

INRAE

> CATI EMPREINTE

Franck Giacomoni

franck.giacomoni@inrae.fr

empreinte-contact@groupes.renater.fr

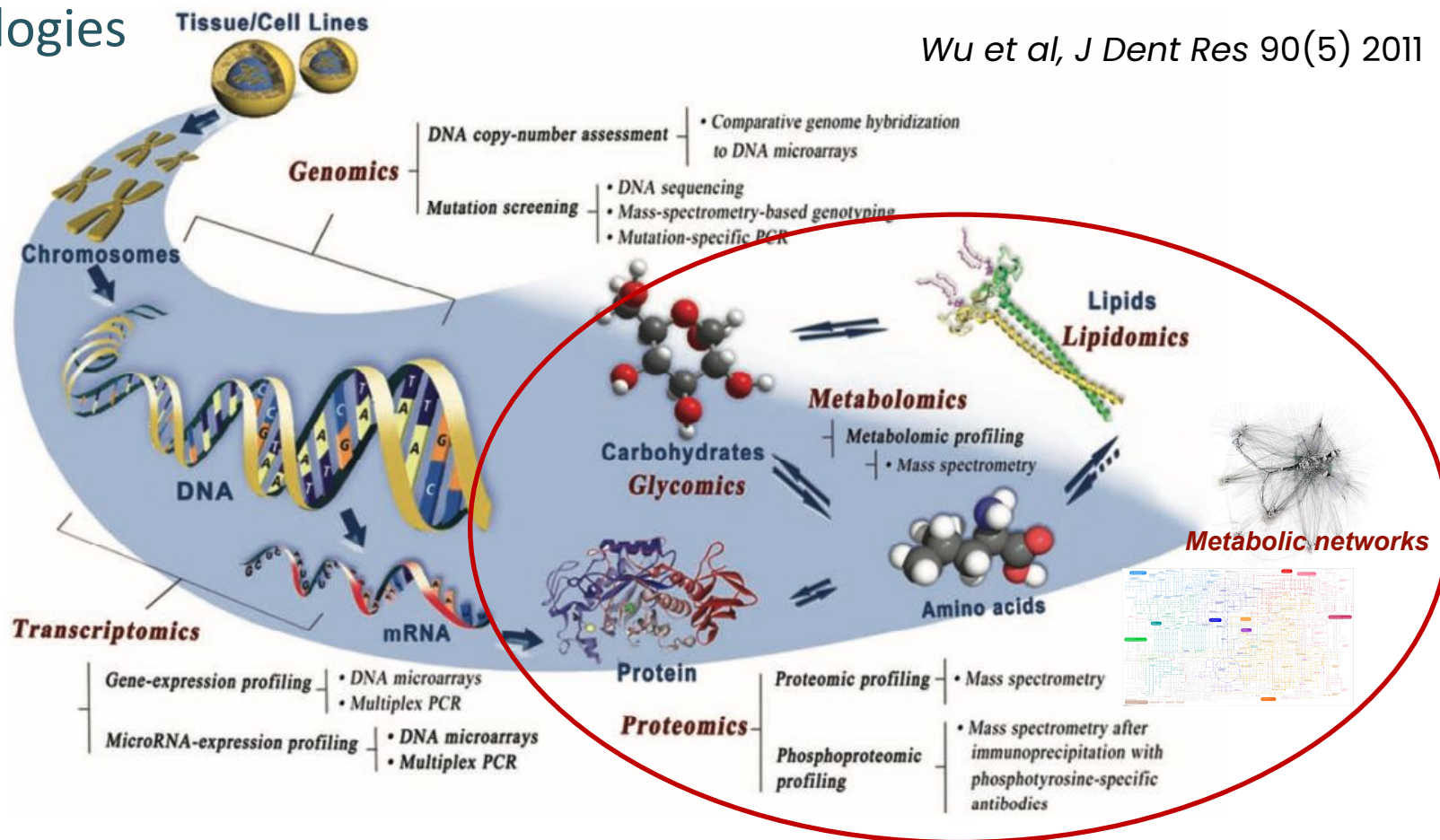
EMPREINTE « in a nutshell »

- Nouveau CATI 3G – Biologie des systèmes
- EMPREINTE OU « Molecular PhEnotyping and biochemical daTa Engineering »
- Porté par ALIMH / TRANSFORM + IR MetaboHUB
- ~20 agents (Bioinformaticiens, statisticiens, informatique scientifique, biochimie...)
- Des Animations techniques/scientifiques
- Des actions intra/inter CATI :
 - Groupe FORUM - Outils et approches du web sémantique appliqués à la métabo et à la santé
 - Groupe GalaxyExtraOmicCommunity (GEOC) : avec les CATI CODEX et IUMAN
 - Groupe « Approches intégratives » avec CATI SYSMICS
 - CHEMOMICS - École mixte interdisciplinaire Chimiométrie et Métabolomique
 - ...

Les données d'EMPREINTE

- Typologies

Wu et al, J Dent Res 90(5) 2011



- Espèces : Humain (santé, toxico, ...), souris, rat, ..., bactéries, ...
- Volumétrie : Échelle production/stockage/sauvegarde: des Tera-octets

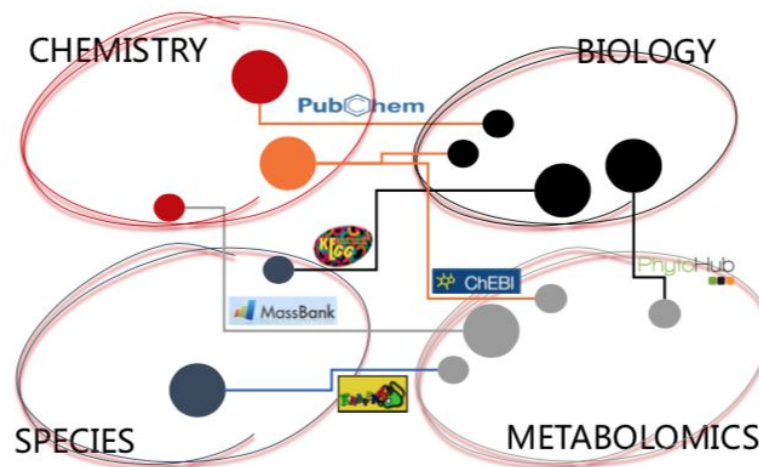
Les données d'EMPREINTE

• Ressources :

- Banques de composés chimiques et de peptides/protéines
 - *RefSeq (NCBI), UniprotKB, ...FlyBase, TAIR db, ...ChEBI, PubChem*
- Base de données spectrales : *MassBank, GNPS, MZCloud, ...*
- Dépôts de référence : NeXtprot (Swiss Institute of Bioinformatics), ProteomeXChange, MetaboLights, ...
- Réseaux Métaboliques (*KEGG, RECON, ...*)
- Bases de connaissances (PhytoHUB, Exposome Explorer, ...)

• Ontologies :

- Pride controlled vocabulary, ...
- Study plan – ISA-TAB
- ChEBI ontology (EBI)
- Ontologies spécialisée via BioPortal



Source : MySelf

Les données d'EMPREINTE

- Consortium – réseaux
 - Human Proteome Organization (HUPO)
 - Metabolomics society, Metabolic Profiling Forum, Réseau Francophone de fluxomique et de métabolomique (RFMF), IR MetaboHUB + PROFI
- Standardisations :
 - Développement de formats standard (mzML, nmrML, mzIdentML, mzQuantML, TraML, ...) et des outils open source incitant le dépôt, la réutilisation et les analyses comparatives des données.
 - Développement du [SBML](#)^{*1} pour les réseaux métaboliques
 - Designé et maintenu par [Proteomics Standards Initiative PSI](#) ou **The Metabolomics Standards Initiative** (MSI) et le Core Information for Metabolomics Reporting (CIMR)
+ projets **FP7 COSMOS**, **H2020 PHENOMENAL**.

Outils utilisés

Orienté traitement de données : DEV

Bioinformatique, statistiques, traitement signal, intégration de données

Logiciels commerciaux (acquisition données, traitement, ...)

Les langages : **R, MatLab, Perl, Python**, ... + libraires académiques

!!!! **Galaxy** !!!! Avec WorkFlow4Metabolomics.org

<https://github.com/workflow4metabolomics> & <https://bioconda.github.io/>

Orienté Système d'information : DEV

Gestion des données (projet, laboratoire, résultats, connaissances, ...)

Portail des bases de données de connaissances

Php, Java - SQL, Elasticsearch, ... RDF

Orienté visualisation de données : DEV

Visualisation de bases de connaissances (Web portal, Askomics, Sparklis, ...)

Visualisation et exploration de données (outils statistiques et graphes)

Visualisation de réseaux métaboliques (MetExplore coming soon...)

<https://metexplore.toulouse.inrae.fr>

Organisation des travaux réalisés

Le collectif EMPREINTE:

Nouveau CATI mais base solide (W4M, MetaboHUB, metabo-proteo, ...)

Distribution des agents geo + PFs/UMR + interDépart.

Hétérogénéité des expertises et domaines applications

Animations scientifiques & techniques + session « allez plus loin »

ex : tester une approche ou un package sur dif. jeux de données

Actions intra CATI

Agréger pls membres du CATI autour d'un mini projet de R&D avec/sans financement

Monter des workshops, hackathons thématiques interCATI avec le soutien d'Ingenium et de DISPO

Ecole chercheurs en collaboration avec des collectifs plus larges

Organisation des travaux réalisés

Gestion des envt d'exécution:

Poste bureautique -> infrastructures virtualisées avec serveurs applicatifs
Vmware – ProxMox
Docker
Galaxy – Clusters GenoToul & IFBcore
Pas IDE commun intraCATI

Les ressources mises à disposition par le CATI :

(WorkFlow4Metabolomics.org)

POC infrastructures pour construire des triplestores orientés
métabolomique/réseaux métaboliques & santé

Enrichir les compétences transversales :

Usage de github et gitlab pour les projets communs
Communiquer, former sur les aspects “génies logiciels”