y[aux,
+                   2]
+                 auxII <- auxII + 1
+             }
+             if (y[aux, 1] != y[aux + 1, 1]) {
+                 matVoti[auxI, auxII] <- y[aux,
+                   2]
+                 rownames(matVoti)[auxI] <- y[aux,
+                   1]
+                 auxI <- auxI + 1
+                 auxII <- 1
+             }
+         }
+         if (aux == nrow(y)) {
+             if (y[aux, 1] == y[aux - 1, 1]) {
```

```
+                 matVoti[auxI, auxII] <- y[aux,
+                   2]
+                 auxII <- auxII + 1
+                 rownames(matVoti)[auxI] <- y[aux,
+                   1]
+             }
+             if (y[aux, 1] != y[aux - 1, 1]) {
+                 matVoti[auxI, 1] <- y[aux, 2]
+                 rownames(matVoti)[auxI] <- y[aux,
+                   1]
+             }
+         }
+     }
+     matVoti
+ }
```

The built function was then applied to our database. The list of the parent images was acquired from the column 20 of the `DescrTable` object created during **section 3**.

```
                      Mean  Skewness Kurtosis Q15 Q20 Q25
1[c]01-MN1_A.tif 193.2694 -1.458412 2.995726 167 175 181
1[c]01-MN1_B.tif 191.8352 -1.665833 3.624617 163 173 180
1[c]01-MN1_C.tif 201.0577 -1.737660 3.632982 172 183 190
                 Q30 Q35 Q40 Q45 Q50 Q55 Q60 Q65 Q70 Q75
1[c]01-MN1_A.tif 186 190 194 197 199 202 205 208 210 214
1[c]01-MN1_B.tif 185 190 193 197 199 202 205 208 211 214
1[c]01-MN1_C.tif 196 200 204 207 210 213 214 217 220 223
                 Q80 Q85 Canny     Parent
1[c]01-MN1_A.tif 216 221 12965 1[c]01-MN1
1[c]01-MN1_B.tif 215 220 15092 1[c]01-MN1
1[c]01-MN1_C.tif 226 229 14170 1[c]01-MN1
```

The `clus512_F1_k2$clustering` F1 clustering vector generated in **section 4.1.1** will be exploited in this classification step. For clarity's sake, we remind that a vote of "1" represents untransformed images ("NT"), while a vote of "2" represents transformed images ("T").

```
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2
 [28] 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 2 2 2 2 2
 [55] 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [82] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1
[109] 1 1 1 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 1 2 2 2 2 2 1 2 1
[136] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 1 1 1 2 1
[163] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
[190] 1 1 1 1 1 1 1
```

For the 196 MD derived *tiles*, the parent images names (column 1) and the scores of classification (column 2) were united in the `SchedaVoti` table.

```
> SchedaVoti <- cbind(as.character(DescrTable[1:196,
+     20]), clus512_F1_k2$clustering[1:196])
```

Finally, the `Voti` function was applied to the `SchedaVoti` table.

- `x`=82 (number of parent images)
- `y`=`SchedaVoti`
- ”0” characters are used as neutral characters when less than 4 *tiles* were extrated from a single parent image

The generated table will feature:

- by **rows** the parent images
- by **columns** the corresponding *tiles* scores

```
> F1_1024x1024 <- Voti(82, SchedaVoti)
```

The first 10 records are shown for exemplificative purposes.

```
                        [,1] [,2] [,3] [,4]
1[c]01-MN1              "1"  "1"  "1"  "1"
1[c]02-MN2              "1"  "1"  "1"  "1"
2[c]02-MN1piastra1      "1"  "1"  "1"  "0"
2[c]03-MN2piastra1      "1"  "1"  "0"  "0"
2[c]04-MNpiastra1       "1"  "1"  "1"  "1"
2[c]05-Tipo II_bordo_f1 "1"  "1"  "0"  "0"
2[c]07-Tipo II_bordo_f2 "2"  "1"  "0"  "0"
2[c]08-Tipo II_bordo_f3 "1"  "2"  "1"  "0"
2[c]09-Tipo II_bordo_f4 "1"  "1"  "2"  "0"
```

4.1.2 Applying the majority of voting criterion

The `MoV` function was built to extract a vector containing the majority of voting for each parent image from the table generated in **section 4.1.1**. In case of draw (e.g. if 2 ”NT” *tiles* and 2 ”T” *tiles* were generated from the same parent image), the highest corresponding biological damage class was chosen.

```
> MoV <- function(Scheda_MoV) {
+     mov <- c(NULL)
+     for (aux in 1:nrow(Scheda_MoV)) {
+         L1 <- length(which(Scheda_MoV[aux, ] ==
+             1))
+         L2 <- length(which(Scheda_MoV[aux, ] ==
+             2))
+         ifelse(L2 >= L1, mov <- c(mov, 2), mov <- c(mov,
+             1))
+     }
+     mov
+ }
```

Finally, the `MoV` function was applied to the `F1_1024x1024` table. The obtained `clus_F1_1024` vector is the clustering vector for parent images in F1.

```
> clus_F1_1024 <- MoV(F1_1024x1024)

 [1] 1 1 1 1 1 1 2 1 1 2 2 2 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 1
[29] 1 1 1 1 1 1 2 2 2 2 2 2 1 1 1 1 1 1 2 2 2 2 1 2 2 2 2 2
[57] 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1
```

4.1.3 F1 parent images confusion matrix

In order to obtain the confusion matrix, the expert scores for the parent images were copied in an `Exp_parent_F1` vector.

- untransformed images are scored as "**1**"
- transformed images are scored as "**2**"

```
 [1] 1 1 1 1 1 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 2 1
[29] 1 1 1 1 1 1 2 2 2 2 2 2 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
[57] 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

Finally, the confusion matrix was obtained, featuring the expert scores (by row) versus the parent images scores obtained through the majority of voting for F1 *tiles* scoring (by column).

```
> confusioni1024_F1_k2_2x2 <- table(Exp_parent_F1,
+     clus_F1_1024)

             clus_F1_1024
Exp_parent_F1  1  2
            1 35  2
            2  5 40
```

Agreement of classification, sensitivity and specificity are as follows.

Summary	Value
Agreement	91.46%
Sensitivity	94.59%
Specificity	88.89%

4.2 Merging in F2

The procedure described in the previous section was applied to the classification scores obtained in F2 in **section 4.2** in order to obtain the F2 scores for parent images.

4.2.1 Grouping the classification scores in F2

A `DescrTable_Ext` table was built starting from the `DescrTable` table created in **section 3**, by adding a column (column 21) containing the classification scores obtained in F2. Classification scores will be extracted from the `DescrTable_C1` and `DescrTable_C2` tables and modified:

- C'1 images are flagged as 1 (from the `clus512_F2_k2_MNT1$cluster` vector)
- C'2 images are flagged as 2 (from the `clus512_F2_k2_MNT1$cluster` vector)
- C'3 images are flagged as 3 (from the `clus512_F2_k2_T2T3$cluster` vector)
- C'4 images are flagged as 4 (from the `clus512_F2_k2_T2T3$cluster` vector)

```
> DescrTable_Ext <- cbind(DescrTable, rep(0, nrow(DescrTable)))
> colnames(DescrTable_Ext)[21] <- "Voti_F2"
> for (aux in 1:nrow(DescrTable_C1)) {
+     tmp <- which(rownames(DescrTable) == rownames(DescrTable_C1[aux,
+         ]))
+     DescrTable_Ext[tmp, 21] <- as.numeric(clus512_F2_k2_MNT1$cluster[aux])
+ }
> for (aux in 1:nrow(DescrTable_C2)) {
+     tmp <- which(rownames(DescrTable) == rownames(DescrTable_C2[aux,
+         ]))
+     DescrTable_Ext[tmp, 21] <- as.numeric((2 +
+         clus512_F2_k2_T2T3$cluster[aux]))
+ }
```

The obtained classification vector is shown.

```
> DescrTable_Ext[1:196, 21]
```

```
  [1] 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2 3 2 2 3 2 2 2 3
 [28] 3 3 3 3 3 3 3 1 2 2 1 2 2 2 2 2 2 2 3 3 2 4 4 4 4 4 4
 [55] 4 4 4 4 4 1 4 3 4 4 4 4 2 1 1 2 2 1 1 2 1 1 1 1 2 2 2
 [82] 4 3 3 3 3 3 3 4 4 3 4 4 4 4 4 3 4 2 2 2 2 2 2 1 1 1 1
[109] 2 2 2 4 3 3 3 2 3 3 3 3 3 2 2 3 4 4 1 4 3 3 4 3 1 4 2
[136] 3 4 3 4 4 4 4 4 2 4 4 4 4 4 4 4 4 1 4 4 4 1 1 1 1 3 2
[163] 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 4 2
[190] 2 2 2 2 2 2 2
```

4.2.2 Majority of voting criterion extraction modified

In order to apply the `Voti` function generated in the previous section to divide the scores by parent images, we built a `SchedaVoti_F2` table featuring:

- the parent images names (`DescrTable_Ext` column 20)
- the scores of classification for F2 (`DescrTable_Ext` column 21)

```
> SchedaVoti_F2 <- DescrTable_Ext[1:196, 20:21]
```

The `Voti` function was then applied to the `SchedaVoti_F2` table, and the `F2_1024x1024` table obtained. The first 10 records are shown for exemplificative purposes.

```
> F2_1024x1024 <- Voti(82, SchedaVoti_F2)
```

```
   [,1] [,2] [,3] [,4]
1     1    1    1    1
2     1    2    1    1
4     1    1    1    0
5     1    1    0    0
6     1    1    1    1
7     2    2    0    0
8     3    2    0    0
9     2    3    2    0
10    2    2    3    0
```

The `MoV_F2` function represents a modified version of the `MoV` function, developed in order to extract the majority of voting out of four classes.

```
> MoV_F2 <- function(Scheda_MoV) {
+     mov <- c(NULL)
+     for (aux in 1:nrow(Scheda_MoV)) {
+         L1 <- length(which(Scheda_MoV[aux, ] ==
+             1))
+         L2 <- length(which(Scheda_MoV[aux, ] ==
+             2))
+         L3 <- length(which(Scheda_MoV[aux, ] ==
+             3))
+         L4 <- length(which(Scheda_MoV[aux, ] ==
+             4))
+         L1234 <- c(L1, L2, L3, L4)
+         MaxL1234 <- which(L1234 == max(L1234))
+         if (length(MaxL1234) == 1) {
+             mov <- c(mov, MaxL1234)
+         }
+         if (length(MaxL1234) != 1) {
+             mov <- c(mov, max(MaxL1234))
+         }
+     }
+     mov
+ }
```

The `MoV_F2` function was applied to the `F2_1024x1024` table.

```
> clus_F2_1024 <- MoV_F2(F2_1024x1024)
```

The obtained clustering vector for parent images in F2 is shown.

```
 [1] 1 1 1 1 1 2 3 2 2 3 3 3 1 2 2 2 2 3 4 4 4 4 4 4 4 4 4 1
[29] 1 2 1 1 2 2 3 3 4 4 4 4 2 2 2 2 1 2 3 3 3 3 2 3 4 4 3 4
[57] 4 3 4 4 4 4 4 4 4 4 4 4 1 3 2 2 2 1 1 1 1 1 4 2 2 2 2
```

4.2.3 F2 parent images confusion matrix

The expert classification for the parent images was copied in the `Exp_parent_F2` object with the following modifications:

- monolayer images were flagged as 1
- Type I *foci* images were flagged as 2
- Type II *foci* images were flagged as 3
- Type III *foci* images were flagged as 4

```
 [1] 1 1 1 1 1 3 3 3 3 3 3 3 1 2 2 2 3 3 4 4 4 4 4 4 4 4 4 1
[29] 1 2 1 1 2 2 3 3 4 4 4 4 2 2 2 2 1 2 3 3 3 3 3 3 3 3 3 3
[57] 3 3 4 4 4 4 4 4 4 4 4 4 1 2 2 2 2 1 1 1 1 1 2 2 2 2 2
```

The confusion matrix for this step of classification was obtained obtained by crossing the expert scores in `Exp_parent_F2` with the parent images scores obtained through the majority of voting for F2 *tiles* scoring contained in the `clus_F2_1024` vector.

```
             clus_F2_1024
Exp_parent_F2  1  2  3  4
            1 17  0  0  0
            2  0 18  1  1
            3  0  5 14  4
            4  0  0  0 22
```

Summary	Value
Agreement	86.58%
Sensitivity	87.72%
Specificity	95.48%

5 The quantitative index

In this last section we will describe how a quantitative index suitable for MU scoring was extracted from the MD images classification.

5.1 Calculating the centers of the canonical classes

In order to assess the distance between the MU images and the canonical classes, the coordinates of each class center were calculated.
A `stdTab` function was built to standardize `y` columns in the descriptors table (`x`).

```
> stdTab <- function(x, y) {
+     tmp <- matrix(0, nrow(x), ncol(x))
+     for (aux in c(y)) {
+         tmpM <- sum(x[, aux])/nrow(x)
+         tmpV <- var(x[, aux])
+         for (auxI in 1:nrow(x)) {
+             tmp[auxI, aux] <- (x[auxI, aux] -
+                 tmpM)/tmpV
+         }
+     }
+     tmp <- cbind(tmp[, y], x[, -y])
+     colnames(tmp) <- colnames(x)
+     rownames(tmp) <- rownames(x)
+     tmp
+ }
```

The described function was applied to the `DescrTable_Ext` table, excluding the columns 20 and 21 due to their non-numerical nature.

```
> Descr_Std <- stdTab(DescrTable_Ext, c(1:19))
```

The first 3 records of the normalized table are shown for exemplificative purposes.

```
                        Mean  Skewness  Kurtosis        Q15
1[c]01-MN1_A.tif 0.03439151 -1.198333 0.4516872 0.03257787
1[c]01-MN1_B.tif 0.03354178 -1.400649 0.5651086 0.03100364
1[c]01-MN1_C.tif 0.03900589 -1.470708 0.5666173 0.03454565
                          Q20        Q25        Q30
1[c]01-MN1_A.tif 0.03051734 0.02912371 0.02822432
1[c]01-MN1_B.tif 0.02977477 0.02876278 0.02786610
1[c]01-MN1_C.tif 0.03348763 0.03237209 0.03180653
                          Q35        Q40        Q45
1[c]01-MN1_A.tif 0.02751783 0.02713246 0.02665857
1[c]01-MN1_B.tif 0.02751783 0.02676451 0.02665857
1[c]01-MN1_C.tif 0.03112865 0.03081205 0.03045085
                          Q50        Q55        Q60
```

```
1[c]01-MN1_A.tif 0.02604615 0.02583097 0.02571289
1[c]01-MN1_B.tif 0.02604615 0.02583097 0.02571289
1[c]01-MN1_C.tif 0.03040288 0.03041956 0.02971506
                        Q65        Q70        Q75
1[c]01-MN1_A.tif 0.02570930 0.02512342 0.02560046
1[c]01-MN1_B.tif 0.02570930 0.02565581 0.02560046
1[c]01-MN1_C.tif 0.03005059 0.03044736 0.03102482
                        Q80        Q85         Canny
1[c]01-MN1_A.tif 0.02449218 0.02473621 -8.013620e-05
1[c]01-MN1_B.tif 0.02379288 0.02391011 -6.694615e-05
1[c]01-MN1_C.tif 0.03148511 0.03134496 -7.266370e-05
                     Parent Voti_F2
1[c]01-MN1_A.tif 1[c]01-MN1       1
1[c]01-MN1_B.tif 1[c]01-MN1       1
1[c]01-MN1_C.tif 1[c]01-MN1       1
```

The *tiles* with the normalized descriptors were divided on the basis of their scores of classification as obtained in F2, thus creating four different tables, one for each cluster group (namely C'1, C'2, C'3 and C'4). For each table, the class center was calculated as the mean values for the 19 featured descriptors. The class center will act as a virtual paragon *tile* for each class, from which to calculate the distance for all the MU tiles.

```
> estr_MD_MN <- which(Descr_Std[1:196, 21] == 1)
> estr_MD_T1 <- which(Descr_Std[1:196, 21] == 2)
> estr_MD_T2 <- which(Descr_Std[1:196, 21] == 3)
> estr_MD_T3 <- which(Descr_Std[1:196, 21] == 4)
> tabMN <- Descr_Std[estr_MD_MN, ]
> tabT1 <- Descr_Std[estr_MD_T1, ]
> tabT2 <- Descr_Std[estr_MD_T2, ]
> tabT3 <- Descr_Std[estr_MD_T3, ]
```

The `ClassCenter` function was built to extract from a table (`x`) the mean value (thus, the class center) for the desired descriptors (`y`).

```
> ClassCenter <- function(x, y) {
+     tmp <- c(NULL)
+     for (aux in y) {
+         tmp <- c(tmp, mean(x[, aux]))
+     }
+     tmp
+ }
```

5.2 Assessing the distances

The four class centers were then united in a comprehensive table.

```
> centerMN <- ClassCenter(tabMN, c(1:19))
> centerT1 <- ClassCenter(tabT1, c(1:19))
```

```
> centerT2 <- ClassCenter(tabT2, c(1:19))
> centerT3 <- ClassCenter(tabT3, c(1:19))
> tabCenter <- rbind(centerMN, centerT1, centerT2,
+     centerT3)
> colnames(tabCenter) <- colnames(Descr_Std[, 1:19])
```

```
               Mean   Skewness    Kurtosis          Q15
centerMN  0.02982862 -1.3841773  0.57690088  0.028297083
centerT1  0.01246654 -0.2281752 -0.17059001  0.005524537
centerT2 -0.01632886  0.4061967 -0.25242681 -0.015053341
centerT3 -0.02747331  1.0254952 -0.08899969 -0.019901621
                Q20          Q25          Q30         Q35
centerMN  0.026869618  0.025805680  0.025063183  0.02449647
centerT1  0.006645006  0.007646646  0.008522186  0.00935677
centerT2 -0.014377366 -0.013847130 -0.013408881 -0.01305369
centerT3 -0.020060952 -0.020434325 -0.020925017 -0.02154504
                Q40         Q45         Q50         Q55
centerMN  0.02409197  0.02377777  0.02359332  0.02344520
centerT1  0.01011099  0.01087848  0.01163374  0.01237422
centerT2 -0.01269901 -0.01238768 -0.01200922 -0.01167723
centerT3 -0.02220397 -0.02296610 -0.02380971 -0.02462643
                Q60         Q65         Q70         Q75
centerMN  0.02330222  0.02322977  0.02320867  0.02314733
centerT1  0.01308875  0.01382882  0.01450511  0.01513122
centerT2 -0.01134427 -0.01105767 -0.01097586 -0.01118140
centerT3 -0.02549860 -0.02635663 -0.02718158 -0.02780674
                Q80         Q85         Canny
centerMN  0.02265193  0.02131588 -8.204021e-05
centerT1  0.01558266  0.01569506  3.690068e-05
centerT2 -0.01214301 -0.01399846  3.528646e-05
centerT3 -0.02751239 -0.02589160 -6.682832e-05
```

A `tabMU` table was created, by joining the MU *tiles* (52, in our case) and the four class centers. The distances between the *tiles* and the class centers table were calculated in the `tabDist` table.

```
> tabMU <- rbind(Descr_Std[197:248, 1:19], tabCenter)
> tabDist <- as.matrix((dist(tabMU)))[53:56, -c(53:56)]
> colnames(tabDist) <- parent[197:248]
```

The first 3 records of table `tabDist` are shown for exemplificative purposes.

```
         2[c]01-Incerto_bordo_h 2[c]01-Incerto_bordo_h
centerMN              1.6554299              1.7447085
centerT1              0.2879688              0.3786066
centerT2              0.3683948              0.2831108
centerT3              1.0036536              0.9211005
         2[c]11-Tipo II_centro1_f
centerMN               1.98940693
```

```
centerT1                0.68120000
centerT2                0.07467855
centerT3                0.59183746
```

A `Distanze` function extracting from an `x` table the mean distances for each of the composing `y` parent MU images from the distances of the derived *tiles* was built.

```
> Distanze <- function(x, y) {
+     mat <- matrix(0, 4, 4)
+     auxI <- 1
+     auxII <- 1
+     vec <- matrix(0, 4, y)
+     colnames(vec) <- rep(0, y)
+     for (aux in 1:ncol(x)) {
+         if (aux != ncol(x)) {
+             if (colnames(x)[aux] == colnames(x)[aux +
+                 1]) {
+                 mat[, auxI] <- as.numeric(x[,
+                   aux])
+                 auxI <- auxI + 1
+             }
+             if (colnames(x)[aux] != colnames(x)[aux +
+                 1]) {
+                 mat[, auxI] <- as.numeric(x[,
+                   aux])
+                 for (auxIII in 1:4) {
+                   vec[auxIII, auxII] <- (sum(mat[auxIII,
+                     ])/length(which(mat[auxIII,
+                     ] != 0)))
+                 }
+                 colnames(vec)[auxII] <- colnames(x)[aux]
+                 auxI <- 1
+                 auxII <- auxII + 1
+                 mat <- matrix(0, 4, 4)
+             }
+         }
+         if (aux == ncol(x)) {
+             if (colnames(x)[aux] == colnames(x)[aux -
+                 1]) {
+                 mat[, auxI] <- as.numeric(x[,
+                   aux])
+                 for (auxIII in 1:4) {
+                   vec[auxIII, auxII] <- (sum(mat[auxIII,
+                     ])/length(which(mat[auxIII,
+                     ] != 0)))
+                 }
+                 colnames(vec)[auxII] <- colnames(x)[aux]
```

```
+             }
+             if (colnames(x)[aux] != colnames(x)[aux -
+                 1]) {
+                 vec[, auxII] <- as.numeric(x[,
+                   aux])
+             }
+             colnames(vec)[auxII] <- colnames(x)[aux]
+         }
+     }
+     vec
+ }
```

The `Distanze` function was applied to the `tabDist` table. The obtained distances are the mean distance between a parent image and the four class centers, and by extension, to the canonic classes themselves.

```
> DistMU <- Distanze(tabDist, 20)
```

The first 3 records of the `DistMU` table are shown for exemplificative purposes.

```
     2[c]01-Incerto_bordo_h 2[c]11-Tipo II_centro1_f
[1,]              1.7000692                1.9254976
[2,]              0.3332877                0.6051854
[3,]              0.3257528                0.0939840
[4,]              0.9623770                0.6716120
     2[c]15-Tipo II-incerto_bordo_h
[1,]                       2.1984268
[2,]                       0.8931625
[3,]                       0.2615870
[4,]                       0.3981819
```

5.3 Quantitative Index of Dissimilarity (QIoD) in graphics

The mean distances between the parent images and the class centers represent a quantitative value of how much each of the MU images is distant (thus, *dissimilar*) from a typical image belonging to one of the canonical classes (monolayer, Type I, Type II, Type III). The smaller the dissimilarity, the closer the image is to that canonical class.
A barchart graphic of the dissimilarity index was built featuring:

- on the **x axis** the ID of the MU parent images (from 1 to 20)
- on the **y axis** the correspoinding distances. Bars represent, from bottom to top in grayshades, classes MN, T1, T2 and T3 respectively.

By definition, the sum of the four distances was set to 1. The `QiD` function was built to achieve this last feature and applied to the `DistMU` table.

```
> QiD <- function(x) {
+     mat <- matrix(0, nrow(x), ncol(x))
+     for (aux in 1:ncol(x)) {
+         tmp <- sum(x[, aux])
+         mat[, aux] <- x[, aux]/tmp
+     }
+     mat
+ }
> QiD_MU <- QiD(DistMU)
```

An extract of the obtained table `QiD_MU` is shown.

```
         [,1]       [,2]       [,3]
MN 0.51183984 0.58414278 0.58603488
T1 0.10034292 0.18359654 0.23809043
T2 0.09807439 0.02851215 0.06973128
T3 0.28974285 0.20374853 0.10614341
```

The barchart for the `QiD_MU` table was built with the R basic function `barplot`.

```
> barplot(QiD_MU, names.arg = c(1:20), xlab = "Parent Images",
+     ylab = "Quantitative index of Dissimilarity (QiD)")
```

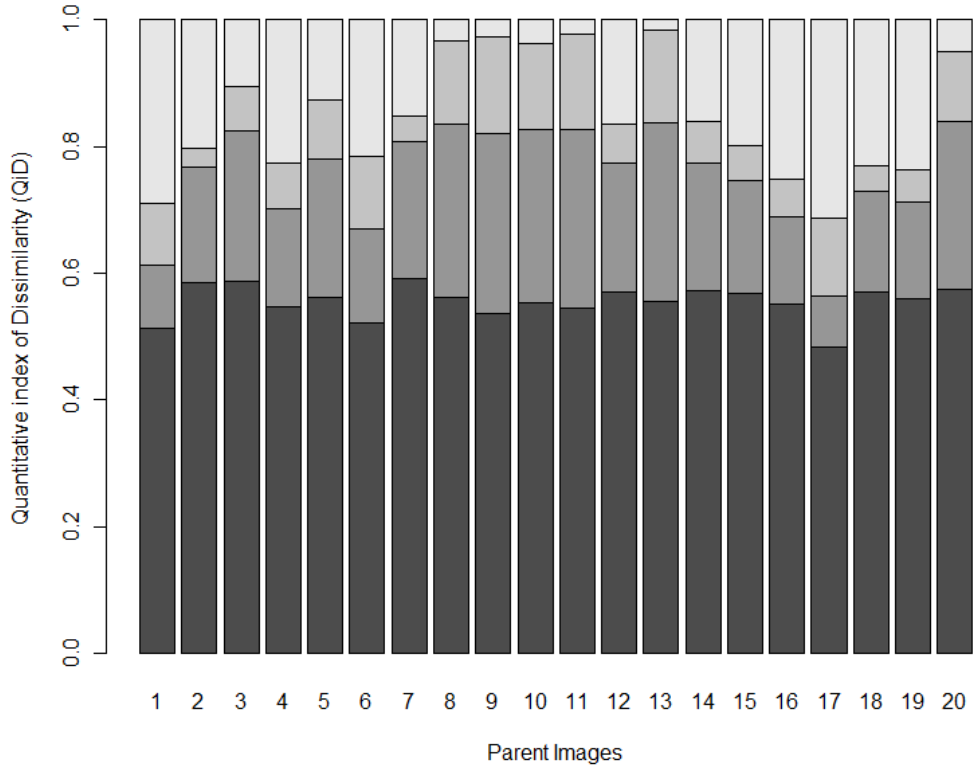


Figure 4: Barchart for the QIoD. In grayshades (from dark to light): MN, T1, T2 and T3

Publications

Urani C., Crosta G.F., Procaccianti C., Melchioretto P., Stefanini F.M. Image classiers for the cell transformation assay: a progress report. Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues VIII, Ed. by Farkas D. L., Nicolau D. V., Leif R. C. Proceedings of the SPIE 7568: 75681F-75681F-11 (2010).

Procaccianti C., Stefanini F.M., Urani C. The cell transformation assay: toward a statistical classification of mixed and intermediate foci images, Alternatives to Laboratory Animals 39(1):23-36 (2011).

Procaccianti C., Urani C., Stefanini F.M. The assessment of mixed and intermediate foci images using the R software environment, Working Papers del Dipartimento di Statistica “G. Parenti”, Università degli Studi di Firenze, available at http://www.ds.unifi.it/ricerca/pubblicazioni/working_papers/ls_working_paper.php#2011 (2011).

Fabbri M., Urani C., Sacco M.G., Procaccianti C., Gribaldo L. Whole genome analysis and microRNAs regulation in HepG2 cells exposed to cadmium (Submitted).