

# DSTEP: 生命データサイエンス教育における計算機環境



鈴木 絢子, 鈴木 穰 (東京大学 大学院新領域創成科学研究科 メディカル情報生命専攻)

## プログラムのねらい



新領域創成科学研究科メディカル情報生命専攻および先端生命科学専攻では、DSTEP (Data Scientist Education/Training Program) と称して大規模生命データを扱う教育プログラムを実施している。ビッグデータ時代への転換に対応できていない基礎医学分野の逼迫する現状に鑑み、人材の育成の観点から現状の打開を試みるものである。製薬企業、バイオテクノロジー企業、バイオインフォマティクス企業、合計12社との間にコンソーシアムを結成し、オンザジョブトレーニング(OJT)による社会人のリカレント教育を行う点に特色がある。OJTとして実際の問題に直結したデータを産生し、その解析を志向した演習科目を設けているが、この遂行には高度な計算機環境の整備が喫緊の課題である。今回、柏地区に整備される一連の計算機環境について、その生命データサイエンス教育への活用について意見交換を行いたい。

## カリキュラム

### コアカリキュラム概要

#### 実践演習 (On the Job Training; OJT)

- 企業の抱える問題を模式的ではあるが実践的に解決
- 必要な実験データの取得・解析手法の開発と実践
- 博士研究として実施

#### 演習科目群

- 企業・アカデミア研究機関へのインターンシップ
- 計算機プログラム初等演習
- バイオインフォマティクス最新プログラム群に関する演習

#### 講義科目群

- メディカルゲノム基礎・発展講義
- バイオインフォマティクス基礎講義

## 特色ある演習科目

### 「バイオデータプログラミング演習I」

必要に迫られてプログラミングを書籍で独学すると基礎知識の不足に悩むことが多い。そのような学生が過去に躓いたポイントを調べ、必要な基礎知識を大量に補った独自のオンラインプログラミング学習教材を作成した。生物学・医科学のプログラミングに必要な概念をゼロから広く浅く学び、研究に必要な高度なプログラミングを将来的に独学できるような「基礎体力」を養う。Linuxコマンドライン, Python/Rプログラミングなど。

### 「バイオデータプログラミング演習II」

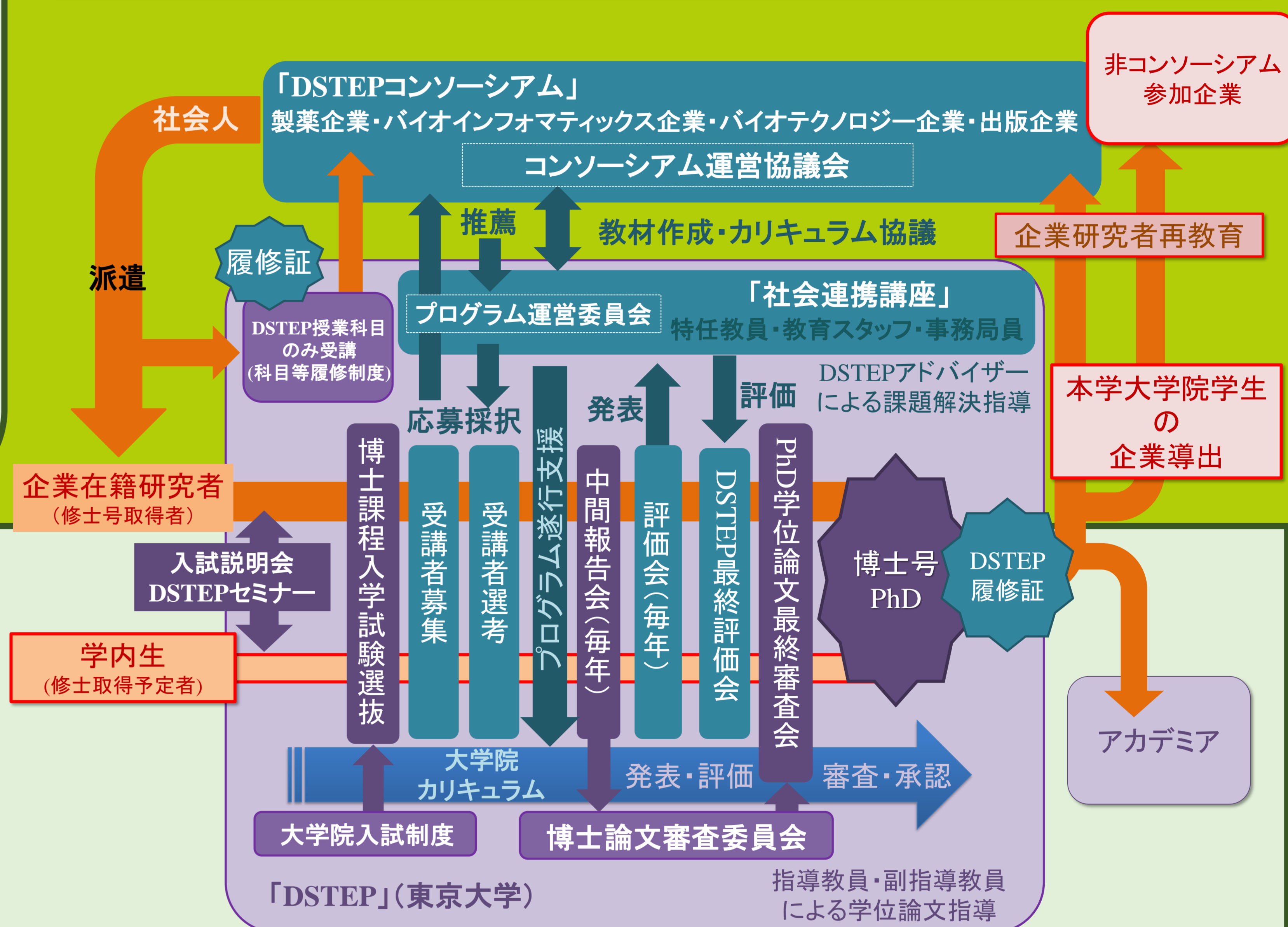
先駆的に開発され試用が開始されている情報解析ツール群を用い、その利用法について、実データを用いた実践的演習形式で習熟する。オミクス解析, シングルセル解析, ロングリード解析など。



DSTEP では、東京大学医科学研究所ヒトゲノム解析センターの協力のもと、演習科目で同センターのスーパーコンピュータ SHIROKANE を利用しています。

東京大学 大学院新領域創成科学研究科  
DSTEP: Data Scientist Education/Training Program  
<http://dstep.cbms.k.u-tokyo.ac.jp/>

## 協賛企業 (DSTEPコンソーシアム)



## プログラム履修モデル

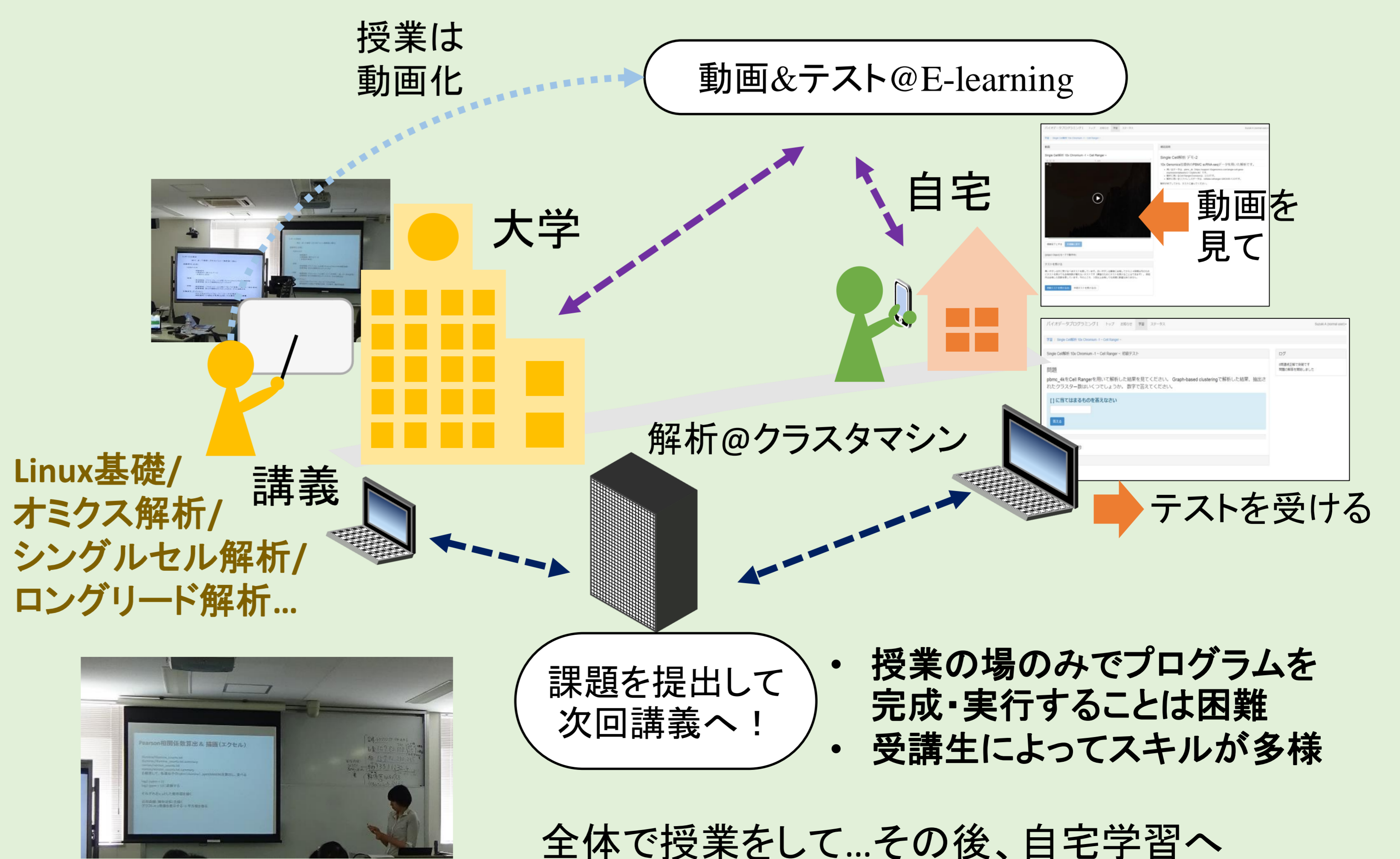
**プログラム受講生 (学内):** 博士課程の在来学生および社会人学生

**A群:** 実践演習を履修し、プログラム運営委員会の承認した研究課題 (OJT 課題) で博士号取得を目指す。

**B群:** プログラム科目を履修するが、該当課題以外で博士号取得を目指す。

A・B群ともに、修了時、博士取得者に対して修了証を発行する。

**科目等履修生 (学外):** 科目単発での受講。成績証明書を発行。



全体で授業をして...その後、自宅学習へ