

Les Données de la recherche dans l'écosystème des publications scientifiques



- Le 12 février 2014
 - Institut Pasteur
 - Les besoins des chercheurs en Sciences de la Vie et de la Santé: exemple de l'Immunologie
 - GC Faure, MD, PhD, PUPH
 - Faculté de Médecine, Univ Lorraine
 - UMR CRAN/CNRS 7039
-

Qui suis-je?



- Un utilisateur final, « enduser » de l'IST depuis longtemps
 - SCICOM Nancy
 - Chargé de Mission à la documentation UHP Nancy
 - ASSIM, SFI, IUIS, ICSU
 - Website, Scientific Curation
 - Un Immunologiste clinicien engagé en cytomique
 - Des Workshops HLDA
 - aux CTCs et CECs

Quelles données ? « Big Data »

- Mathématiques, Physique, Astronomie...
 - Des chiffres et des images (numérisées!)
 - En SBS?
 - Sciences de la Vie: Gènes, Molécules, Cellules, Tissus... Données sans risque mais pas sans impact
 - Sciences de la Santé: Données à risque?!
-

En Europe

- **IMGT/Montpellier**

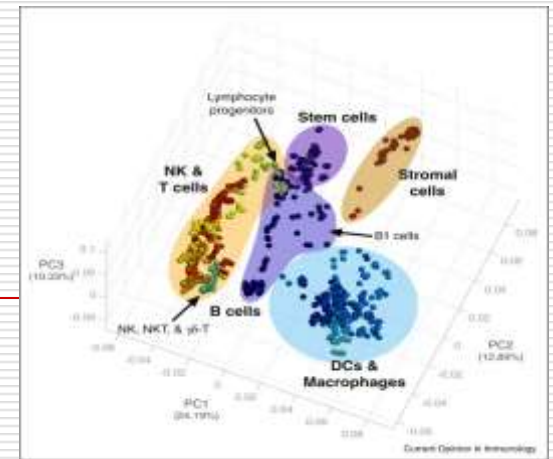
IMGT®, the international ImMunoGeneTics information system®

<http://www.imgt.org>, is the global reference in immunogenetics and immunoinformatics, created in 1989 by Marie-Paule Lefranc ([Université Montpellier 2](#) and [CNRS](#)). IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T cell receptors (TR), major histocompatibility (MH) of human and other vertebrate species, and in the immunoglobulin superfamily (IgSF), MH superfamily (MhSF) and related proteins of the immune system (RPI) of vertebrates and invertebrates.

Estimated size of immune repertoire: 10^{10}



Genes, aux USA



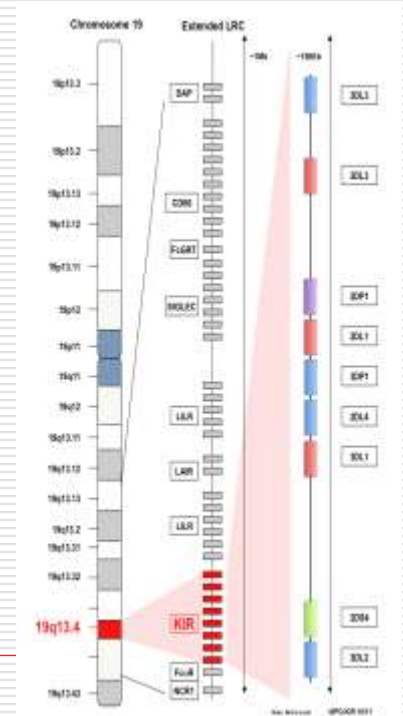
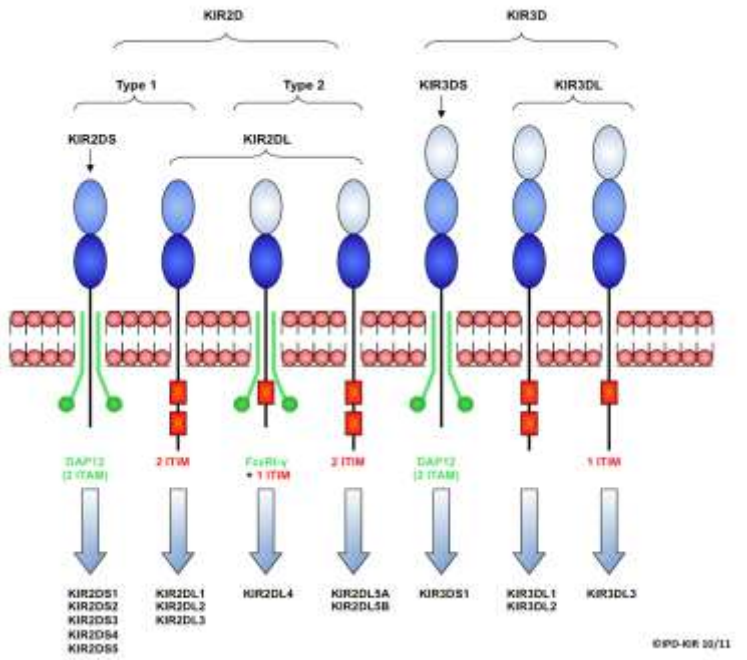
- ENCODE
 - 75% and not 3% of genome can be transcribed
- Immunological GENOME project
 - Curr Opinion Immunol 2013;25:593-7
 - Trends immunol 2013;34:602-9
 - Microarray-based transcriptomes for most of the well-defined leukocyte subsets of the C57BL/6 mouse

Immunité innée et cellules NK

IPD-KIR

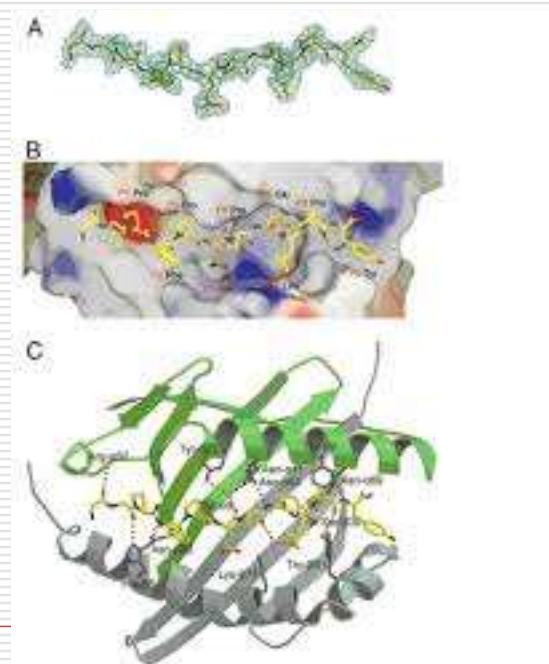
- Molécules:

The database provides a centralised repository for human KIR sequences. Killer-cell Immunoglobulin-like Receptors (KIR) have been shown to be highly polymorphic at the allelic and haplotypic level. KIRs are members of the immunoglobulin superfamily (IgSF) formerly called Killer-cell Inhibitory Receptors.



Molecules

- Molecules:
Immunopeptidome
 - Full repertoire of peptides bound to the HLA molecules, in health and disease
 - Mol Cell Proteomics 2011;10
 - Mass spectrometry technologies



Cellules

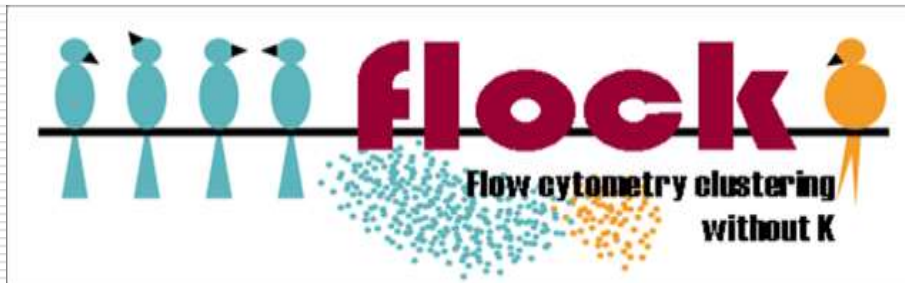
Projets internationaux



- Immport: Bioinformatics for the future of Immunology

The Immunology Database and Analysis Portal (ImmPort) system was developed under the Bioinformatics Integration Support Contract (BISC) Phase II by the Northrop Grumman Information Technology Health Solutions team for the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID), Division of Allergy, Immunology, and Transplantation (DAIT)

MHC + Cytometry...FLOCK: Flow cytometry clustering without K; + CyTOF Analysis of human T cells

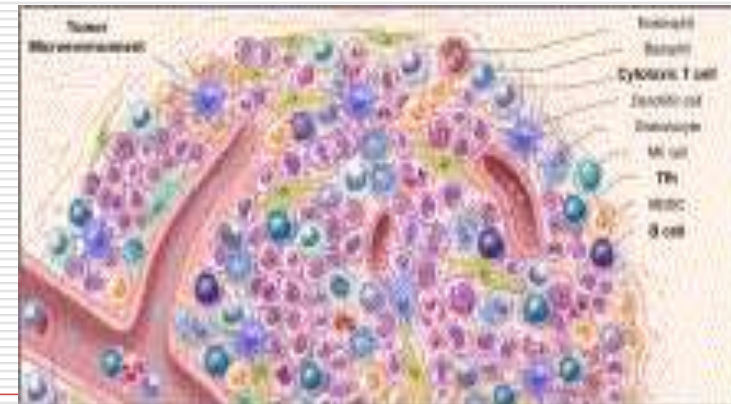
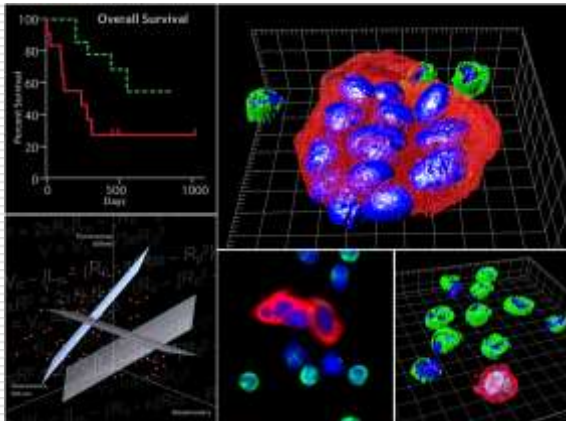


Cellules et Tissus

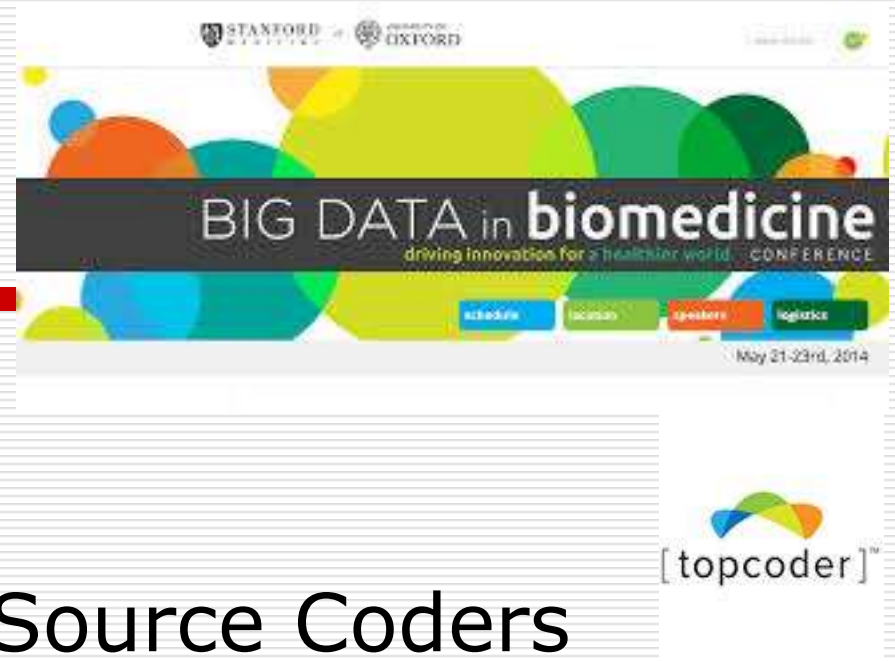
EPIC Sciences (San Diego)

A new kind of cytology detecting rare cells (CTCs and CECs) on blood slides, allowing to introduce the concept of fluid biopsy.

- Immunome (tumeurs):
 - Immunity, [Volume 39, Issue 4](#), 631-632, Oct 2013
 - Tumors are complex masses containing not just neoplastic cells but also stromal cells, neovasculature, and a gamut of immune cells.



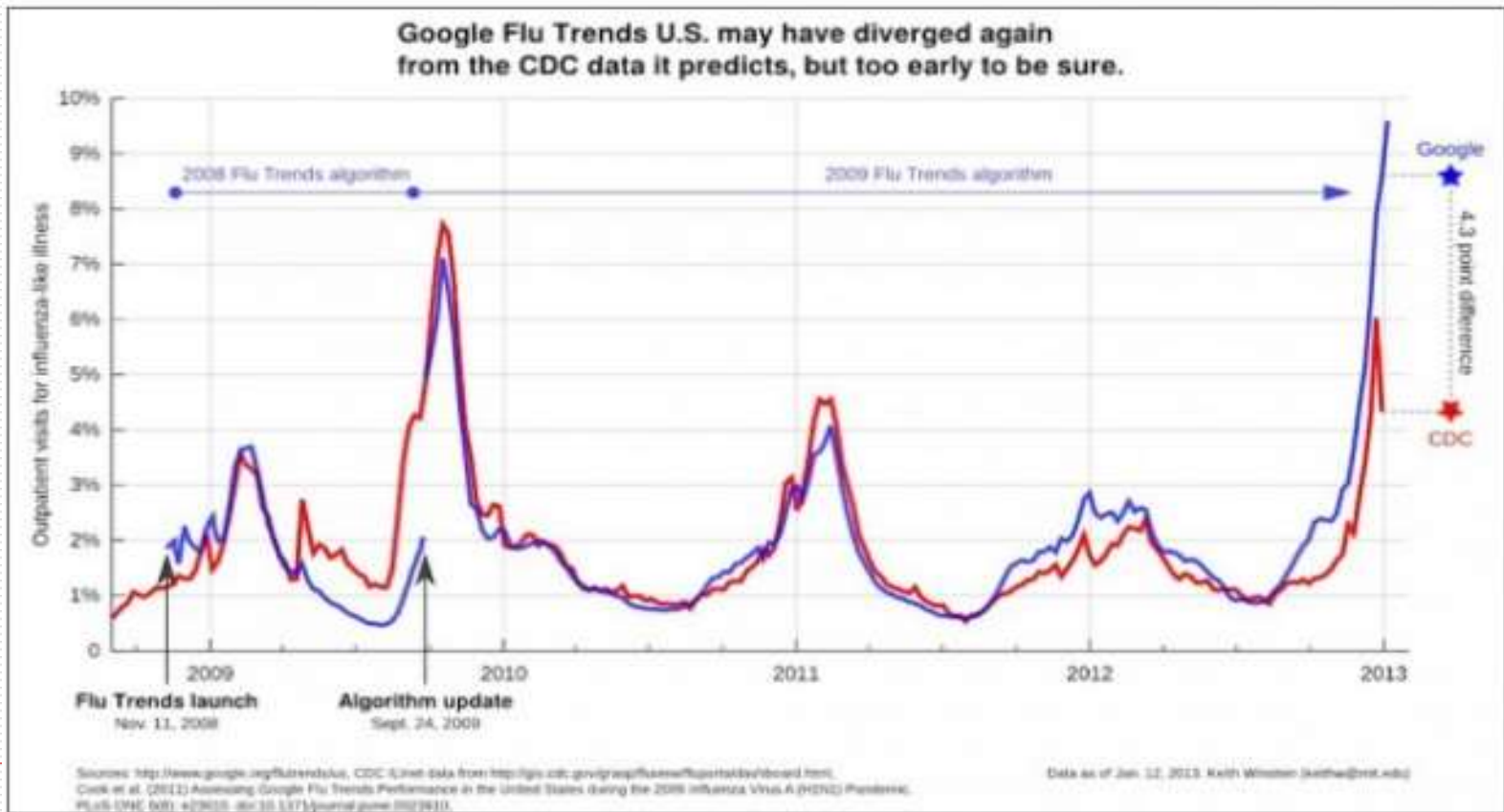
« Open... »



- Data
- Sources
- Software: CrowdSource Coders
 - Nature Biotechnology Feb 2013
 - TopCoder.com
- « Big Data in biomedicine » FF Costa
- Drugdiscovery Today 2014, in press
 - Opportunity for future personalized medicine programs

Flu : Google trends

- CDC



Google in blue, CDC in red. Note the dramatic divergence toward 2013. (Keith Winstein, MIT)

Sciences de la Santé: Recherche clinique et big data

- Du papier aux e-CRFs
- GEIL, EGIL, ELN...
- GOELAMs
 - e-CRFs
 - ARCs, TECs
- A risque! Médical, données patients
 - Secret, confidentialité
- Impact
 - Évaluation de protocoles cliniques, données pharmacoéconomiques



UK/(USA): Clinical research

- General Practice Research Database
 - 15 years..
- CPRD / NHS
 - Asthma
 - Cardiovascular disorders



- Patientslikeme!
- ~~SAGE/Bionetworks.. Parkinson~~

Big Data and HIV Research

AIDS 2013;27:835-8



- HIVsystemsbiology.org
 - Gene Overlapper
 - 39 lists of genes.. 500billion subsets
 - 3 genome wide siRNA screens
 - CDNA overexpression screen
 - Crowd sourcing for Mason-Pfizer monkey virus proteases
 - HIV replication Cycle Site
 - AIDS Vu project (géographie de la maladie #GoogleMap)
- ~~Publication 15091 papers in 2011:>40/day~~

Recherche clinique

Médecine personnalisée



MMRF
Multiple Myeloma
Research Foundation

- Myélome: MMRF Multiple Myeloma Research Foundation
 - Sept 2013... first and only open access data platform in myeloma

The groundbreaking MMRF CoMMpass study is now enrolling NEWLY DIAGNOSED multiple myeloma patients.

Landmark 10-year study, 1000 patients, Molecular analysis of every patient

- Personalized or Precision medicine
-

Les Besoins des chercheurs

- Outils informatiques!?

E-BIOTHON: cloud dédié à la bioinformatique

CNRS, IBM, Inria, the Institut Français de Bioinformatique and the innovative start-up SysFera are rolling out E-Biothon, an experimental Cloud platform (1) to help speed up and advance research in biology, health and the environment.

- Maîtrise des logiciels
- Accès aux données publiques et privées (Dossier Médical)
- Maîtrise des concepts de l'information scientifique
- Maîtrise de l'évolution des concepts

Besoins des futurs chercheurs

Les publications: big data and information overload

- Teaching and training of students
 - Data: Lab Archives and Notebooks
 - Papers: Back to Basics
 - Databases of scientific papers
 - From PubMed, to Researchgate, from World of science to Scopus
 - Grey Literature! (Google!)
 - Impact Factor: from citations to altmetrics
 - Knowledge: Concepts and ideas
-
- Curation?...

A few other references

- <http://www.hsph.harvard.edu/news/magazine/spr12-big-data-tb-health-costs/>
-