O.D.I.N.

Omics Data Integration Network

Kévin LEBRIGAND, PhD
Institut de Pharmacologie Moléculaire et Cellulaire
Université Côte d'Azur, CNRS, France.
lebrigand@ipmc.cnrs.fr
@kevinlebrigand
@kevinlebrigand@genomic.social

Omics Data Integration Network

Overview

UNIVERSITÉ CÔTE D'AZUR

- Inter-UCA institutes of biology project 
- Inter-disciplinary multi-omics: genomics, proteomics, metabolomics, cytometry, imaging

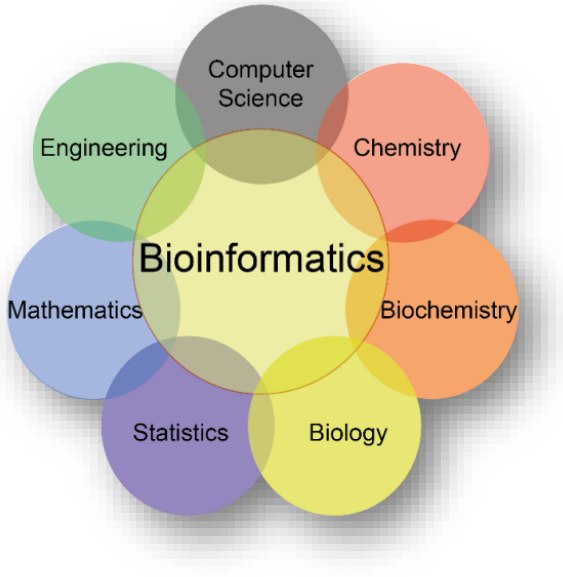
- Document and structure expertise's and resources dedicated to omics data analysis

- ODIN will be a multi-omics research project data manager (FAIR)
- ODIN will store and share reference data analysis guidelines
- ODIN will allow easy-to-start data analysis through 4D-MICS EquipEx for Biological Data infrastructure

- Optimize bioinformatics for UCA research teams scientific production

Bioinformatics

Discipline at the crossroads of biology, computer science and new technologies

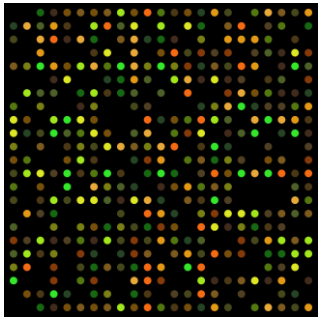


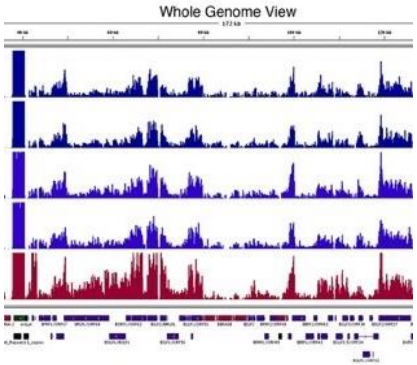
Characterized by the application of mathematical, statistical, computational methods to the analysis of biological, clinical, biochemical and biophysical data.

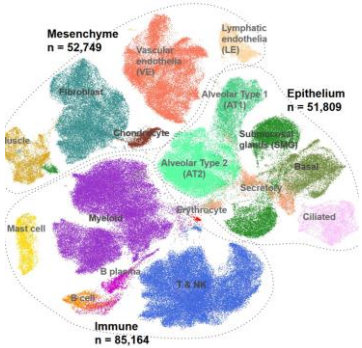
- development and implement tools to **store** and **manage** information
  manage data flows and infrastructure, database storage, web services implementation
- **analysis** and **interpretation** of data to identify relevant information
  analysis workflows setup, scripting analysis, figures production for publications, PI support
- development of new algorithms and statistical tools (**computational biology**)
  complex programing language, new signal mining, academics research in mathematics (AI)

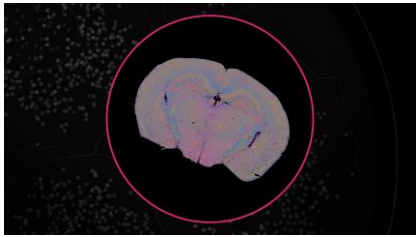
Example of 20 years of transcriptomics

Driven by microfluidics technological developments









Early 2000's: DNA microarray

- Large-scale transcriptome
- Oligonucleotide probe tilling
- Fluorochrome signal analysis
- Bulk resolution

Cost : 4k€
20 samples
25k genes
0,5M matrix

Late 2000's: RNA sequencing

- Whole transcriptome
- Next Generation Sequencing
- Full-transcript coverage
- Bulk resolution

Cost : 4k€
20 samples
50k genes
1M matrix

Mid 2010's: Single-cell

- Whole transcriptome
- Microfluidics + NGS
- 3p-end gene signal (UMI)
- Sensitivity (6%)
- Single-cell / state resolution

Cost : 4k€
5k cells
50k genes
250M matrix

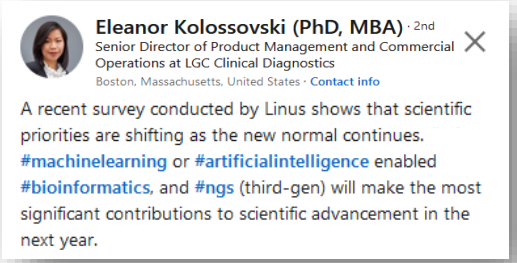
2020's : Spatial

- 500-1000 gene targets
- Imaging analysis
- Multiplexing FiSH (single molecule)
- Sensitivity (30-80%)
- Sub-cellular resolution

Cost : 4k€
250k cells
1k genes
250M matrix
+ Spatial dimension

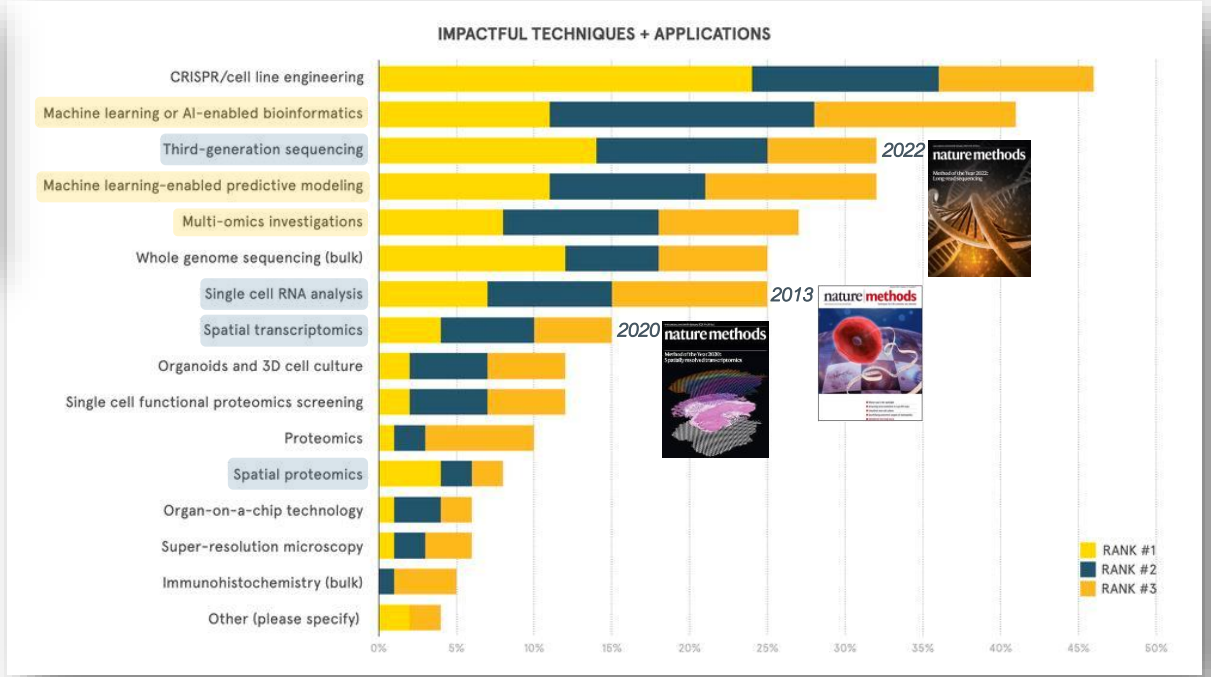
Academics labs scientific priorities

Requires expertise in bioinformatics and computational biology



Eleanor Kolossovski (PhD, MBA) · 2nd
Senior Director of Product Management and Commercial Operations at LGC Clinical Diagnostics
Boston, Massachusetts, United States · Contact info

A recent survey conducted by Linus shows that scientific priorities are shifting as the new normal continues. #machinelearning or #artificialintelligence enabled #bioinformatics, and #ngs (third-gen) will make the most significant contributions to scientific advancement in the next year.



Are UCA core facilities and research teams ready to process and analyze all those data ?

4D-Omics (Equipex+, PIA3)

Instrument numérique pour la biologie quantitative multi-échelle en région Sud

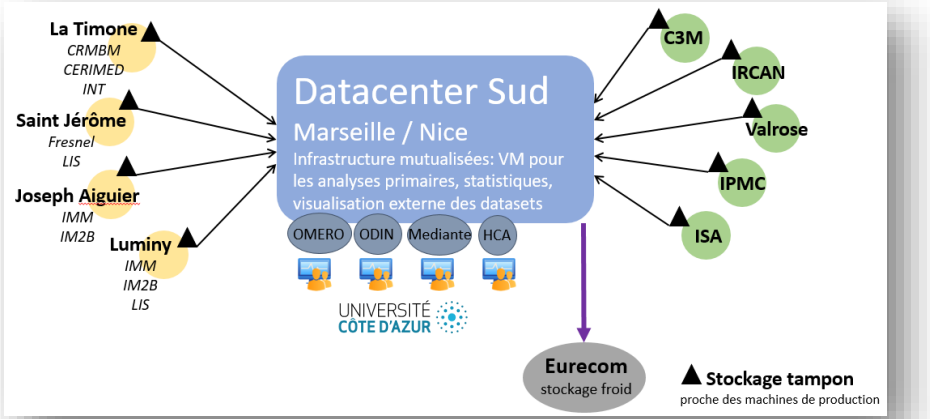
UNIVERSITÉ CÔTE D'AZUR

4D-OMICS EquipEx for Biological Data — FRANCE 2030

Structurer la composante biologique de DATASUD

Université Côte d'Azur, Université d'Aix Marseille

- Coordinateur: Pascal Barbry
- Projet: oct.2021 - jun.2029
- Budget: 6.595.999 €



✓**Convaincre les laboratoires de biologie des avantages d'une migration vers des ressources numériques distribuées**:

- Économie de moyens, optimisation des moyens en personnel
- Recentrage des laboratoires sur l'exploitation de la donnée

✓**Illustration de la démarche sur quelques programmes phares**:

- *Human Cell Atlas*: un consortium international fortement structuré, basé sur la distribution rapide et large de données potentiellement sensibles
- *Omero*: un système d'information pour les données d'imagerie
- *Mediante*: un système d'information pour les données de génomique
- *ODIN*: centralisation des scripts et objets d'analyses des données de biologie

✓**Entrepôts de données**:

- Assurer la pérennité des données du site (5 Pb bandes Eurecom, may 23)
- Travail sur les données de santé (CHU)

Omics Data Integration Network

Data analysis sustainability and expertise sharing between UCA research teams

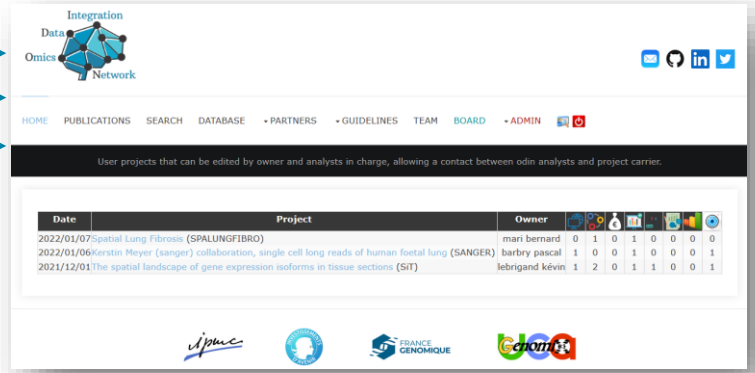
UNIVERSITÉ CÔTE D'AZUR

UCA omics resources map

Reference analysis guidelines

UCA facilities quantitative data

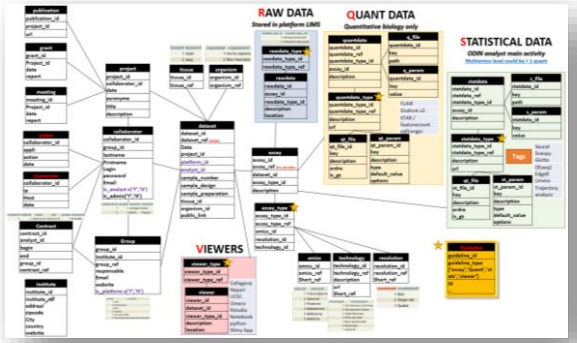
Genomics
Cytometry
Proteomics
Metabolomics
Imaging



PostgreSQL

System files

Researchers
Bioinformaticians
Ready to start interactive analysis sessions



- ❑ User/team-oriented **research project manager**
- ❑ Multi-omics datasets **FAIR** handling
- ❑ Statistical analysis scripts and objects storage (re-use)
- ❑ Reference **guidelines** for "omics" data analysis
- ❑ Internal and external publications integration capability
- ❑ **Expertise sharing within UCA community**
- ❑ External viewers availability

https://www.genomique.info:8443/odin/index

Omics Data Integration Network

Step-by-step

UNIVERSITÉ CÔTE D'AZUR

1. Draw the UCA omics resources map (production systems and expertises)

ipmc

Core facilities / engineer staff		*Omics production systems*	*Omics assays*	*Data analyst expertise*
Genom'UCA	V.Magnone G.Rios M.Couralet		Short-read seq Long-read seq Spatial RNA-seq Single-cell RNA-seq 3', 5' Single-cell multi-ome …	Kevin Lebrigand (IR) Marin Truchi (CDD) Yvon Mbouamboua (CDD) Antoin Collin (PhD) Mariem Ben khedher (CDD) Christophe Becavin (UCA) Romain Gautier (UCA) Dominique Douget (UCA)
CAPABIO	D.Debayle A.S.Gay L.Fleuriot		Metabolomics Proteomics ….	
MICA Microscopie Imagerie Côte d'Azur	F.Brau S.Abelanet		Imaging applications	
Cytometry	J.Cazareth		Tri cellulaire Phénotypage	

Omics Data Integration Network

Step-by-step

UNIVERSITÉ CÔTE D'AZUR

2. Document internal or external reference analysis guidelines

ijpmc

Core facilities / engineer staff

Genomic — V.Magnone, G.Rios, M.Couralet

CAPABIO — D.Debayle, A.S.Gay, L.Fleuriot

MICA Microscopie Imagerie Côte d'Azur — F.Brau, S.Abelanet

Cytometry — J.Cazareth

Omics production systems

Omics assays

Short-read seq
Long-read seq
Spatial RNA-seq
Single-cell RNA-seq 3', 5'
Single-cell multi-ome
…

Metabolomics
Proteomics
….

Imaging applications

Tri cellulaire
Phénotypage

Data analyst expertise

Kevin Lebrigand (IR)
Marin Truchi (CDD)
Yvon Mbouamboua (CDD)
Antoin Collin (PhD)
Mariem Ben khedher (CDD)
Christophe Becavin (UCA)
Romain Gautier (UCA)
Dominique Douget (UCA)

Analysis workflow

- Quality Control
- Data filtration
- Normalization
- Dimension reduction
- Data exploration
- Differential analysis
- Integration
- Figure production
- Public repository sub
- Viewers

Reference publications (external / internal)
Guidelines

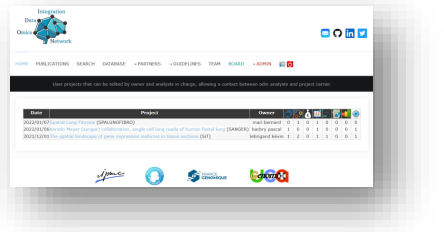
Omics Data Integration Network

Step-by-step

UNIVERSITÉ CÔTE D'AZUR

3. Implement a multi-omics research project FAIR data manager system

ODIN web portal



- Java J2EE development
- Hosted by Tomcat app server
- PostgreSQL database

Storing / Managing

- User/team-oriented research project manager (FAIR)
- Reference guidelines for “omics” data analysis

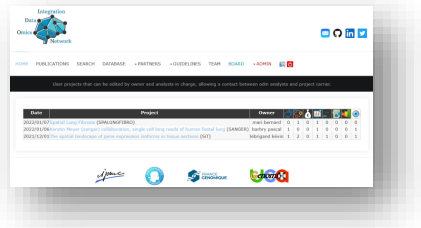
Omics Data Integration Network

Step-by-step

UNIVERSITÉ CÔTE D'AZUR

4. Start bioinformatics analysis session based on new data and shared guidelines

ODIN web portal



- Java J2EE development
- Hosted by Tomcat app server
- PostgreSQL database

Storing / Managing

- User/team-oriented research project manager (FAIR)
- Reference guidelines for "omics" data analysis

Researchers
Bioinformaticians
Start interactive analysis sessions

Bioinformatics analysis

- RStudio server
- Jupyter Notebook server

Technical solutions

- Local export archives (data + analysis scripts)
- Container: Kubernetes, Docker, Singularity
- *Virtual machines*

4D-MICS
EquipEx for Biological Data

UNIVERSITÉ CÔTE D'AZUR

UCA Research Engineer - IDEX academy 4 (September 2023)

Acknowledgments

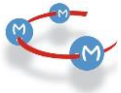
Omics Data Integration Network

Olivier Croce
Gaël Cristofori
Boris Meyer
Claude Philippe
Erwan Simon

Arnaud Hubstenberger
Thierry Lepage
Antoine Fortuné
Luc Martin

Kevin Lebrigand
Marin Truchi

Etienne Danchin
Corinne Rancurel
Martine Da-Rocha
Arthur Péré



Michele Trabucchi
Maria Stathopoulou

IPMC bioinformatics platform

Structure and organize bioinformatics resources and developments @ IPMC level

Committed to support IPMC research teams and facilities

- IR recruitment (CDD, 1 year, march 2023)
- Support experimental design setup, funding requests and paper methods writing
- Constant technological watch in the priority technological areas defined by the institute
- Support bioinformatics @ IPMC (internal seminars)
- French Institute of Bioinformatics (IFB) subscription (visibility, ease for future recruitment)
- Dedicated space to group bioinformaticians recruited by teams and facilities
  - Group work emulation
  - Define standardized workflows for IPMC equipment in Genomics, Proteomics, Cytometry and Imaging
  - Mentoring, best practices and guidelines sharing
  - Expertise / methodological developments shared between teams
  - Provide an optimal working environment, IT infrastructure (4D-MICS EquipEx for Biological Data)

UCA omics ecosystem

Cartography of UCA resources and expertise in omics data production and analysis

UNIVERSITÉ CÔTE D'AZUR

Cartographie des ressources et expertises omiques des instituts de biologie UCA

En relation avec les différents partenaires du réseau, réaliser une cartographie des ressources des instituts partenaires UCA en lien avec la production et l'analyse des données de biologie quantitatives: bioinformaticiens, expertises, systèmes de production des plateaux techniques, infrastructure informatique, équipes de recherches potentielles partenaires.

ipmc

Core facilities / engineer staff		*Omics production systems*	*Omics assays*	*Data analyst expertise*	*Informatics infrastructure*
Genom'UCA	V.Magnone G.Rios M.Couralet		Short-read seq Long-read seq Spatial RNA-seq Single-cell RNA-seq 3', 5' Single-cell multi-ome …	Kevin Lebrigand (IR) Marin Truchi (CDD) Yvon Mbouamboua (CDD) Antoin Collin (PhD) Mariem Ben khedher (CDD) Christophe Becavin (UCA) Romain Gautier (UCA) Dominique Douget (UCA)	*Caire (serveur web)* Intel Xeon - 8 cores 48 Gb RAM 29 To storage *Joyo (calcul cluster)* Intel Xeon - 8 cores 48 Go de RAM, 15 + 22 To storage SGE x12 nodes (20 cores, 64 Gb RAM) *Bego (stats server)* 2 x Intel Xeon Gold 6248 - 40 cores GPU A100 Ampere 2 To RAM 37 To storage
CAPABIO	D.Debayle A.S.Gay L.Fleuriot		Metabolomics Proteomics ….		
MICA Microscopie Imagerie Côte d'Azur	F.Brau S.Abelanet		Imaging applications		
Cytometry	J.Cazareth		Tri cellulaire Phénotypage		

- *Réalisation du même travail dans les autres instituts de biologie UCA*
- *Présentation d'une cartographie évolutive au sein d'un portail web commun*

IRCAN · Institut SOPHIA AGROBIOTECH · Institut de Biologie Valrose