

GalaxyDx : A Web-server dedicated to cancer diagnosis data analysis

Vivien Deshaies

U900 Institut Curie, INSERM, Mines ParisTech
Paris, France

National Context

- INCa : Institut National du Cancer
 - Group of “fight against” cancer partners
- Project : NGS structuring for cancer diagnosis in France
 - First part : NGS implementation in diagnostic labs
 - Second part : Bioinformatics analysis
 - Knowledge sharing
 - Pipeline building and sharing
 - Biologist / clinician Training
- Institut Curie bioinformatics platform
 - Referent team on bioinformatics analysis and training



Local Context : diagnosis at Institut Curie

- 2 diagnostic teams
 - Germ line genetics : Diagnosis on *BRCA 1 – 2*
 - Somatic genetics : Colon / Lung cancer diagnosis
- Technology : IonTorrent (PGM)
 - 1-2 run / week → 3-4 run / week
- No dedicated bioinformatician

- Diagnosis at bioinformatics platform :
 - 1 FTE :
 - 1 bioinformatician for analysis : 50 %
 - 1 bioinformatician for development : 50 %
- Requires easy communication between bioinformaticians and diagnostic teams

Galaxy advantages

Advantages for clinicians

- Easy to use
 - Suitable for training
- Choose most appropriate tools
- Tune tools parameters
- Run test analysis
- Bioinformaticians understanding










Advantage for bioinformaticians

- Ease pipeline building :
 - More reliable feedbacks → Better solutions / adaptations

GalaxyDx tools

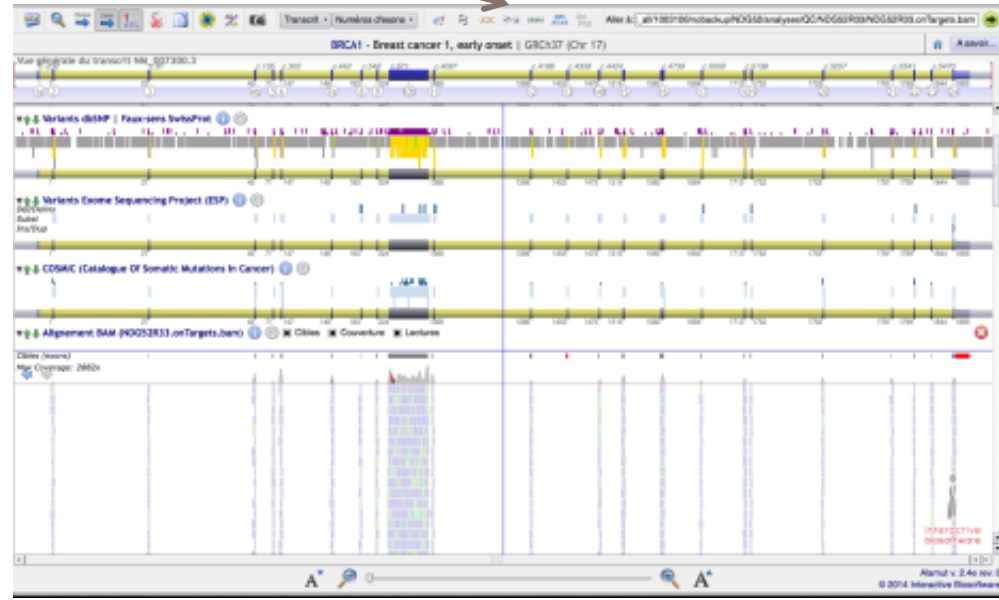
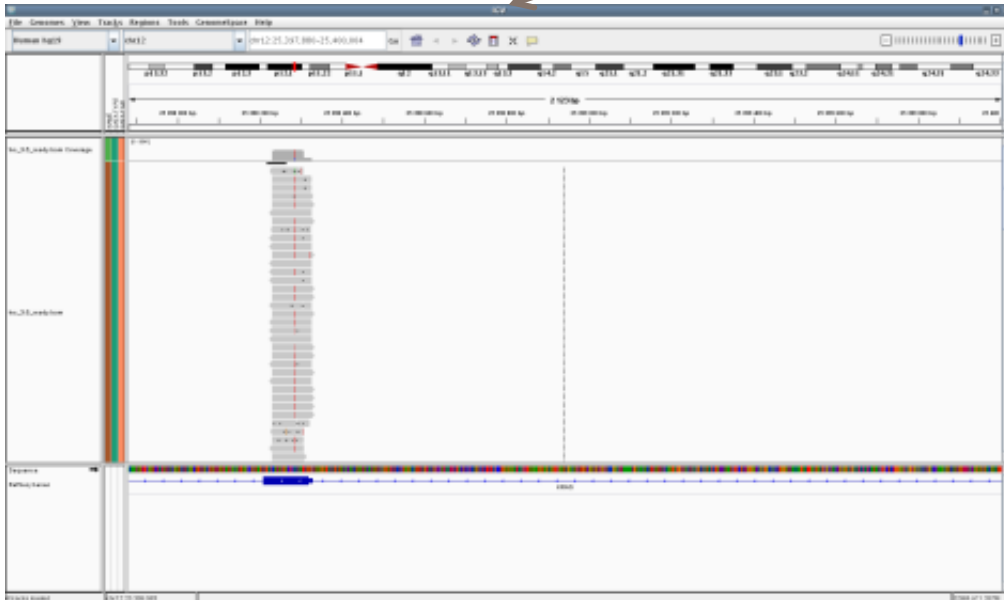
- Classic NGS analysis tools :
 - Bowtie, BWA, GATK, VarScan, Annovar ...
- Torrent suite (3.6.2 and 4.0.2) : tmap and TorrentVariantCaller
- Analysis tool :
 - Large rearrangement detection :
 - Bedtools + DESeq → foldchange graphs
- QC tool :
 - Coverage maker → table of coverage on target
- Summary tool :
 - Table maker :
 - Vcf + annovar + Scores (Grantham, MaxEntScan, ESR) → Human readable variant summary table

GalaxyDx : Visualize aligned data and variants in IGV or Alamut

30: SAM-to-BAM on data 29: converted BAM   
 16.3 MB
 format: bam, database: hg19
 Chargement du fichier de environnement Samtools
 Version: 0.1.19-44428cd SAM file converted to BAM
     
 display with IGV [local](#)
 display with Alamut [local](#)
 Binary bam alignments file

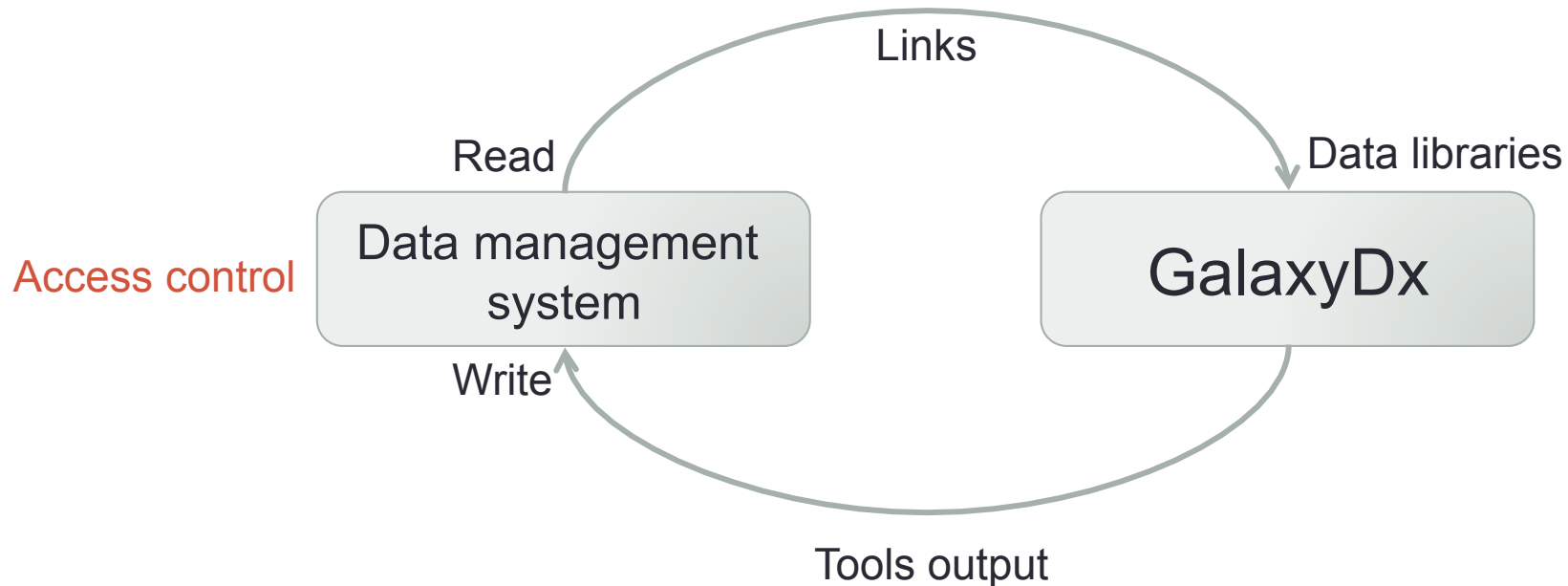


Alamut Visual



Patient data protection in GalaxyDx

- Institut curie data management system:
 - Data management by project
 - Specific access right for each project



GalaxyDx current state

- First time online (local network) : February 2014
 - Trainings :
 - Internals :
 - Past : April 2014
 - Planned : September 2014
 - INCa project : In preparation
- Useful feedbacks
- Already in use locally for pipeline tuning (*BRCA 1-2*)

Future objectives

- Share tools
 - Global tool shed
 - IFB tool shed
- Distribute GalaxyDx :
 - Virtual machine (virtualbox, docker, AMI...)
 - Github / bitbucket repository
- Create screencasts for diagnostic analysis in galaxy
- Integrate tools for diagnostic analysis with RNA-seq data

Galaxy at Institut Curie

- Currently online :
 - Local network :
 - Galaxy Institut Curie
 - GalaxyDx
 - Public instances :
 - Nebula (ChIP-seq)
- Under construction :
 - Galaxy Institut Curie public
 - Classic NGS tools
 - Nebula tools
 - GalaxyDx tools
 - Screencasts, tutorial, toy data

 **Galaxy / Institut Curie** **Galaxy / DX / Institut Curie** **nebula**
curie.fr

Aknowledgements

Genetics team :

Julien Tarabeux

Agnès Collet

Claude Houdayer

Etienne Rouleau

Dominique Stoppa-Lyonnet

Bioinformatics platform :

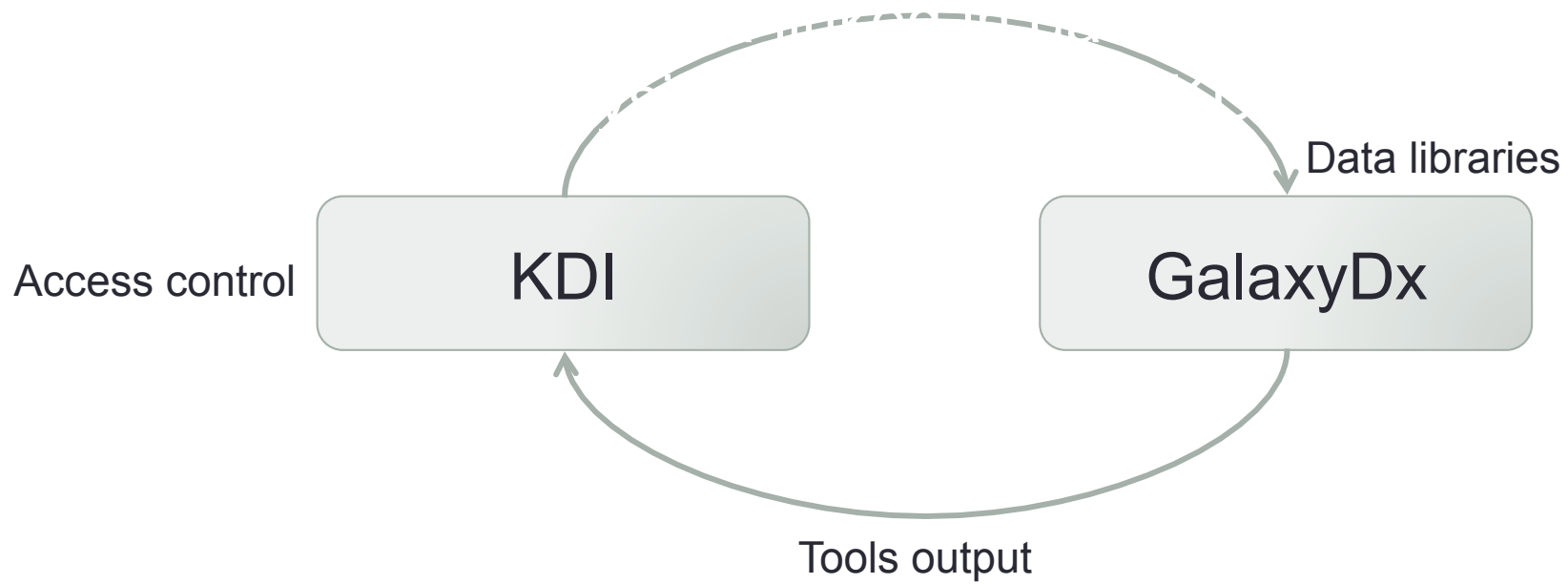
Elodie Girard

Alban Lermine

Nicolas Servant

Philippe Hupé

Emmanuel Barillot



Galaxy benefits

- Easy to use
 - Suitable for training
- Galaxy benefits for clinicians :
 - Understand bio-informaticians
 - Choose most appropriate tools
 - Tune tools parameters
 - Run test analysis
- Ease pipeline building for bio-informaticians :
 - More reliable feedbacks → Better solutions / adaptations