

# DBCLS Galaxyではじめる ゲノムスケールデータの 共同研究

山口 敦子  
中尾 光輝



大学共同利用機関法人 情報・システム研究機構

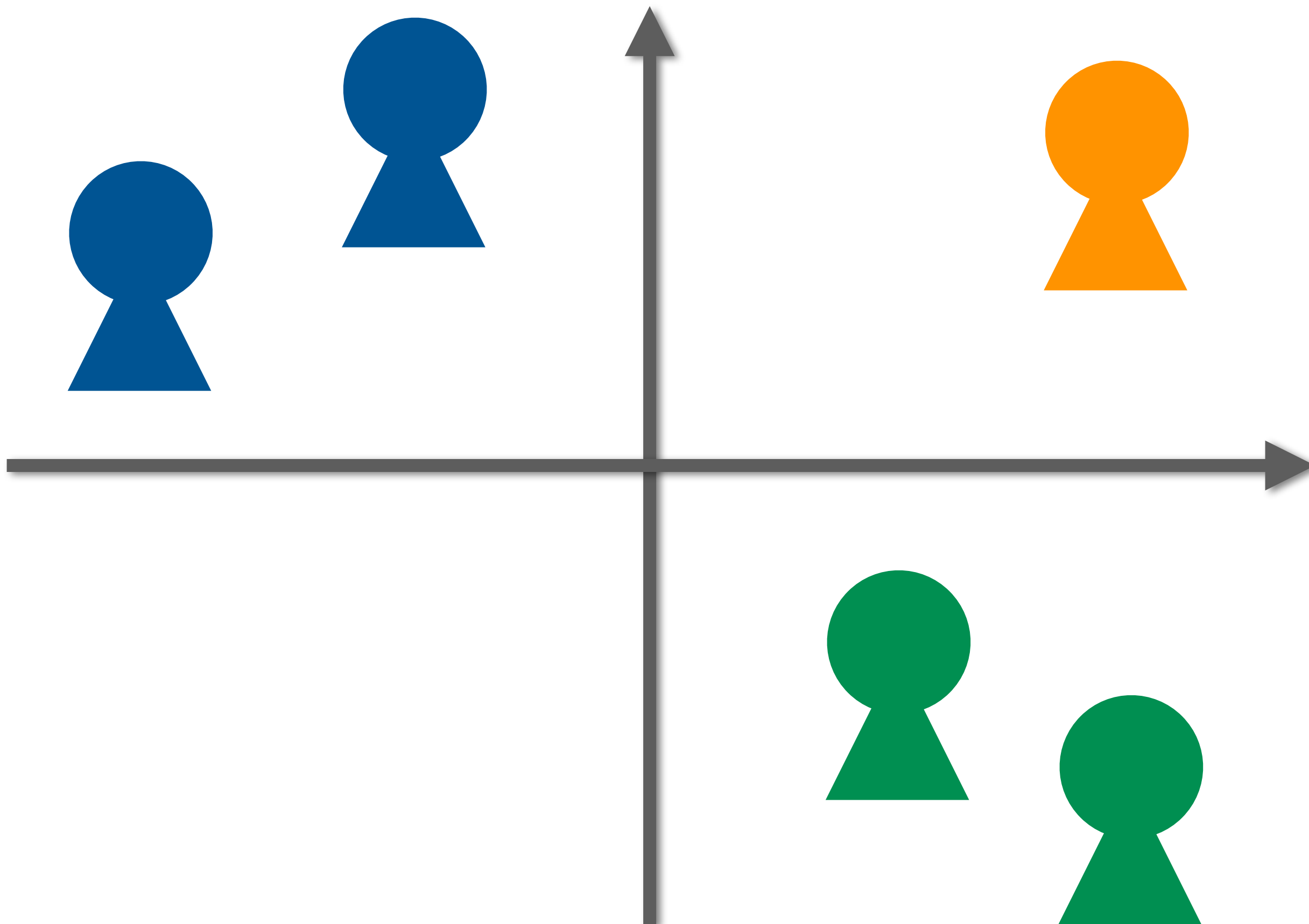
ライフサイエンス統合データベースセンター

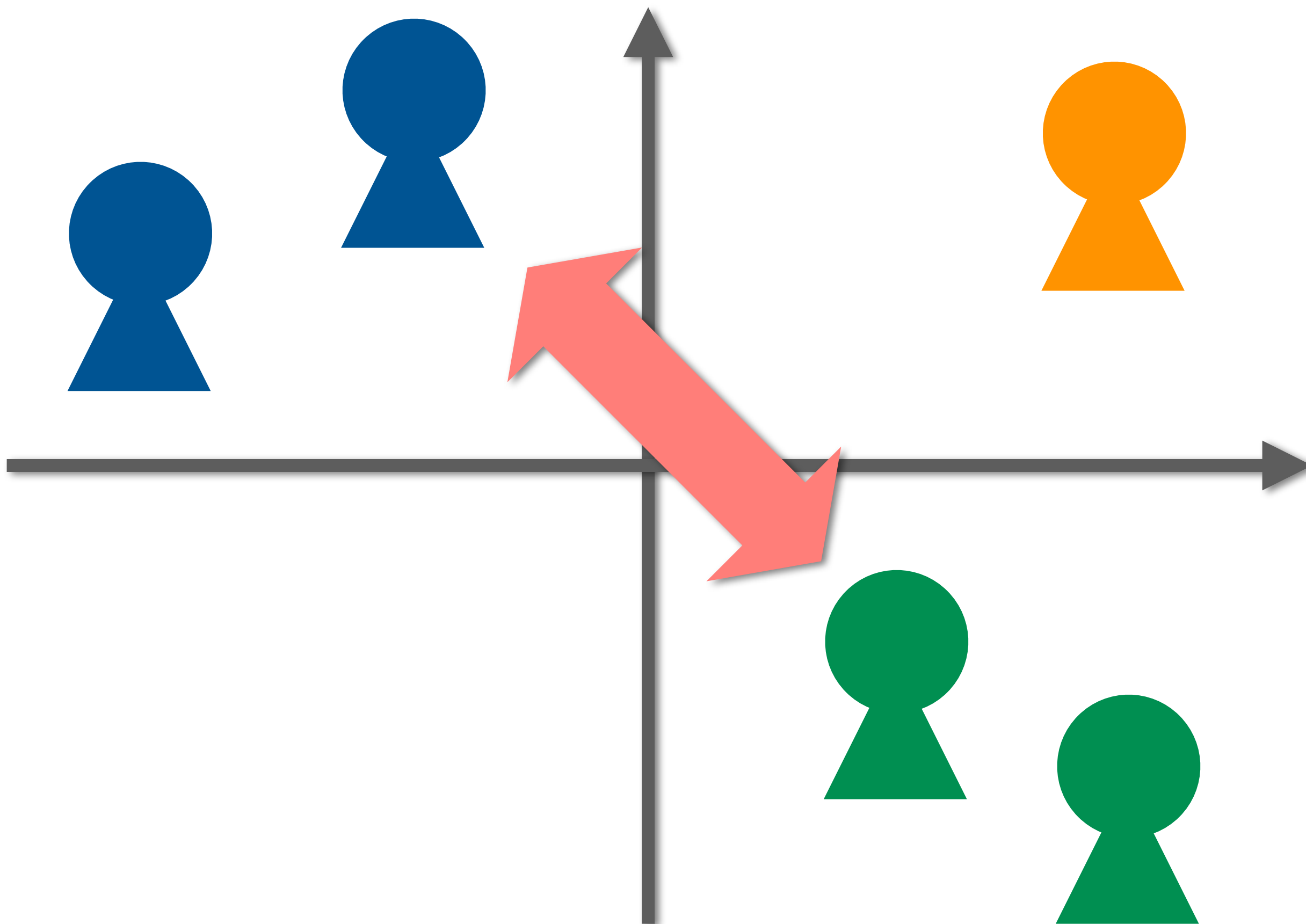
# DBCLS Galaxyではじめる ゲノムスケールデータの 共同研究

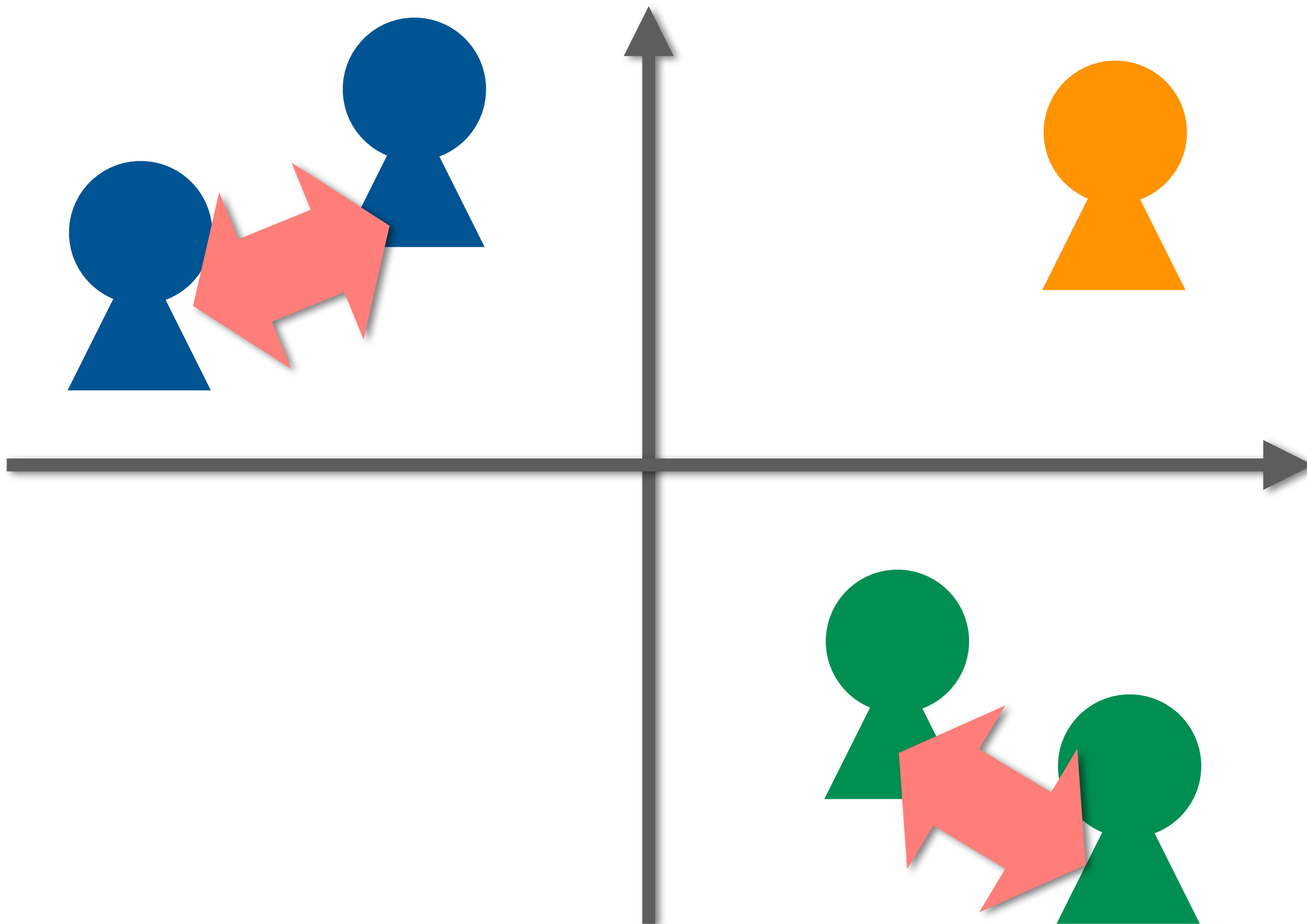
# 共同研究で大切な事

# コミュニケーション

# データ共有









# 300MBのデータファイルの受け渡し

# 最新のデータファイル はどれか？

このデータセット、  
どのように解析され  
たのか？

1000万個のSNPsは  
Excelでは扱えない

# ゲノムスケール共同研究の課題

データの共有

解析履歴の共有

再現性の確保

サイバーインフラストラクチャの問題

# DBCLS Galaxyではじめる ゲノムスケールデータの 共同研究

# Galaxy とは

- ウェブアプリケーション
- ゲノムデータ解析環境：座標比較、多重整列、ゲノムアノテーション比較、メタゲノムサンプルプロファイリングなどなど
- データ交換／公開／共有
- <http://galaxyproject.org>

# Galaxy でできること



# CHiP-seq実験で同定された TAF1結合サイトを含むプロ モーターを見つける

<http://screencast.g2.bx.psu.edu/galaxy/flash/TAF1.html>

**CHiP-seq法**とは、クロマチン免疫沈降法とシーケンシング法を組み合わせ、転写調整因子のゲノム上の結合部位を網羅的に同定する方法。

**TAFI 結合サイト**とは、DNA結合タンパク質かつ転写制御因子、TAFIタンパク質（基本転写因子TFIIDサブユニット1）が結合するゲノム領域。

**プロモーター**とは、遺伝子の転写制御にかかわるシス因子ゲノム領域。

Galaxy | Published Page | Screencasts

◀ ▶ 🔍 🗨 ⌚ ⌚ +

http://main.g2.bx.psu.edu/screencast

リーダー ↻ 🔍 Google

Galaxy

Analyze Data Workflow Data Libraries Help User

Published Pages | aun1 | Screencasts

About this Page

Galaxy Screencasts

Galaxy Screencast

A picture is worth a thousand words. In this screencast, we guide you through the steps of a typical Galaxy workflow.

If you're new: Six steps to get started

1. Finding promoters for a set of genes from a ChIP-seq experiment: we will find genes that contain such features as promoters, exons, etc. You need a file containing genomic coordinates (e.g., BED) downloaded from here.

2. Finding Exons with the help of genomic features: exons are defined by the coordinates of exons and introns. You will follow along as we find exons in a set of genes.

3. Saving your results: we will save the results safely stored and that you can share with colleagues. This is the only required step.

4. Generating a workflow: the result is a history of analysis workflows. In this step, we will generate a workflow from scratch.

5. Generating workflow: a workflow from scratch: dataset in a 6 column format.

6. Extracting Sequences: we will extract genomic sequences and FASTA format as well as obtain corresponding alignment blocks from multiple species whole genome alignments in their native format and also create a pseudo-global alignment in FASTA format, which is suitable for use in existing analysis software.

Metagenomic Analyses with Galaxy

1. Processing and analysis of 454 data in metagenomic studies. Here we show the utility of Galaxy in metagenomic

デモ

19

# データセット、解析履歴 (ヒストリー) を共有する

<http://screencast.g2.bx.psu.edu/galaxy/flash/Sharing.html>

# デモ

## 21

# さらにくわしくは

スクリーンキャスト <http://galaxycast.org>

統合TV <http://togotv.dbcls.jp/20090310.html>

# Galaxy.DBCLS.jp

各種ツールをプリインストール

DBCLSツール入ってる

すぐに使える状態 (OpenID があれば)

日本語メニュー表示



# お持ち帰り Galaxy

<ftp://ftp.dbcls.jp/galaxy/>

Galaxy.DBCLS.jpのVMWareブートイメージ

各種ツールインストール済み

ラボ内用に立ち上げるのに便利



# ゲノムスケール共同研究のシナリオ

データと解析の共有はGalaxyで解決

あたらしいデータやツールのプラットフォーム（サイバーインフラストラクチャ）としてGalaxyが有用

論文作成サポートはつぎのトークで