Options for Protecting Medical Data by IP Rights

Iryna Lishchuk, Marc Stauch Institut für Rechtsinformatik Leibniz Universität Hannover Hannover, Germany

e-mail: lishchuk@iri.uni-hannover.de, stauch@iri.uni-hannover.de

Abstract—This paper investigates different approaches to recognising Intellectual Property (IP) in medical data so as to protect efforts invested in enhancing the usability of such data through data curation. Sui generis database rights, copyrights and related rights, the legal regime of know-how all offer plausible options for protection here. In this paper, we analyse these options including by reference to a specific EC FP7 e-health research project (CHIC) and assess the prospects and potential benefits of applying them in order to protect investments made in data curation for medical research.

Keywords-IP rights; data rights; medical data; data curation.

I. INTRODUCTION

In recent years attention is increasingly focusing on the potential use of clinical health data for medical research. In principle, such data, recorded in patient or research databases can be of tremendous value when analyzed, in revealing linkages, e.g., between environmental and/or genetic factors and diseases. A major advantage too is that such connections can often be identified straight from the records, without the need for further invasive and potentially risky research.

As the potential value of health data becomes better understood, efforts to monopolize clinical data by exclusive IP or proprietary rights are also expanding. For instance, there are cases when the commercial use of health related data has been asserted under the coverage of database rights [1]. Patentable inventions have been derived out of the data of patients and research subjects and successfully commercialized [2]. The property right in data, generated in medical research, may also be claimed under contractual schemes [3]. At some point copyrights may also come to consideration for monopolizing data in medical domain [4].

However, as a precondition for allowing a significant amount of clinical data to be usefully exploited, there is an important initial step required in the form of data curation. In this regard, as we analyze below, most types of IP protection are tailored to protect specific objects that have already passed a certain threshold of maturity (data repositories, confidential information with assignable commercial value, etc.); but, as we discuss, none as such guarantees adequate protection to protect the prior investment made in curating the data.

In what follows, we begin by describing the data curation process in medical research in Section II. Then, in Section III, we consider the key relevant regimes of IP protection that may apply to protect such activity, namely: copyright and related rights, sui generis data base rights, and knowhow protection, as well as reliance on contractual mechanisms. By way of illustration, in Section IV, we consider how those regimes may apply to data curation in the context of a specific medical research project. In Section V the paper then concludes with some suggestions as to how curation activity may be better protected in the future.

II. DATA CURATION

The clinical data provided for e-health research usually comprises a large mass of data of multiple data types, formats. words. figures, numerical parameters, abbreviations, etc. From a technical standpoint, data integration is still a significant challenge for such research. In this regard, a starting point in the context of curation might be to see raw data in terms of the 'given', which as yet lacks semantic meaning, with the latter only emerging through the addition of an interpretive context (which also marks the change in state from data into information). It is arguably the technological development and transformation of raw or incompletely processed data into information (or the uncovering of additional semantic meaning), brought about by the curative process, which presents the suitable object of IP protection.

Data integration is key here, but the format, scope, parameter, structure, context, terminology, completeness, etc., of the individual and heterogeneous data are not standardized, which may affect their quality, and ultimately their interoperability and integration [5]. This could also potentially affect collaboration of the different researchers in this field if they use different semantics and techniques to describe, format, submit, and exchange data.

The curation required here to ensure the data relates to and measures the same phenomena with sufficient accuracy to be usable is a large and painstaking task. It includes the problem of dealing with incomplete data fields and cross-checking that various indices were measured and recorded in a similar way (e.g., images were taken using similar equipment, co-morbidities were classified using the same terminology, etc). It is evident too that considerable expertise and skill is required for it to be performed well: the curator needs to have a real feel and understanding for the subject matter in order to make sensible judgments in resolving various gaps and uncertainties.

III. POTENTIAL IP PROTECTION

A. Copyright and Related Rights

Clinical data comes for the most part from clinical trials, laboratory results, medical examinations, etc. An example of the clinical data from the research project is shown in Figure 1[6]. Such data is usually expressed in some numeric parameters, figures, words, combinations of such items. The representation of clinical data in this format is suitable and useful for digital data processing. However, the isolated items, be they words, keywords, syntax, figures or mathematical concepts as such, will not attract copyright. According to the Court of Justice of European Union (CJEU), items, "considered in isolation, are not as such an intellectual creation of the author who employs them." [7]. In order to be protected by copyright, the data must constitute the expression of the original author's creativity, which is only present when "through the choice, sequence and combination of those words that the author may express his creativity in an original manner and achieve a result which is an intellectual creation" [7]. The protection of clinical data by copyright may be acceptable for the medical reports, written by the physician or the patient and only when the expression of original creativity is achieved.

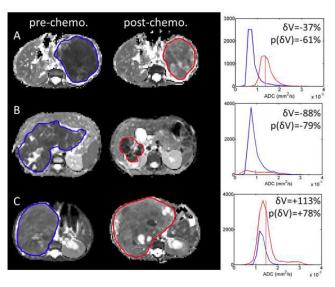


Figure 1. DWI and ADC mapping of nephroblastoma from different patients before and after pre-operative chemotherapy. Presented at the annual meeting of the British Chapter of the ISMRM, September 2012, provided by Prof. Kathy Pritchard-Jones from UCL. Copied from CHIC Deliverable D2-2 "Scenario based user needs and requirements" [6].

As may be seen from the image, some data is presented in visual form and is represented by images. However, medical images are normally produced by technical means (such as X-Ray, Ultrasound, etc.) and lack the creativity – an indispensable pre-requisite for copyright. A similar standard of copyright and requirement of original creativity applies to photographic works as well. According to Recital 16 Directive 2006/116/EC [8], a photographic work is

protected by copyright, if it is original. A work "is to be considered original if it is the author's own intellectual creation reflecting his personality". Other criteria such as merit or purpose are not relevant for copyright. According to the CJEU decision in the case C 145/10 REC of Eva-Maria Painer [9], copyright protects pictures taken by an individual, exercising free and creative choices, thus stamping a picture with his personal touch. It means, only pictures, which are taken by an individual expressing some level of creativity may be protected by copyright. On the other hand, images, generated automatically, will lack the creative input and may not be copyrighted. Since the images, produced in medical domain, are normally taken automatically and the process of recording is mostly completely managed by technical means, such images normally do not express creativity and do not attract the protection by copyright, respectively.

Apart from the rights considered so far, in the field of copyright there are a number of other emerging rights granted as a response to relevant investment. These rights are normally provided to the person, who invests in producing the protectable information. Such rights are referred to as related rights. Protection by related rights does not necessarily link to the intellectual creation (as the case is with traditional copyright), but rather to the economic investment.

The major rationale for protection by related rights tends to shift between intellectual creation and investment [10]. A mixture of artistic creation and investment attracts exclusive rights to performers in fixations of their performances. The economic investment constitutes a major factor, which renders exclusive rights to phonogram producers in their phonograms, to the film producers in respect of first fixations of their films, to broadcasting organizations in fixations of their broadcasts [11].

However, the number of related rights as of now is rather limited (mostly to those, indicated above). Therefore, attaching added value to the data enriching, post-procession, modification, etc., does not constitute the kind of investment protectable by related rights.

Against these considerations, the protection of clinical data, which is normally collected in the course of medical examinations and is represented in some numerical or technical visual format, by copyrights or related rights, may not be considered as a practicable solution, because the requirements for copyright protection in this data would not be met.

B. Sui Generis Database Right

As a rule, clinical institutions, participating in medical research, manage and maintain the clinical data in the clinical data repositories. Some clinical institutions manage their clinical information and store the results of clinical trials using Ontology-based Clinical Trial Management Application (ObTiMA) [12]. Others prefer data management systems specific to their medical activities.

Against this practice, an option of protecting the clinical data under the umbrella of sui generis database rights comes into consideration first.

The legal protection of databases is provided by the Directive 96/9/EC of 11 March 1996 on the legal protection of databases (the Database Directive) [13]. Such protection is granted in recognition of the fact that constructing a database requires "investment of considerable human, technical and financial resources" [13]. The directive aims to reward and protect such investment by providing the maker of a database with a sui generis data base right that places him in a position to prevent unauthorized access and copying of the database contents, which he compiled. In this regard, Article 7 Database Directive states:

"Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database." The object of protection in terms of the Database Directive is a 'database' meaning "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means" [13].

Protection of databases by the sui generis right can be considered as a plausible option for protecting the clinical data repositories, provided such repositories satisfy the criteria for protection. For this, the repository must show significant investment in "the obtaining, verification or presentation" of its contents.

As regards the scope of the database right, it would protect the collected data from being copied as a whole or in substantial part, evaluated "qualitatively and/or quantitatively" and either copied in one action or step by step [13].

Provided the clinical data repository qualifies as a database in the meaning of Database Directive and the clinical institution holds the sui generis database rights, the institution may stipulate the terms of using the repository contents as a whole, grant the rights of use under contractual and enforce the prevent unauthorized extraction/reutilization of the repository contents as a whole or in substantial part. The holder of sui generis database rights may leverage how the contents of its repository may be used, whether the data items may be extracted (downloaded) and in what scope, whether the data may be transferred to external parties or whether the data procession may only be done on its premises.

However, the sui generis protection applies to the contents of the repository as a whole or in substantial part and may apply separately and irrespective of protectability of data items by other rights, such as copyrights. Article 7 (4) makes this explicit, saying that the database right: "shall apply irrespective of eligibility of the contents of that

database for protection by copyright or by other rights. Protection of databases [....] shall be without prejudice to rights existing in respect of their contents".

Thus, the holder of the repository may manage the use of the repository contents as a whole. However, the use of separate data items in the repository may remain governed by the terms, stipulated by the data providers and/or holders of rights in such items. For instance, the access rights to the datasets, handled as confidential, may require signing of non-disclosure agreement (NDA) and the use of such data may be limited and be subject to technical protection measures, etc.

The options of protection, which potentially may apply to separate datasets we consider next.

C. Know-how

Because of the high sensitivity of health related data (and the potential harm from disclosure to the patient's interests in privacy, dignity and autonomy), clinical data in the medical treatment domain is managed under the rules of professional medical secrecy and subject to the fiduciary duties. For preserving the secrecy of clinical data, after such data leaves the medical domain (where it was handled under the rules of professional medical secrecy) and enters the domain of clinical research (where not necessarily all parties are bound by the rules of professional secrecy), protecting such data under the legal regime of know-how (or as undisclosed information) may be advised as a good option.

Protection of undisclosed information is provided by Section 7, Article 39 et seq. TRIPS Agreement [14]. The legal regime of know-how enables natural and legal persons, who are in legitimate possession of such information, to prevent such information "from being disclosed to, acquired by, or used by others without their consent in a manner contrary to honest commercial practices." Unfair practices for these purposes would include the acquisition of information via violation of contractual duties, breach of confidentiality obligations, inducement to breach, etc. [14].

In order to be protectable, the relevant information should have the quality of protectable information within the meaning of Article 39 TRIPS Agreement. Article 39 TRIPS Agreement protects information, which:

"(a) is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;

- (b) has commercial value because it is secret; and
- (c) has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret." [14].

The first weak point of protecting clinical data as knowhow is that as of now the legal framework on know-how protection in the EU is not harmonized [15]. Although, there is a proposal for a draft directive on the protection of undisclosed information in the EU (the Draft Directive) [16], before it is adopted and implemented, protection of know how remains dispersed through the national states of the EU Member States, and subject to varying requirements for and scope of protection

The Draft Directive, which is intended to harmonize the national laws in relation to know-how protection, in many aspects repeats the provisions of the TRIPS Agreement (in particular, it relates to the protectable subject matter and requirements for protection (Article 2), acts of unfair acquisition of information (Article 3), rights and remedies conferred (Article 5 et seq), etc.). In this regard it may also be queried how far the Draft Directive, if adopted, would improve the protection for data, the preparation of which consumed much effort, but which for one or another reason may not reach the level of protectable know-how. Here the key obstacles in applying know-how protection to the clinical data, processed for research, relate to the need (in order to be protected) for such data to be secret, subject to the confidentiality measures and have economic value.

First, to satisfy the criterion of secrecy, the information, sought to be protected, must be accessible to a limited number of persons only. The use of such information must be subject to confidentiality measures. The application of confidentiality measures means that the data must be stamped as "Confidential" and the sharing of such data must be made upon non-disclosure obligation and observation of the confidentiality measures. Disclosure of such datasets without due confidentiality measures might compromise the regime of secrecy so that protection would be forfeited. As regards the requirement of economic value of know-how, this will be considered to be present if through publication, the research investment and competitive standing of the entity doing the work would be undermined [17].

In relation to the volumes of clinical data, made available for research, this requirement, besides being at odds with the underlying culture of academic research, would create further workload. The data, subject to the regime of confidentiality, must first be strictly identified. The confidentiality mark would need to be attached to individual data items and any use and disclosure of such data to any third party must be made upon signing the non-disclosure agreement. This preservation of the confidentiality mark, conclusion of NDA and control over handling such data as confidential would present another challenge.

Against these considerations, the protection of clinical data under the legal regime of know-how might, in principle, be possible in relation to some defined amount of data, but hardly offers a feasible solution, when protection of large amounts of data, processed in medical research is sought. It also may operate against the principle of openness, if optimal use is to be made of the data by the research community, exploiting the full potential of available datasets.

D. Contractual Approaches

Insofar as the IP regimes for protecting the data, produced in medical research projects fail, one further method for regulating rights in data may be by contractual relations. Thus, in third party funded projects, the relations of ownership over the research results are typically governed by contract. The sponsor is typically interested to exploit the project results and funding is typically granted upon condition that the sponsor acquires the ownership and exploitation rights over the research results [3]. This model does not cause problems in practice, because the acquisition of ownership and exploitation rights is typically foreseen by the contract. The participating institutions are bound by these contractual relations and required to procure the ownership over the research results from the persons, whom they engage into the project.

IV. APPLICATION OF IP REGIMES TO DATA CURATION IN CHIC

A. Background

The research project "Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology", is an ICT research project in the clinical domain [18]. CHIC develops clinical trial driven tools and services within a secure infrastructure, which facilitate the creation of multiscale cancer hypermodels (integrative models) by technical means. These composite multiscale constructs of models (hyper-models or integrative models) are intended to synthesize and imitate the biological processes, which occur in course of tumor progression, at several temporal and spatial levels (molecular, cellular, etc.) at once.

In this context too, the study of how individual cancer components interact with each other has led to an explosion in the number of different types of data generated from the patients such as: molecular data, epigenetic data, clinical data, imaging data, pathology data and other laboratory data [19]. These different data types are assembled in order to systematically explore and formalize them in mathematical models.

Subsequently, the models are developed and validated against clinical data either taken from the literature or provided by the clinical partners [20]. The data management systems, used by the clinical partners, differ. Whereas the integration of data from data management system ObTiMA [12] is harmonized, the data from individual clinical data repositories need to be adapted to the requirements of the project. The use of diverging data management systems by the clinical institutions leads to the situation that the data, collected from different sources, is not inter-operable with each other and mostly cannot be used for research as such. The clinical data also needs to be post-processed by the modelers so that it fits into the set of parameters, which the models recognize and can utilize as an input for running the simulations. Data curation is a very important step because

the inputs, outputs and descriptions of processes, simulated by the models, need to be standardized into the set of parameters, acceptable and usable by all cancer models.

B. Applicability of IP Regimes to Project Data Curation

The clinical data, which after the necessary deidentification enters the domain of CHIC, is placed and stored in the CHIC clinical data repository. The CHIC data repository hosts data categorized per data type: imaging data (DICOM etc), descriptive/structural data (age, sex etc), other files (histological reports), links (to other data repositories) etc. The datasets for each type are accessible individually so that the data corresponding to the model parameters may be chosen. The fact that the repository is built "based on the experience already accumulated during the implementation of other data repositories" should be sufficient to prove the requisite investment in "either the obtaining, verification or presentation" of its contents [13]. Against this background, the database right in the CHIC clinical data repository is likely to be granted.

Protection of the CHIC data repository by the sui generis database rights would go to the maker of the database. In the meaning of the Database Directive, the maker of a database is seen as "the person who takes the initiative and the risk of investing", but excluding subcontractors [13]. Thus, the party, who constructed the CHIC repository, would be in a position to manage the use of the repository, such as by allocating the access rights to the project parties or external parties, to define the rights of use (access only, modification, download, etc.), to divide the repository into sections and define different regimes of uses depending on the data stored therein, etc. Grant of the sui generis protection would also entitle the right holder to enforce his rights, once unauthorized copying of the repository contents on the large scale has occurred.

Apart from the protection of the repository contents as a whole by sui generis database rights, the items in the repository may also enjoy protection in their own right. Since the clinical data repository deals with highly sensitive information (meaning that already for that reason, access to the data is strictly limited), application of the legal regime of know-how to some data items at least may be an option. As we saw above, for this, the data items, selected for knowhow protection, must be identified, the access and use of such data shall be limited to a defined number of people only, the management of such data shall be subject to confidentiality measures. In the case of CHIC, the regime of secrecy may be provided to the data via marking it as "Confidential" and making the disclosure of such data subject to the non-disclosure obligation. Considering from the technical side, the confidentiality mark would then need to be placed and borne by the data throughout the whole research process so that the data marked as "confidential" by the input comes out marked "confidential" by the output. This would present an additional workload, but is implementable. Also, disclosure of such data items to the CHIC parties subject to the non-disclosure obligation would not present a significant obstacle, because the project parties are bound by the contractual relations within the project. The factual use of data within the project may also be managed by technical measures, such as granting or denying the access rights, rights of use and extraction, limiting the data procession to the framework of technical infrastructure of CHIC only. Whereas the application of such contractual and technical confidentiality measures to the clinical data in CHIC may be feasible, in how far such technical and confidentiality measures may be implemented in other medical research projects may be questionable.

By contrast, copyrights and related rights offer less plausible options for protecting the clinical data in CHIC. As noted above, the clinical data in CHIC is represented by technical data from clinical trials, which is composed from different parameters. As observed in Section III, isolated items are not protectable by copyright. Copyright will fail against the lack of creativity expressed in such data. The investment, deployed in curating the data for CHIC, does not qualify as investment, protectable by related rights.

V. CONCLUSIONS

As we have seen, there are various ways in which the activity of curating clinical datasets could benefit from IP protection. Thus, collecting, arranging the data into a repository and making it suitable for use may render the investment, deployed in collecting and presenting the data, protectable by sui generis database rights. Similarly, the generation of research data and adoption of additional confidentiality and security measures to keep this data secret to the broader community may render such data protectable as know-how.

However, the present approach that seeks to maintain (commercial) data confidentiality by keeping data secret leads to a fragmented research environment, and reduces the chances for greater data interoperability to be achieved. Here the law - aided by technology should aim to encourage greater openness, while assuring appropriate curation rewards. This could, e.g., take the form of an officially endorsed mechanism or system for measuring and tagging changes produced in a given data set (or the merging of several data sets) resulting from curation efforts, as the reward-trigger. At the same time, as another crucial policy element, the law needs – especially in the case of the curation of sensitive health data – to ensure that privacy and other interests of patients and research subjects are and remain adequately protected.

In particular, it will here be necessary to take account of (and compensate for) the knock-on effects of IP changes, where data-holders are no longer (also) motivated by commercial considerations to keep their data secure and confidential. This concern is all the greater here since the activities of data sharing and curation being encouraged, also by their nature present enhanced risks to personal privacy. The point of curation is precisely to uncover new

connections and patterns in data that help generate robust inferences (usable – for good or ill) about the relevant data subjects. Accordingly, it is submitted that any system for rewarding investment in data curation should also require (as a condition for such rewards) that the data curator takes every appropriate measure to counterbalance the associated enhanced risks to privacy.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement No 600841.

REFERENCES

- [1] D. M. Gitter, "Informed Consent and Privacy of De-Identified Information and Estimated Data, Lessons from Iceland and the United States in an Era of Computational Genomics," (Published Conference Proceedings style), in Proc. ALLDATA 2016, The Second International Conference on Big Data, Small Data, Linked Data and Open Data (includes KESA 2016), Lissabon, 2016, pp. 7-12.
- [2] Moore v. Regents of University of California, Supreme Court of California, July 9, 1990, 51 Cal. 3d 120.
- [3] H.-D. Lippert, "Wem gehören Daten, die im Rahmen von Forschungsprojekten gewonnen werden?"/"To whom belongs the data generated in research projects?" in: Geistiges Eigentum: Schutzrecht oder Ausbeutungstitel?/in: Intellectual Property: Protection right or title to exploit, Springer, Volume 5, 2008, pp. 359-369.
- [4] Haimo Schack, "Zur Rechtfertigung des Urheberrechts als Ausschliesslichkeitsrecht"/"On justification of copyright as exclusive right," in: Geistiges Eigentum: Schutzrecht oder Ausbeutungstitel?/in: Intellectual Property: Protection right or title to exploit, Springer, Volume 5, 2008, pp. 124-140.
- [5] European Commission, "E-Health Task Force Report Redesigning health in Europe for 2020," Luxembourg 2012, ISBN 978-92-79-23542-9
- [6] CHIC, Deliverable D2-2 "Scenario based user needs and requirements", http://chic-vph.eu/uploads/media/D2-2_Scenario-based_user_needs_and_requirements.pdf> 16.08.2016.
- [7] CJEU, Judgment of 16 July 2009, Case C 5/08, Infopaq International A/S v Danske Dagblades Forening, Ref. 45.
- [8] Directive 2006/116/EC on the term of protection of copyright and certain related rights (codified version), OJEU, L 372/12, 27 December 2006.
- [9] CJEU, Judgment of 7 March 2013, Case C 145/10 REC, Eva-Maria Painer v. Standard VerlagsGmbH, Axel Springer AG, Süddeutsche Zeitung GmbH, Spiegel-Verlag Rudolf Augstein GmbH & Co. KG, Verlag M. DuMont Schauberg Expedition der Kölnischen Zeitung GmbH & Co. KG.
- [10] H. Zech, "Information als Schutzgegenstand," Tübingen, 2012, ISSN: 0940-9610 (Jus Privatum).
- [11] Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJEU L 167/10 - L 167/19, 22.6.2001.
- [12] H. Stenzhorn, et al, "The ObTiMA system ontology-based managing of clinical trials," Stud Health Technol Inform. 2010;160(Pt 2):1090-4.
- [13] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJEU, L 77/20 - L 77/28, 27.3.96.

- [14] Agreement on Trade-Related Aspects of Intellectual Property Rights, the TRIPS Agreement, Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization, Marrakesh, Morocco, 15 April 1994.
- [15] Hogan Lovells International LLP, "Report on Trade Secrets for the European Commission – Study on Trade Secrets and Parasitic Copying (Look-alikes), MARKT/2010/20/D", 2011.
- [16] Proposal for a Directive on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, Brussels, 28.11.2013, COM(2013) 813 final, 2013/0402 (COD).
- [17] K. Lodigkeit, Intellectual Property Rights in Computer Programs in the USA and Germany, Peter Lang GmbH, 2006, pp. 98-101.
- [18] Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology http://chic-vph.eu/project/> 16.08.2016.
- [19] C. Coveney, J. Gabe, and S. Williams, "The sociology of cognitive enhancement: medicalisation and beyond," Health Sociol. Rev., 20 (2011), pp. 381–393.
- [20] J. Dejaegher, L. Solie, S. De Vleeschouwer, and S. W. Van Gool, "Dendritic Cell Vaccination for Glioblastoma Multiforme: Clinical Experience and Future Directions," In G. Stamatakos and D. Dionysiou (Eds): Proc. 2014 6th Int. Adv. Res. Workshop on In Silico Oncology and Cancer Investigation – The CHIC Project Workshop (IARWISOCI), Athens, Greece, Nov.3-4, 2014 (www.6thiarwisoci.iccs.ntua.gr), pp.14-18. (open-access version), ISBN: 978-618-80348-1-5.