



First edition 2023 in Fréjus



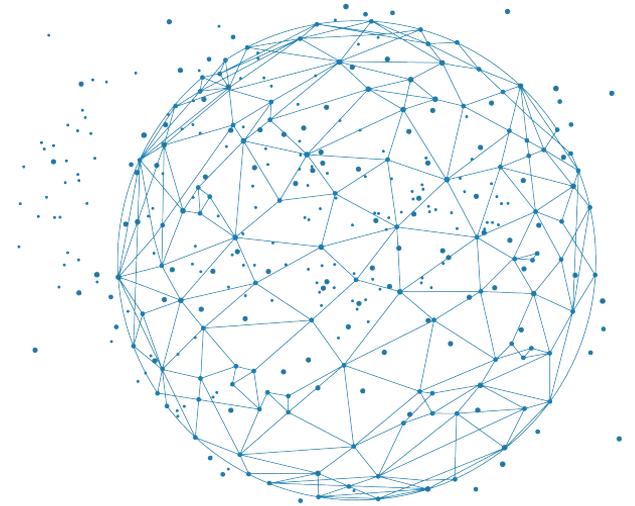
# École thématique en bioinformatique Intégrative

Hélène Chiapello, Olivier Sand  
& Lucie Khamvongsa-Charbonnier

DOI version final



# Introduction





## Buts

- **Intégrer différents jeux de données**, qui proviennent de différents niveaux d'analyse dans la cellule : (meta)génomés, (meta)transcriptomes, (meta)protéomes, métabolomes, réseaux d'interactions...
- **Extraire les informations pertinentes** à partir de ces différents jeux de données, pour ensuite les interpréter en utilisant des modèles intégratifs

> Un domaine de recherche en bioinfo

> L'intégration de données nécessite de maîtriser **des approches bioinformatiques qui résultent de différents domaines**

- Analyses multivariées
- Réseaux
- Web sémantique

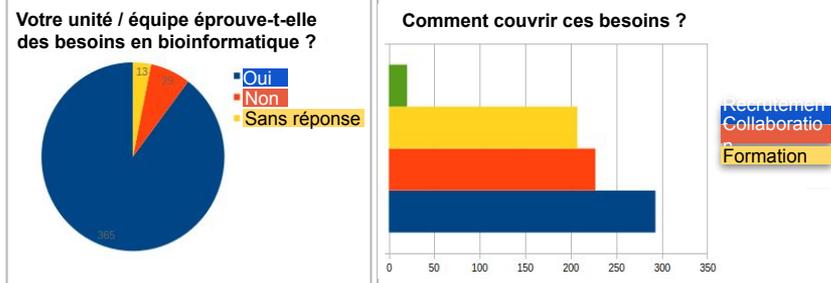
## Une **forte demande** de formation en bioinformatique intégrative

- pour les biologistes et les bioinformaticiens
- pour tous les métiers et niveaux
- dans différents domaines d'application : biologie, santé, agronomie, ...

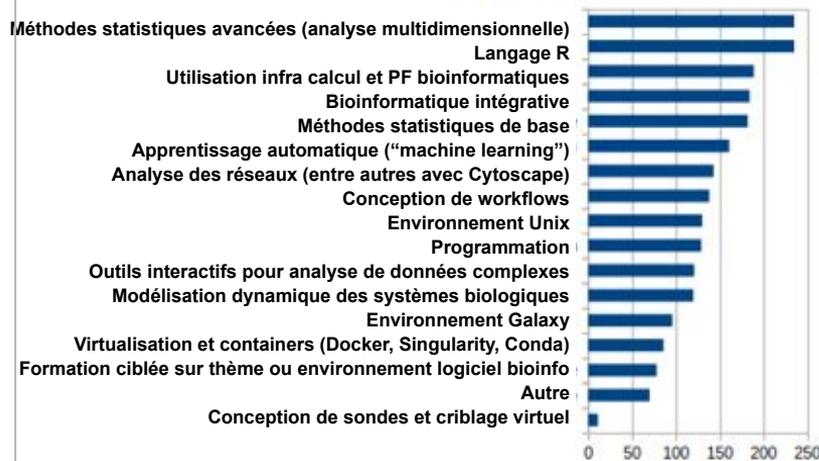
## Stratégie :

- former en priorité
  - des bioinformaticiens/biostatisticiens
  - des futurs formateurs
- mutualiser des ressources pédagogiques

## Enquête IFB 2019 auprès des équipes/unités en France (N=407)



## Vos besoins en matière de formations à l'analyse bioinformatique





La formation a pour but :

- d'introduire les **concepts de bases** et les différents types d'approches utilisées en bioinformatiques intégrative
- de proposer un **approfondissement et une mise en pratique d'une de ces approches** sur un/des jeux de données intégrant différents types de données omiques, via 3 ateliers en parallèle : analyses multivariées, réseaux, web sémantique
- de **créer, améliorer et partager les ressources pédagogiques** (supports de formation, jeux de données, tutoriels) sur le thème de la bioinformatique intégrative



A la fin de cette formation les participants :

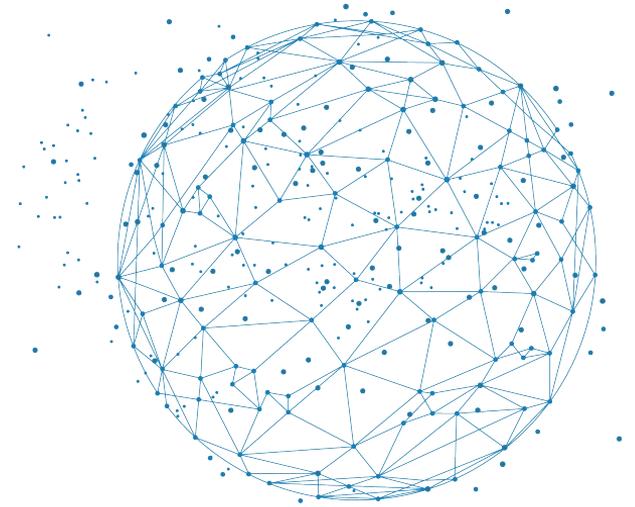
- auront acquis **un socle de connaissances générales en bioinformatique intégrative**
- auront **mis en oeuvre une analyse intégrative** depuis la préparation des données jusqu'à l'analyse critique de résultats sur un/des jeux de données proposés lors de la formation
- auront **contribué à constituer du matériel pédagogique partagé** sur le sujet



Librement inspiré de : <https://openlifesci.org/code-of-conduct>

- Soyons amicaux et patients. Nous sommes ici entre collègues, demain vous serez (peut-être) formateur
- Soyons accueillant, pas de discrimination d'aucune sorte ici
- Soyons prévenants, respectez le travail des autres
- Soyons respectueux, restons calme et courtois en cas de désaccord
- Essayons de comprendre pourquoi nous ne sommes pas d'accord
- Essayons de résoudre les points de désaccord et divergence de manière constructive
- Par ailleurs l'erreur est humaine... et reproductible (en général)
  
- **Toutes les questions sont les bienvenues !**

# Présentation rapide de l'IFB



# IFB – une fédération d'équipes et de plateformes

## Missions

- Services
- Vision stratégique
- Représentation internationale (ELIXIR...)

## 36 plateformes et équipes associées

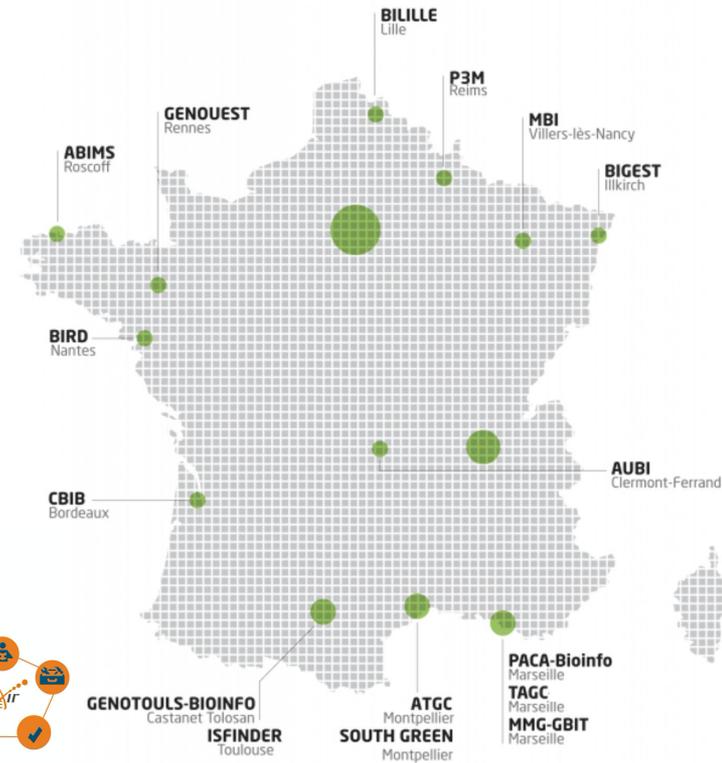
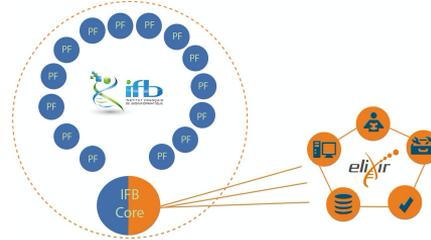
- En lien avec de nombreux instituts et universités

## De nombreux services

- Calcul
- Stockage
- Formations
- Accompagnement de projets

## IFB-core (UMS 3601)

~400 experts  
(~200 ETPs)



### RÉGION PARISIENNE

**EBIO** Paris  
**INSTITUT CURIE** Paris  
**IGR** Villejuif  
**MICROSCOPE** Evry  
**MIGALE** Jouy-en-Josas

### C3BI

**RPBS** Paris  
**URGI** Versailles  
**ORPHANET** Paris  
**ICONICS** Paris  
**IFB CORE** Evry

### RÉGION LYONNAISE

**INCA-SLC** Lyon  
**PRABI-HCL** Lyon  
**PRABI-AMSB** Villeurbanne  
**PRABI-Lyon-Grenoble** Villeurbanne  
**PRABI-Lyon-Gerland** Lyon



# L'infrastructure de calcul de stockage (ou NNCR)

- 2 infrastructures complémentaires: **Cloud** et **Cluster**
- Services répartis sur 12 sites
  - 2 serveurs nationaux (IFB-core-cluster & IFB-core-cloud)
  - 10 serveurs régionaux
- Mutualisation des efforts humains : **NNCR task force**
  - Chaque plateforme partage des ingénieurs (10% à 50% ETP)
- **Bénéfices partagés**
  - Eviter les redondances
  - Gestion collective des problèmes
  - Partage d'expérience
  - Procédures robustes et transposables
  - Forte émulation des ingénieurs
  - Rompre la solitude de l'ingénieur info dans son labo bio

8 cloud sites  
 > 8,000 vCPUs  
 > 1.5Pb storage  
 > 36 TB RAM

6 clusters  
 > 16,000 vCPUs  
 > 16Pb storage  
 > 92 TB RAM



- Hardware**
- 6 SLURM Clusters
  - 16,900 vCPU
  - 92 TB RAM
  - 10.5 PB
  - 10 GPU cards
  - 4935 yearCPU in 2020

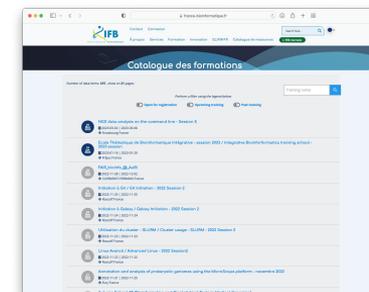
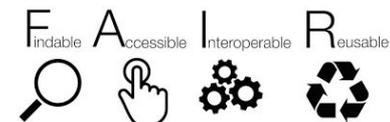
Plateforme	Localisation	Calcul (#CPU HT*)	Stockage (#TB)	RAM
<b>ifb-core-cloud</b>	<b>Lyon (CC-IN2P3)</b>	<b>3936</b>	<b>408</b>	<b>20 408</b>
AuBi	Clermont-Ferrand	200	20	800
GenOuest	Rennes	600	350	2600
PRABI	Lyon	696	144	3300
BiRD	Nantes	860	150	2 500
BIGEst	Strasbourg	1024	500	4 000
BILILLE	Lille	192	0	768
CBP-PSMN	Lyon	768	24	1920
<b>Total fédération de clouds</b>		<b>8276</b>	<b>1596</b>	<b>36296</b>

Plateforme	Location	Compute (#CPU HT*)	Storage (#TB)	RAM (#GB)	RAM/core (#GB)	RAM/core (#Cards)
<b>ifb-core-cluster</b>	<b>Orsay (IDRIS)</b>	<b>4300</b>	<b>2400</b>	<b>20008</b>	<b>4,65</b>	<b>9</b>
ABiMS	Roscoff	2600	2500	10600	5,50	2
GENOTOUL	Toulouse	6224	8000	36500	5,56	1
GenOuest	Rennes	1866	2300	11616	6,23	7
BiRD	Nantes	902	600	6800	6,79	-
MIGALE	Jouy en Josas	1016	350	7000	6,89	-
<b>Total IFB clusters</b>		<b>16908</b>	<b>16550</b>	<b>92524</b>	<b>35,64</b>	<b>19</b>

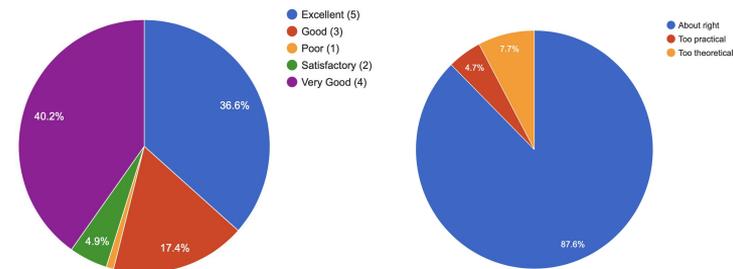
- Des formations régionales, nationales et internationales
- Un catalogue national des formations
- Un cycle de formation autour des principes FAIR en bioinformatique et pour la gestion des données
- Des actions de promotion
  - des bonnes pratiques de formation
  - de la mutualisation de matériel pédagogique (supports, workflows, jeux de données)
- Des actions ciblées pour les formateurs
  - ex : LMS moodle, GT e-formation

	2019	2020	2021
<b>Training events</b>	138	127	176
International	7	16	
Europe/Elixir	3	3	
National	53	76	
Regional	60	19	
University training	35	18	
Professional training	91	207	
Average satisfaction rate	86%	79%	
<b>Learners</b>	2169	2326	4458
<b>Trainers</b>	NA	272	379

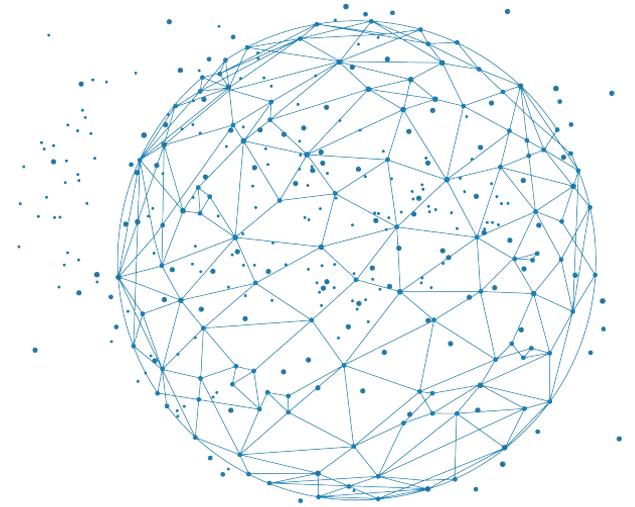
**Science  
Ouvverte et  
principes FAIR**



- ELIXIR-FR Training Coordinator (O. Sand) & deputy (H. Chiapello)
- Adoption de BioSchemas (scraping par TeSS)
- Meilleures annotations des formations dans TeSS
- Curation de TeSS
- FAIR Training Focus Group  
(<https://elixir-fair-training.github.io/FAIR-training-handbook>)
- Formation en collaboration avec d'autres noeuds  
(<https://fpsom.github.io/2021-06-ml-elixir-fr/>)
- e-Learning
- projet CONVERGE WP2 (Training and Capacity Building)
- Training Metrics Database  
(<https://training-metrics-dev.elixir-europe.org>)



# Présentation de l' équipe pédagogique ETBII



# Lucie Khamvongsa-Charbonnier



- ORCID : <https://orcid.org/0000-0002-1194-0546>

- Page Web :

- Institut :

- Institut Français de Bioinformatique

- Équipe / GT:

- Training (All-star IFB)

- GT IFB Formation

- GT IFB e-Formation

- Ville :

- Marseille

- Compétences

- Analyse single omic (ChIP, ATAC, RNA-seq)

- Analyse multi-omic

- Reproductibilité

- Formation

- Interventions dans ETBII

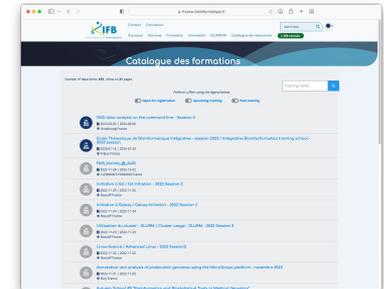
- Organisation

- Techno (Cloud,Cluster)

- Principes FAIR appliqués au matériel pédagogique



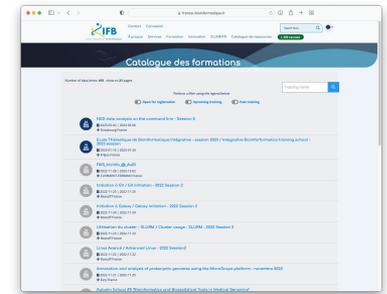
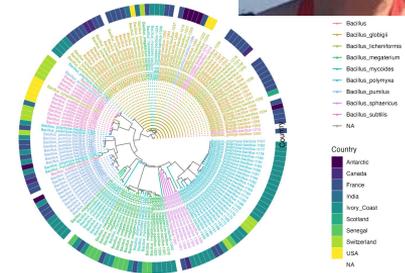
Science Ouverte et  
principes FAIR



# Hélène Chiapello



- **ORCID** : <https://orcid.org/0000-0001-5102-0632>
- **Page Web** : <https://cv.hal.science/helene-chiapello>
- **Institut** :
  - INRAE, unité MaIAGE
  - Institut Français de Bioinformatique
- **Équipes / GT** :
  - Equipe INRAE StatInfOmics
  - GT IFB Formation, e-formation et Science Ouverte
- **Ville** :
  - Jouy-en-Josas
- **Compétences**
  - Génomique et bioinformatique microbienne
  - Phylogénie et phylogénomique
  - Formation
  - Science ouverte, principes FAIR
- **Interventions dans ETBII**
  - Organisation
  - Principes FAIR appliqués au matériel pédagogique



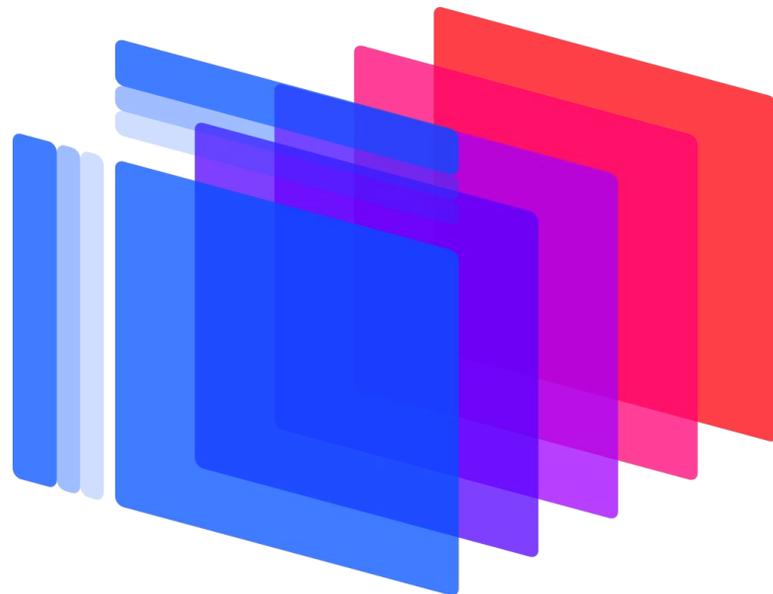
Science Ouverte et principes FAIR



# Anaïs Baudot

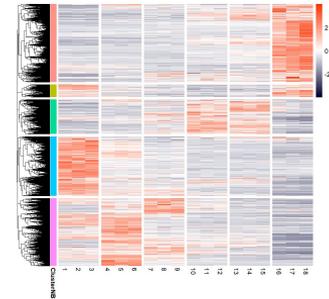
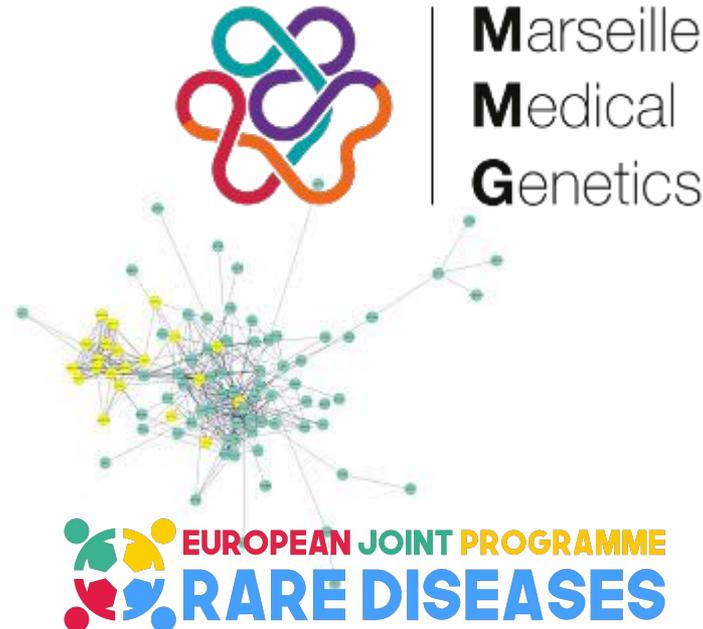


- **ORCID** : <https://orcid.org/0000-0003-0885-7933>
- **Page Web** :  
<https://www.marseille-medical-genetics.org/a-baudot/>
- **Institut** :
  - Marseille Medical Genetics
  - Barcelona Supercomputing Center
- **Équipe** :
  - Networks and Systems Biology for Diseases
- **Ville** : Marseille
- **Compétences**
  - Biologie des réseaux
  - Biologie des systèmes
  - Intégration de données
- **Interventions dans ETBII**
  - Organisation
  - Principes FAIR appliqués au matériel pédagogique





- **ORCID** : <https://orcid.org/0000-0002-4090-2573>
- **Page Web** : <https://www.marseille-medical-genetics.org/a-baudot/>
- **Institut** :
  - Marseille Medical Genetics
- **Équipe** :
  - Networks and System Biology for Diseases
- **Ville** :
  - Marseille
- **Compétences**
  - Analyse OMICs
  - Intégration de données
- **Interventions dans ETBII**
  - Atelier réseaux



# Samuel Chaffron

- **ORCID** : 0000-0001-5903-617X
- **Page Web** : <https://pagespersowp.ls2n.fr/samuelchaffron/>
- **Institut** :
  - ❑ Laboratoire des Sciences du Numérique de Nantes (LS2N), CNRS UMR 6004 et Nantes Université.
  - ❑ Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE.

## ■ **Équipe** :

- ❑ ComBi (Combinatoire et Bioinformatique)

## ■ **Ville** :

- ❑ Nantes

## ■ **Compétences**

- ❑ (Méta)génomique et Microbiomes
- ❑ Intégration de données (multi-omiques)
- ❑ Biologie des (éco)systèmes

## ■ **Interventions dans ETBII**

- ❑ Atelier graphes / réseaux
- ❑ Journal club
- ❑ Travail matériel pédagogique





■ ORCID : <https://orcid.org/0000-0003-1465-1640>

■ Institut :

- ❑ CNRS
- ❑ Institut Français de Bioinformatique

■ Équipe :

- ❑ IFB-core, Bilille
- ❑ GT IFB Formation, e-formation, Prospective & Innovation, Interopérabilité
- ❑ ELIXIR Training Platform, FAIR training FG, CONVERGE WP2, TeSS Club

■ Ville :

- ❑ Villeneuve d'Ascq (Lille)

■ Compétences

- ❑ Analyses NGS (variants, ChIP-seq)
- ❑ Analyses régions régulatrices
- ❑ Formation

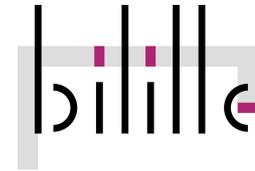
■ Interventions dans ETBII

- ❑ Organisation
- ❑ Principes FAIR appliqués au matériel pédagogique



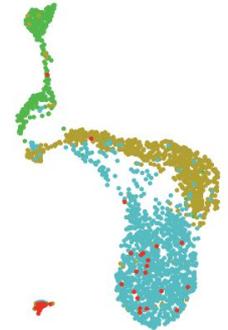
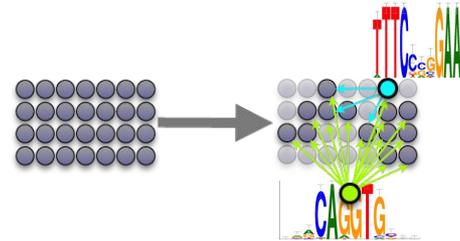


- **ORCID** : 0000-0003-0189-0220
- **Page web** : <https://wikis.univ-lille.fr/bilille/>
- **Institut** :
  - UAR 2014 - US 41 Plateformes Lilloises en Biologie & Santé
- **Équipe** :
  - Plateforme de bioinformatique bilille
- **Ville** :
  - Lille
- **Compétences**
  - single-omic (transcriptomique, protéomique)
  - intégration de données omiques
- **Interventions dans ETBII**
  - Principes généraux intégration de données omics
  - Journal club



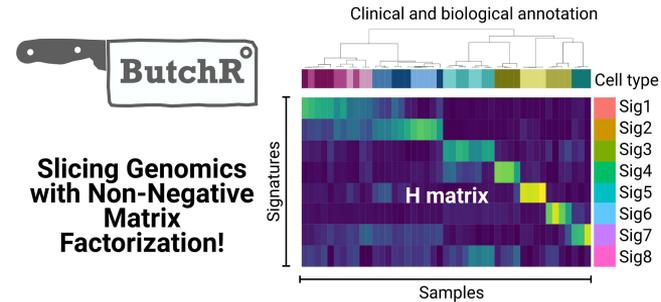


- **ORCID** : <https://orcid.org/0000-0003-1684-9888>
- **Page Web** :
- **Institut** :
  - ❑ CIML (jusqu'au 31/03/23)
  - ❑ CRCM (à partir du 01/04/23)
- **Équipe** :
  - ❑ Instabilité génomique et hémopathie humaines
  - ❑ Facteurs épigénétiques dans l'hématopoïèse normale et pathologique
- **Ville** :
  - ❑ Marseille
- **Compétences**
  - ❑ Omics
  - ❑ Inférence de GRN
- **Interventions dans ETBII**
  - ❑ Atelier analyses multivariées
  - ❑ Journal club
  - ❑ Animation groupes de travail sur le matériel pédagogique





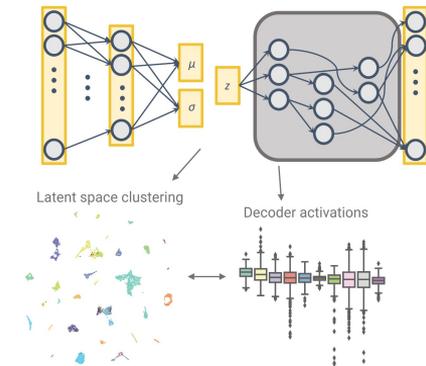
- ORCID : <https://orcid.org/0000-0003-4989-4722>
- Page Web : <https://www.hdsu.org/>
- Institut :
  - Medical Faculty Heidelberg
  - Health data science unit & BioQuant
- Équipe :
  - Biomedical Genomics Group
- Ville :
  - Heidelberg
- Compétences
  - Omics integration
  - Cis-regulation
  - ML for omics data
- Interventions dans ETBII
  - Atelier et cours d'analyses multivariées
  - Journal club
  - Animation groupes de travail sur le matériel pédagogique



healthdata  
scienceunit

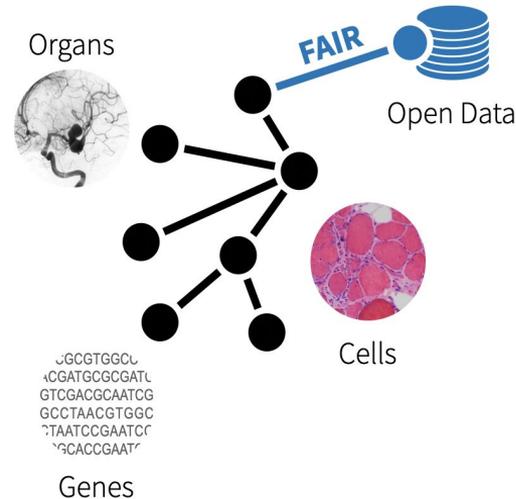
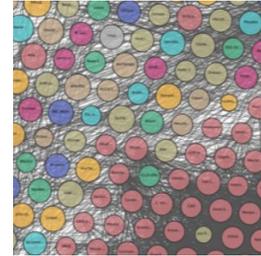


HEIDELBERG  
FACULTY OF  
MEDICINE



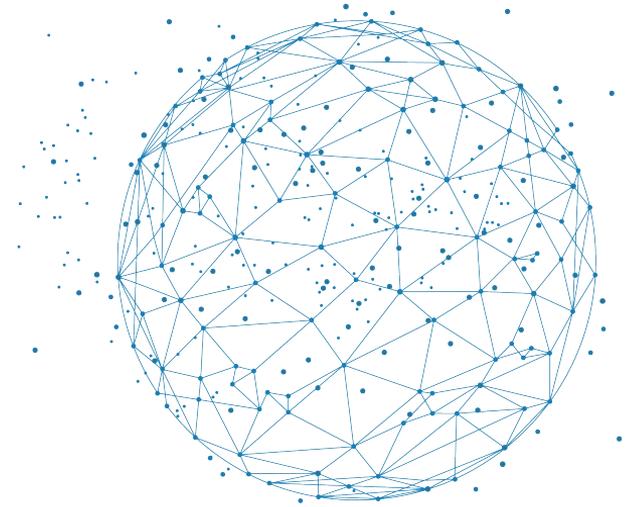
# Alban Gaignard

- ORCID : <https://orcid.org/0000-0002-3597-8557>
- Page Web : <http://albangaignard.github.io>
- Institut :
  - Institut du Thorax
  - LS2N (collaborateur)
- Équipe :
  - équipe génétique
  - Plateforme BiRD
  - IFB - GT interop - GT santé
  - Elixir - Bioschemas
- Ville :
  - Nantes
- Compétences
  - Semantic web / Knowledge graphs
  - Workflows / Provenance
- Interventions dans ETBII
  - Cours / atelier semantic web
  - Journal club / matériel pédagogique





# Présentation des participants ETBII 2023





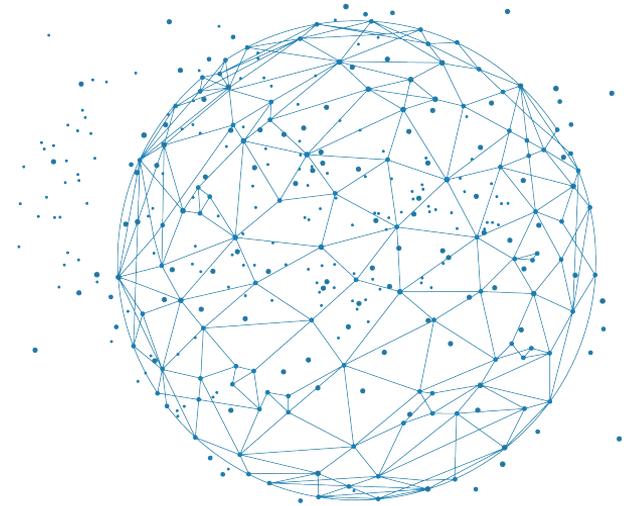
Vous ? Votre équipe / unité ?

Votre première expérience de la bioinformatique intégrative ?

Vos attentes pour cette formation ?

Vos besoins en termes de matériel pédagogique partagé ?

# Planning et informations pratiques



# Planning synthétique de la semaine



lundi 16/01	mardi 17/01	mercredi 18/01	jeudi 19/01	vendredi 20/01
<i>Sur la route</i>	9h-10h30 Théorie analyses multivariées	9h-10h30 Ateliers réseaux/web sém./analyses multivariées	9h-10h principes FAIR appliqués au matériel pédagogique	9h-10h30 Travail en parallèle sur le matériel pédagogique
	<i>10h30-11h pause</i>	<i>10h30-11h pause</i>	<i>10h-10h30 pause</i>	<i>10h-10h30 pause</i>
	11h-12h30 (1h30) Théorie réseaux	11h-12h30 Ateliers réseaux/web sém./analyses multivariées	10h30-12h30 Travail parallèle sur le matériel pédagogique	10h30-12h Restitution des sous groupes en plénière & évaluation finale
	<i>12h30-14h déjeuner</i>	<i>12h30-14h déjeuner</i>	<i>12h30-14h déjeuner</i>	<i>panier repas</i>
14h-15h Introduction	14h-15h30 (1h30) Théorie web sémantique	14h-15h30 Ateliers réseaux/web sém.analyses multivariées	14h-16h Travail en parallèle sur le matériel pédagogique	<i>Sur la route</i>
15h-15h30 Concepts généraux	<i>15h30-16h pause</i>	<i>15h30-16h pause</i>	16h-16h30 Bilan d'avancement en plénière	
<i>15h30-16h00 pause</i>	16h-17h00 Présentation des ateliers	16h-17h30 Ateliers réseaux/web sém./analyses multivariées	16h30-17h pause	
16h00-18h00 Journal club 1/2	17h-18h Ateliers réseaux/web sém./analyses multivariées	17h30-18h30 Ateliers réseaux/web sém./analyses multivariées	17h-18h30 Journal club 2/2 restitution	
18h00-18h30 Présentation technique & tests	18h-1830 - La bioinformatique intégrative et PICD			
	Introduction et Fondamentaux	Ateliers en parallèle	Séances sur matériel pédagogique	

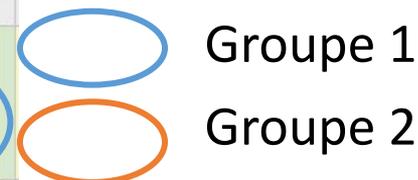
- ETBII 2023 → 1<sup>er</sup> édition
- Tous à 2 ateliers
  - analyses multivariées MV1/MV2
  - réseaux R1/R2
  - web sémantique WS1/WS2
- Section moodle “A faire avant la formation”
  - ↳ Présentation technique & tests 18h-18h30

Web sémantique		Réseaux		Analyses multivariée	
WS1	WS2	R1	R2	MV1	MV2
Cariou	Denecker	Bihouée	Simon	Messak	Bihouée
Delpuech	Messak	BIDAUT	Aloui	ORJUELA	Cariou
Laghrissi	LAMER	Denecker	ORJUELA	Chaparro	Delpuech
COGNAT	Zanzoni	Montagne	Chaparro	Figeac	Laghrissi
Aloui	Brito	Guille	Figeac	Hennequet -Antier	COGNAT
Simon	Poirot	Moine-Fra nel	Hennequet -Antier	Oger	BIDAUT
		LAMER	Oger	Zanzoni	Montagne
		HELIGON	Mariadass ou	Mariadass ou	Guille
		Poirot	Viard	Viard	Moine-Fra nel
		Da Costa	Berthelier	Brito	HELIGON
		Jamilloux		Berthelier	Da Costa
					Jamilloux



Web sémantique		Réseaux		Analyses multivariées	
WS1	WS2	R1	R2	MV1	MV2
Cariou	Denecker	Bihouée	Simon	Messak	Bihouée
Delpuech	Messak	BIDAUT	Aloui	ORJUELA	Cariou
Laghrissi	LAMER	Denecker	ORJUELA	Chaparro	Delpuech
COGNAT	Zanzoni	Montagne	Chaparro	Figeac	Laghrissi
Aloui	Brito	Guille	Figeac	Hennequet -Antier	COGNAT
Simon	Poirot	Moine-Fra nel	Hennequet -Antier	Oger	BIDAUT
		LAMER	Oger	Zanzoni	Montagne
		HELIGON	Mariadass ou	Mariadass ou	Guille
		Poirot	Viart	Viart	Moine-Fra nel
		Da Costa	Berthelie	Brito	HELIGON
		Jamilloux		Berthelie	Da Costa
					Jamilloux

mardi 17/01	mercredi 18/01
9h-10h30 (1h30) Théorie analyses multivariées	9h-10h30 (1h30) Ateliers réseaux/web sémantique/analyses multivariées WS1 / R1 / MV1
10h30-11h Pause café	10h30-11h Pause café
11h-12h30 (1h30) Théorie réseaux	11h-12h30 (1h30) Ateliers réseaux/web sémantique/analyses multivariées WS1 / R1 / MV1
12h30-14h (1h30) Déjeuner	12h30-14h (1h30) Déjeuner
14h-15h30 (1h30) Théorie web sémantique	14h-15h30 (1h30) Ateliers réseaux/web sémantique/analyses multivariées WS2 / R2 / MV2
15h30-16h Pause café	15h30-16h Pause café
16h-17h (1h) Présentation des ateliers	16h-17h30 (1h30) Ateliers réseaux/web sémantique/analyses multivariées WS2 / R2 / MV2
17h00-18h00 (1h) Ateliers réseaux/web sémantique/analyses multivariées WS1 / R1 / MV1	17h30-18h30 (1h) Ateliers réseaux/web sémantique/analyses multivariées WS2 / R2 / MV2



A ajouter  
nom des salles pour  
les TPs  
et répartition atelier  
salle



Les petits-déjeuners de 8h à 10h

Les déjeuners à partir de 12h30

Les dîners de 19h30 à 20h30

Jeudi soir dîner de gala

Vendredi:

1. Il faut libérer les chambres pour 10h
2. un local sera disponible pour stocker les bagages
3. les plateaux repas seront disponibles à partir de 12h dans la salle de restauration



INSTITUT FRANÇAIS DE BIOINFORMATIQUE

