Bioinformatics for Precision Medicine in Oncology: principles and application to the SHIVA clinical trial

Nicolas Servant[1,2,3]*, Julien Roméjon[1,2,3], Pierre Gestraud[1,2,3], Philippe La Rosa[1,2,3], Georges Lucotte[1,2,3], Séverine Lair[1,2,3], Virginie Bernard[4], Bruno Zeitouni[1,2,3], Fanny Coffin[1,2,3], Gérôme Jules-Clément[1,2,3,10], Florent Yvon[1,2,3], Alban Lermine[1,2,3], Patrick Poullet[1,2,3], Stéphane Liva[1,2,3], Stuart Pook[1,2,3], Tatiana Popova[1,9], Camille Barette[1,2,3,5], François Prud'homme[1,2,3,4,5], Jean-Gabriel Dick[5], Maud Kamal[6], Christophe Le Tourneau[2,3,7], Emmanuel Barillot[1,2,3] and Philippe Hupé[1,2,3,8]*

1 Institut Curie, Paris, France
2 INSERM U900, Paris, France
3 Mines ParisTech, Fontainebleau, France
4 Institut Curie, Sequencing facility ICGex, Paris, France
5 Institut Curie, Informatic Department, Paris, France
6 Institut Curie, Translational Research Department, Paris, France
7 Institut Curie, Department of Medical Oncology, Paris, France
8 CNRS UMR144, Paris, France, 9 INSERM U830, Paris, France
9 INSERM U830, Paris, France
10 INSERM U932, Paris, France

Correspondence: Philippe Hupé, Plateforme de Bioinformatique, Unité 900 : Institut Curie - INSERM - Mines ParisTech, UMR144: Institut Curie - CNRS, 26, rue d'Ulm, 75248 Paris Cedex 05 – FRANCE. philippe.hupe@curie.fr

Introduction

Personalized medicine requires a strong interdisciplinary collaboration between several stakeholders covering a large range of expertise from medical, clinical, biological, translational, biotechnological and bioinformatics fields. The variety of actors and cultures, together with the diversity of constraints make it very challenging to apply personalized medicine in daily clinical practice, to meet expected short deadlines to deliver the results. Personalized medicine strongly depends on our ability to collect, disseminate and process complex information. Indeed, every stakeholder produces information during the healthcare pathway at different time and different places and all these informations have to be gathered, integrated and summarized in a digested report to make easy the therapeutic decision according to rules defined a priori.

Materials and Methods

The KDI system

To tackle the challenge of data integration, we have developed a dedicated information system named KDI (Knowledge and Data Integration) able to handle the heterogeneity and the complexity of the data (Servant et al., 2014). The KDI system ensures information sharing, cross-software interoperability, automatic data extraction, and secure data transfer. Using state-of-the-art informatics technologies, KDI acts as a hub to allow all the data being referenced such that it knows exhaustively which data is available for a given patient and where the raw and processed data are physically stored.

The KDI system consists of different modules dedicated to the storage, processing, analysis and visualization of each type of data (clinical, biological, microarray, NGS, etc.). High modularity associated with an efficient interoperability makes our system able to retrieve any relevant information. To facilitate the developments of these modules, we have retained a classical n-tiers architecture implemented with the JAVA/J2EE language. The core of each module of the KDI system can be presented as the association of different layers (Figure 3B).
Data layer. Data are stored in a relational database using the Entity-Attribute-Value (EAV) pattern. This conceptual modeling provides a data model plasticity required to handle the heterogeneity and the scalability of the variables of interest. Therefore, with EAV modeling, same concepts managed by different projects (with specific requirements by project) can be stored in a unique database without any modification of the data model. MySQL has been chosen as database provider for all web applications of the system. Complementary solutions such as NoSQL databases are currently evaluated for particular requirements (ontologies storage, specific queries, etc.).
Data access layer. Data access is supported by the DAO (Data Access Object) pattern. By using HibernateDaoSupport superclass provided by Spring Framework, we promote the standardization of database access for all standard queries (findAll, findById, save, delete). Moreover, Hibernate mapping through JPA annotations associated with use of Hibernate Criteria provides a homogeneous frame for this critical layer. Database sessions and transactional aspects are also delegated to Spring Framework.
Business layer. Business core of our web applications has two main objectives: i) provide structured data for presentation layer, and ii) make data available for remote and secured access by other applications and technical users. Standard services are developed using core functionalities of Spring framework (Aspect-Oriented Programming - AOP, Inversion of Control - IoC, JavaBeans Factory). Web services are published (server side) and invoked (client side) through Apache CXF framework. To respect Web Services Security (WS-Security) standards, we use the Apache WSS4J project provided by CXF (with interceptors chain process) to set up a username token authentication on each web application in the system.
Front-end layer. Presentation layer is based on JSF (Java Server Faces) which is a component oriented framework for building user interfaces for web applications. To enrich the basic component set provided by JSF, we use additional component libraries such as Apache Trinidad and Primefaces. By this systematic approach for each user interface, we aim to build a visual identity, ergonomic, easily usable, for the whole information system. All data available within KDI can be browsed and retrieved from a user-friendly bioinformatics web portal.
Client layer. This layer represents the web browser through which end-users access KDI system.

The SHIVA clinical trial design

The SHIVA clinical trial is a randomized proof-of-concept phase II trial comparing molecularly targeted therapy based on tumour molecular profiling versus conventional therapy in patients with refractory cancer (Le Tourneau et al., 2012, 2014). For each patient, a biopsy from the metastasis is performed and the molecular profiles are assessed using both the Cytoscan HD technology (Affymetrix) for the detection of DNA copy number alterations and loss of heterozygosity (LOH), and the Ion Torrent™ PGM sequencing technology (Life Technology) for the detection of somatic mutations. Immunohistochemistry (IHC) is used for the assessment of hormone receptor status, including oestrogen, progesterone and androgen receptors, as well as for the validation of focal gene amplifications detected with Cytoscan HD. DNA copy number amplifications (Affymetrix Cytoscan HD microarray) and mutations (next-generation sequencing with IonTorrent) in a subset of 76 genes along with biomarkers detected by IHC are considered for the decision-making.

Data preprocessing

DNA copy number amplifications (Affymetrix Cytoscan HD microarray) and mutations (next-generation sequencing with IonTorrent) are analyzed by in-house *ad-hoc* pipelines. The raw data as long as the results of the pipelines are stored in the KDI system. Clinical data are sored in the system too.

Integrative analysis: the report for the Molecular Biology Board (MBB)

The last step of the bioinformatics workflow is the production a technical report for the MBB. This task is crucial and must be complete and precise on one hand, and summarized on the other to allow a quick decision of the board. To answer this need, a report is automatically generated for each patient from the data stored in KDI. This report first presents the clinical information of the patient and the overall molecular profiles per gene, with the DNA copy number alterations, LOH (Loss of Heterozigosity) status, and number of mutations. This first section provides the MBB with a rapid overview of all detected alterations. If needed, the MBB can also have access to more detailed results, with graphical views of the copy number profiles for each gene, as well as the list of mutations with detailed annotation as previously described. This name-blinded technical report is sent to the members of the MBB for scientific validation and prioritization of the identified molecular abnormalities.

Results

We have developed a seamless information system named KDI that fully supports the essential bioinformatics requirements for PM. The system allows management and analysis of clinical information, classical biological data as well as high-throughput molecular profiles. It can deliver in real-time information to be used by the medical and biological staff for therapeutic decision-making. KDI makes it possible to share information and communicate reports and results across numerous stakeholders, representing a large continuum of expertise from medical, clinical, biological, translational, technical and biotechnological know-hows. The system relies on state-of-the-art informatic technologies allowing cross-software interoperability, automatic data extraction, quality control and secure data transfer. KDI has been successfully used in the framework of the SHIVA clinical trial for more than 18 months. As of June 2014, 730 patients have been included and 152 randomized in the SHIVA trial. KDI is used for other trials in the framawork of european FP7 projects (RAIDs for cervival cancer – http://www.raids-fp7.eu/ ; MAARS for allergy and autoimmune diseases related to skin – http://www.maars.eu/).The KDI is curently used to manage all the high-throughput data in Institut Curie.

References

Le Tourneau C, Kamal M, Trédan O, Delord JP, Campone M, Goncalves A, Isambert N, Conroy T, Gentien D, Vincent-Salomon A, Pouliquen AL, Servant N, Stern MH, Le Corroller AG, Armanet S, Rio Frio T, Paoletti X. Designs and challenges for personalized medicine studies in oncology: focus on the SHIVA trial. Target Oncol. 2012 Dec;7(4):253-65.

Servant N, Roméjon J, Gestraud P, La Rosa P, Lucotte G, Lair S, Bernard V, Zeitouni B, Coffin F, Jules-Clément G, Yvon F, Lermine A, Poullet P, Liva S, Pook S, Popova T, Barette C, Prud'homme F, Dick J, Kamal M, Le Tourneau C, Barillot E and Hupé P. Bioinformatics for Precision Medicine in Oncology: principles and application to the SHIVA clinical trial. 2014. Front. Genet. 5:152. doi: 10.3389/fgene.2014.00152

Le Tourneau C, Paoletti X, Servant N, Bièche I, Gentien D, Rio Frio T, Vincent-Salomon A, Servois V, Romejon J, Mariani O,1 Bernard V, Hupé P, Pierron G, Mulot M, Callens C, Wong J, Mauborgne C, Rouleau E, Reyes C, Leroy Q, Henri E, Gestraud P, La Rosa P, Escalup L, Mitry E, Trédan O, Delord JP, Campone M, Goncalves A, Isambert N, Gavoille C , Kamal M. Randomized proof-of-concept phase II trial comparing therapy based on tumor molecular profiling versus conventional therapy in patients with refractory cancer: Results of the feasibility part of the SHIVA trial. Br J Cancer. 2014 Apr 24. doi: 10.1038/bjc.2014.211.