

# Séminaire interne du CRCM

## La plateforme Disc

Bernard Chetrit

Datacentre IT and Scientific  
Computing

Centre régional de lutte contre le cancer Provence-Alpes-Côte d'Azur



INSTITUT PAOLI-CALMETTES



Instituts  
thématiques



**Inserm**

Institut national  
de la santé et de la recherche médicale



# Disc un outils pour la recherche

- Un lieu et un outil d'échange autour du calcul et de l'exploitation des données scientifiques.
- Une infrastructure d'exploitation et de tests dans des conditions de calcul fortement parallélisés.
- Une offre de serveurs virtuels permettant de mettre rapidement en place un environnement de calcul ou de développement spécifique.
- De l'espace pour le stockage et la sauvegarde des données scientifiques.
- Des outils de visualisation de qualité permettant l'affichage déporté sur des postes simples même dépourvus de cartes graphiques performantes.

# Ressources

- **Alambic cluster (236 CPU-CORES, 4 TFLOPS)**
- **Agui cluster (260 CPU-CORES, 4,2 TFLOPS)**
- **Serveurs de virtualisation (86 CPU-CORES 2 TFLOPS)**
- **2 noeuds de visualisation 3D (20 CPU-CORES, 2xQuadro K5200, 256G RAM)**
- **Systeme de stockage des données Scientifiques 2x145 To**
- **Reseau d'interconnexion 10GBs point à point.**

# Services sur les clusters



AGUI  
260 CPU-CORES  
1608 Gb RAM  
QDR infiniband  
4,2 Tflops

Alambic  
236 CPU-CORES  
1168 Gb RAM  
QDR infiniband  
4 Tflops

# Logiciels et applications sur les clusters

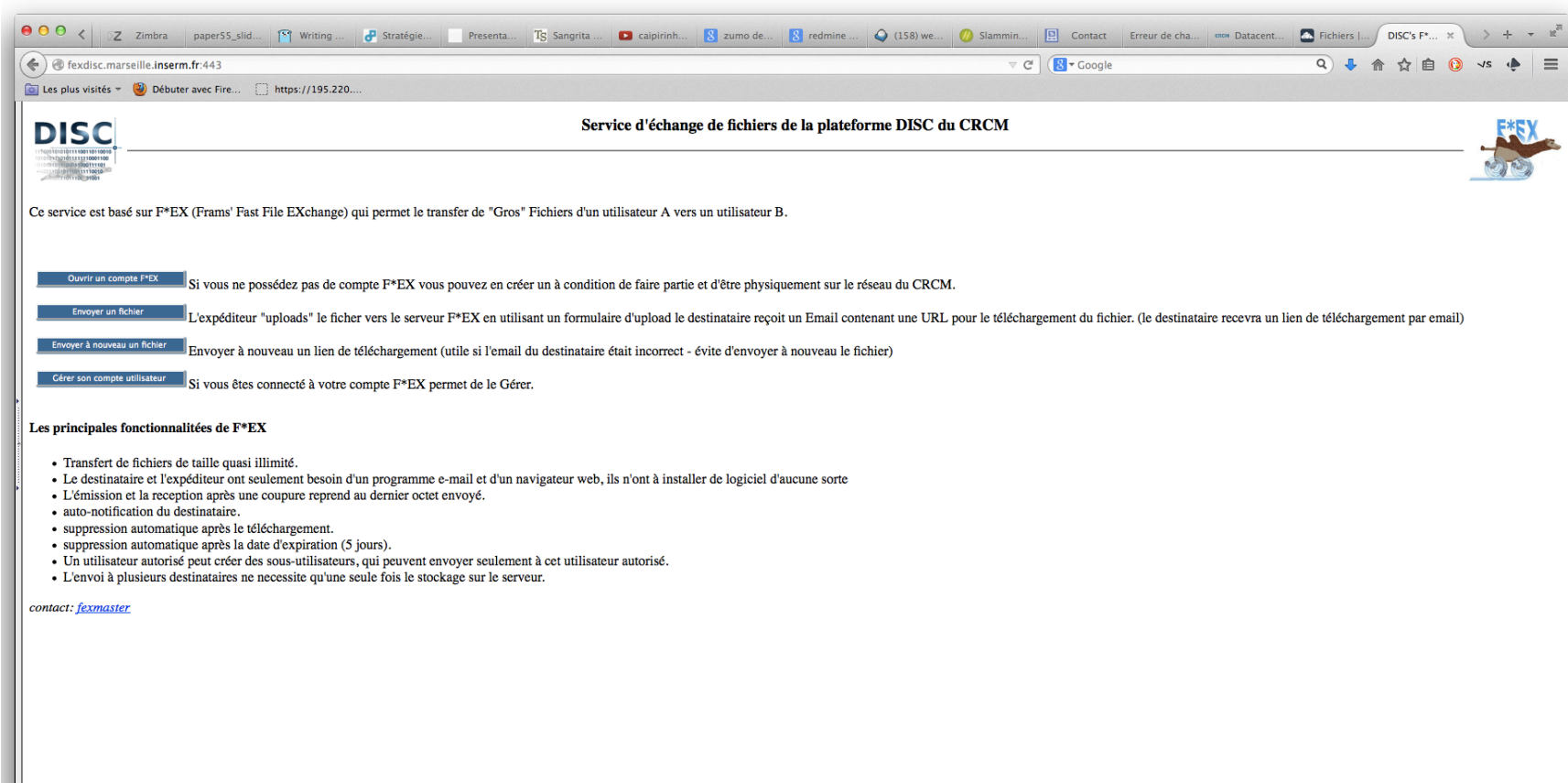
- Analyse
- Docking
- dynamic
- Imagerie
- ngs
- NMR
- SAXS
- Statistique
- Cristallographie

# Services

- Machine virtuelles à façon:
  - Serveurs de licences: Matlab, Sybyl, Topspin
  - Serveurs Web: Mobyli, 2p2idb, Gfscore, Moltarpred,USR,trget, iscb, Cibi,Disc,dart
  - Web services: Fexdisc, ForgeCRCM, Rstudio Serveur
  - Passerelle de connexion SSH: charon
  - Serveurs de calcul:  
Dynam (R,Namd,Charm,Matlab)  
Hotnet (R/Bioconductor, Matlab, Hotnet, Cytoscape,..)

# Fexdisc Transfert de fichiers

<http://fexdisc.marseille.inserm.fr>



The screenshot shows a web browser window with the URL [fexdisc.marseille.inserm.fr:443](http://fexdisc.marseille.inserm.fr). The page title is "Service d'échange de fichiers de la plateforme DISC du CRCM". The DISC logo is in the top left, and the F\*EX logo is in the top right. The main text explains that the service is based on F\*EX (Frams' Fast File EXchange) for transferring files between users. Below this, there are four blue buttons with corresponding instructions:

- Ouvrir un compte F\*EX**: Si vous ne possédez pas de compte F\*EX vous pouvez en créer un à condition de faire partie et d'être physiquement sur le réseau du CRCM.
- Envoyer un fichier**: L'expéditeur "uploads" le fichier vers le serveur F\*EX en utilisant un formulaire d'upload le destinataire reçoit un Email contenant une URL pour le téléchargement du fichier. (le destinataire recevra un lien de téléchargement par email)
- Envoyer à nouveau un fichier**: Envoyer à nouveau un lien de téléchargement (utile si l'email du destinataire était incorrect - évite d'envoyer à nouveau le fichier)
- Cérer son compte utilisateur**: Si vous êtes connecté à votre compte F\*EX permet de le Gérer.

Below the buttons, the section "Les principales fonctionnalités de F\*EX" lists the following features:

- Transfert de fichiers de taille quasi illimité.
- Le destinataire et l'expéditeur ont seulement besoin d'un programme e-mail et d'un navigateur web, ils n'ont à installer de logiciel d'aucune sorte.
- L'émission et la réception après une coupure reprend au dernier octet envoyé.
- auto-notification du destinataire.
- suppression automatique après le téléchargement.
- suppression automatique après la date d'expiration (5 jours).
- Un utilisateur autorisé peut créer des sous-utilisateurs, qui peuvent envoyer seulement à cet utilisateur autorisé.
- L'envoi à plusieurs destinataires ne nécessite qu'une seule fois le stockage sur le serveur.

At the bottom, there is a "contact: [fexmaster](#)" link.

# ForgeCRCM site de travail collaboratif

<http://forgecrcm.marseille.inserm.fr>

The screenshot shows a web browser window displaying the ForgeCRCM website. The address bar shows the URL: [forgecrcm.marseille.inserm.fr/projects/demandes-de-service-sur-la-plateforme-disc/wiki/Demandes\\_de\\_service](http://forgecrcm.marseille.inserm.fr/projects/demandes-de-service-sur-la-plateforme-disc/wiki/Demandes_de_service). The page title is "Demandes de services et de support à la plateforme DISC". The page content includes instructions on how to submit a service request to the DISC platform, a list of documents for download, and a list of links for further information. The page is displayed in a dark theme with a sidebar on the left containing navigation links like "Aperçu", "Activité", "Demandes", "Nouvelle demande", "Gantt", "Calendrier", "Annonces", "Documents", "Wiki", "Fichiers", "Dépôt", "Statistiques", and "Configuration".

ACCUEIL MA PAGE PROJETS AIDE CONNECTÉ EN TANT QUE CHETRIT MON COMPTE DÉCONNEXION

CRCM » DISC » Demandes de service sur la plateforme DISC Recherche: » Demandes de service sur la plateforme DISC

**Demandes de services et de support à la plateforme DISC** Modifier Surveiller Verrouiller Renommer Supprimer Historique

**Aperçu**

**Activité**

**Demandes**

**Nouvelle demande**

**Gantt**

**Calendrier**

**Annonces**

**Documents**

**Wiki**

**Fichiers**

**Dépôt**

**Statistiques**

**Configuration**

Pour déposer une demande de service à la plateforme DISC veuillez remplir le document [Demande-service-Plateforme-DISC](#) et le renvoyer en utilisant l'onglet [Nouvelle demande](#) de ce projet.

Les demandes seront étudiées et une réponse vous sera faite ASAP

Accessoirement, vous pouvez télécharger selon vos besoins les documents suivant:

- [Quelques commandes de bases pour Linux](#)
- [Service d'échange de fichiers de la plateforme DISC](#)
- [Comment générer une clé SSH pour se connecter aux serveurs de la plateforme DISC](#)
- [Lancer des calculs sur les clusters de la plateforme DISC, utilisation de OAR](#)

Nouveau fichier

Formats disponibles : PDF | HTML | TXT

**Wiki**

[Page de démarrage](#)

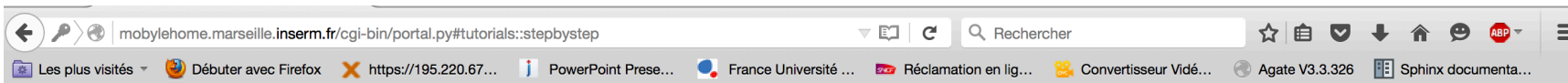
[Index par titre](#)

[Index par date](#)



# Le portail mobyle

## http://mobylehome.marseille.inserm.fr



## Mobyle

(*guest*)  
set email | sign-in | sign-out  
refresh workspace

[\[more\]](#)

**Programs**

- alignment
  - ITI
    - generateHTMLReport
    - optStep-integrateSubnets
    - step1-detectSubnetworks
    - step2-subnetworksStatisticalValidation
    - step3-intersectSubnetworks-getGO-computeP
    - step4-generateHTMLReport
- NGS
  - reporting
- Plot
  - plotScoreHist
- Screening
  - HTS-Net
- transcriptome

**Data Bookmarks** [\[overview\]](#)

**Jobs** [\[overview\]](#)

**Tutorials**

- How to use Mobyle? A step by step tutorial
- Registration information
- Sequence formats
- Alignment formats

- Welcome
- Forms
- Data Bookmarks
- Jobs
- Tutorials**

### Step-by-step x

## Step by step tutorial

Welcome to Mobyle, a portal to run bioinformatics analyses. This tutorial is interactive: it helps you to open forms, navigate through the portal and fill some fields. At any time, you can come back here by clicking on the Tutorials tab above.

### Running an analysis

Let's start by running a first analysis: predict a protein membrane topology with the **toppred** program. Clicking [here](#) opens the **toppred** form. The first field is for the protein sequence. You can paste the following data:

```
>sp|P02914|MALK_ECOLI_MALTOSE/MALTODEXTRIN_TRANSPORT_ATP-BINDING_PROTEIN_MALK.
MASVQLQNVTKAWGEVYVSKDINLDIHEGEFVVFVGPSCGCKSTLLRMIAGLETITSGDL
FIGEKRMNDTPPAERGVGMVFSYALYPHLSVAENMSFGLKPAGAKKEVINQRVNVQVAEV
LQLAHLDRKPKALSGGQRQVVAIGRTLVAEPSVFLLEPLSNLDAALRVQMRTEISRLL
KRLGRTMIYVTHDQVEAMTLADKI VVLDAGRVAQVGKPLELYHYPADRFVAGF IGSPKMN
FLPVKVTATAIDQVQVELPMPNRQQVWLPVESRDVQVGANMSLGI RPEHLLPSDIADVIL
EGEVQVVEQLGNETQIHQIPSI RQNLVYRQNDVVLVEEGATFAIGLPPERCHLFREDGT
ACRRLHKEPGV
```

Then launch the program by clicking on the Run button.

When finished, the results are available under the **Jobs** tab. The output starts by some details on the job, followed by the output files. You can return to the form anytime by clicking on its tab, at the top of the page.

### Using the databox

You can input data either by copy/paste, file loading, database fetch, or selection of the output from a previous analysis. Let us try this with some examples.

#### - file loading

If you don't have any sequence file, you can save the following sequences into a file on your disk.

# Rstudio serveur

<http://disc.marseille.inserm.fr>

The screenshot displays the RStudio web interface. The top navigation bar includes the R logo, menu items (File, Edit, Code, View, Plots, Session, Build, Debug, Tools, Help), and user information (chetrir). The browser address bar shows the URL [disc.marseille.inserm.fr/rstudio/](http://disc.marseille.inserm.fr/rstudio/). The main workspace is divided into several panels:

- Environment Panel:** Shows the current environment with 177 observations and 2 variables. The 'Values' section lists variables: EN (List of 6), ENV ("rsthda"), FR (List of 6), and LANG ("FR").
- Files Panel:** Displays a list of installed and available packages.
- Console Panel:** Shows the R prompt and output, including the R license and a list of helpful commands like `contributors()`, `citation()`, `demo()`, `help()`, `help.start()`, and `q()`.

**Package List (from Files Panel):**

Name	Description	Version
gridExtra	Miscellaneous Functions for "Grid" Graphics	2.2.1
abind	Combine Multidimensional Arrays	1.4-3
acepack	ace() and avas() for selecting regression transformations	1.3-3.3
ade4	Analysis of Ecological Data : Exploratory and Euclidean Methods in Environmental Sciences	1.7-3
affy	Methods for Affymetrix Oligonucleotide Arrays	1.48.0
affyio	Tools for parsing Affymetrix data files	1.40.0
affyPLM	Methods for fitting probe-level models	1.46.0
AnnotationDbi	Annotation Database Interface	1.32.3
aplpack	Another Plot PACKage: stem.leaf, bagplot, faces, spin3R, plotsummary, plothulls, and some slider	1.3.0



# Les Projets qui utilisent la plateforme

**Support de calcul pour la plateforme d'Oncogénomique Moléculaire (Chaffanet, Birnbaum). Pipeline de sélection des variants dans le cancer du sein écrit par A. Guille.**

**ALGOPRED : Dr JM- BOHER (IPC) - Increase random forest survival adjustment on a Cox model.**

**Suivi des dynamiques conformationnelles de protéines flexibles (calcul d'ensembles de conformations, simulations des données de SAXS et RMN, etc...) Véronique-Receveur:**

- Interactions entre le domaine intrinsèquement désordonné de BCR et les domaines SH3 et SH2 d'ABL dans la leucémie myéloïde Chronique
- Domaine N-terminal intrinsèquement désordonné de l'Androgen Receptor (AR), et ses liens avec le Cancer de la Prostate (Projet ANR n° ANR-13-BSV5-0013, Coll. Marc-André Delsuc, IGBMC, Strasbourg)
- Mode de régulation fonctionnelle de la Protéine CP12 intrinsèquement désordonnée, régulant la GAPDH (Coll. Brigitte Gontero-Meunier, BIP, Marseille)
- Maturation d'une molybdoenzyme lors de l'insertion du cofacteur (coll. Axel Magalon, LCB, Marseille)
- Conformations actives de l'Acétyl et de la Butyryl cholinesterases (coll. Pascale Marchot, AFMB, Marseille)

# Les Projets qui utilisent la plateforme

## Projets Cibi:

- Analyses courantes en cytométrie en flux, microarrays et protéomique.
- Projet PL-BIO 2012-2015 coordonné par Dr *Christophe Ginestier* (CRCM). **Criblage haut débit d'une banque d'ARN interférence pour l'identification de voie de signalisation régulant les cellules souches cancéreuses mammaires. (20kE investis en infrastructure)**
- - Projet Plan Cancer 2013-2016 coordonnée par Dr *Sophie Vasseur* (CRCM): **Identification of metabolic pathways contributing to pancreatic adenocarcinoma progression and metastatic dissemination. (50kE investis en infrastructure)**

## Projet ISCB:

- Projet PDZ (Synténine) - Financement AMIDE
- Projet PDZ (Grasp55) - Financement AMIDEX
- Projet PDZ (Erbin) - Financement AMIDE
- Projet dCK - Financement ARC
- Projet Bromodomaine (BRD4-BD1) - NA

# Les Projets qui utilisent la plateforme

## Projets cancéropole:

- Rôle de la protéine hsp27 dans le cancer de la prostate-**Palma Rocchi**
- Understanding EZH2 recruitment to chromatin in hematopoietic cells-**Estelle Duprez**
- Localisation du facteur de transcription Spp1 chez la levure-**Vincent Géli**
- Genome-wide localisation of RPA (Replication Protein A) and its mutants in *Saccharomyces cerevisiae*-**Vincent Géli**
- Role de cdc13 et ses interaction lors de la réplication-**Vincent Géli**
- Impact des introns sur la formation des R-loop et transcription-**Vincent Géli**
- Site de fixation de proteines de recombinaison meiotique-**Bertrand Llorente**
- Identification of *eme1<sup>65A</sup> rqh1 $\Delta$*  suppressor mutations from the hybridization pattern of genomic DNA on oligonucleotide tiling arrays by means of the SNPscanner algorithm- **Pierre-Marie DEHE**
- Identification des sites préférentiels de SLX4 à la chromatine par CHIP seq. Une fois identifiés, des modulations de la cartographie de ces sites seront analysées au cours du cycle cellulaire et en réponse à des agents génotoxiques.-**PH Gaillard/Sarah Scaglione**
- Décrypter le rôle de XIST dans l'homéostasie de l'épithélium mammaire et le développement du cancer du sein - **Christophe Ginestier**
- JAM-C expression identifies Leukemic Initiating Cells in Acute Myeloid Leukemia Stem Cell compartment-**Maria De Grandis**
- Role de l'hypoxie dans la régulation de l'expression des inhibiteurs de l'apoptose-**Aurélie Tchoghandian (UMR911-CRO2, équipe 4, Timone)**
- .....

# Plus d'informations

<http://disc.marseille.inserm.fr>

Prochain mardi des technologies gourmandes!

***Arnaud Guille***

Utilisation des outils de la plateforme DISC dans un pipeline NGS

22 Mars 2016 à 10H30