

平成 27 年度

バイオインフォマティクス人材育成カリキュラム

NGS ハンズオン講習会

実施報告書

門 田 幸 二

(東京大学 大学院農学生命科学研究科 特任准教授)

## 概要

近年の次世代シーケンサ(NGS)解析技術の飛躍的な発展は、ライフサイエンス分野におけるデータ量の急増を生み、データの整備や活用がボトルネックとなっている。研究現場では、NGSデータを自在に解析できるバイオインフォマティクス人材が求められている。これらの状況を踏まえ、バイオサイエンスデータベースセンター（NBDC）運営委員会 人材育成分科会では、NGS を利用した様々な研究目的に対応可能な人材を効率的に育成するためのカリキュラム（<http://biosciencedbc.jp/gadget/chousa/generation-sequencer.pdf>）の策定が行われた。このNGS用カリキュラムは、最低限必要とされる知識・技術を2週間程度で身につけることを想定した「速習」と、時間をかけて習得することを想定した「速習以外」に分かれており、平成26年度に「速習」コースが試行的に実施された。

本共同研究では、平成26年度の速習コース実施結果を踏まえ、NBDCと連携してハンズオン講義に特化した講習会の実施および関連教材の整備を行い、受講者アンケート等によりその有効性を検証する。これらの活動を通じて、NGSデータの解析において必要な人材の育成に資する教材の整備・充実や公開および知識の共有を目指す。

## 1. 実施体制およびスケジュール

本研究では、NGS解析に特化したバイオインフォマティクス人材育成を目的として、1) NGSハンズオン講習会の実施、および2) 教材の整備について研究開発を行う。

グループ名	研究代表者氏名	所属機関・部署・役職名	研究題目
東京大学	門田幸二	東京大学・大学院農学生命科学研究科・特任准教授	次世代シーケンサ解析に特化したバイオインフォマティクスハンズオン講習会の実施および教材の整備

全体スケジュールは以下の通りである。

研究項目	H27年度
1) NGSハンズオン講習会 ・講師の選定、講義内容の検討など ・講習会実施(7/22~8/6, 8/26~28)	
2) 教材の整備 ・ポータルサイトの構築支援 ・「(Rで)塩基配列解析」の内容充実	

## 2. 実施内容及び結果

### 2-1. NGS ハンズオン講習会の概要

本講習会は、平成 26 年度の速習コース講習会実施結果を踏まえ、以下の日時及び体制で実施した（表 1）。

実施日	実施時間	項目	レベル	形式	講師(所属)
A日程					
7月22日	10:30-18:15	PC環境の構築（Bio-Linux 8とRのインストール状況確認）	-	自習	門田幸二(東京大学) 寺田 透(東京大学)
7月23日	10:30-18:15	Linux基礎	初・中級	実習	門田幸二(東京大学)
7月24日	10:30-18:15	スクリプト言語(シェルスクリプト)	中級	実習	服部恵美(Amelieff)
7月27日	10:30-18:15	スクリプト言語(Perl)	中級	実習	服部恵美(Amelieff)
7月28日	10:30-18:15	スクリプト言語(Python)	中級	実習	服部恵美(Amelieff)
7月29日	10:30-18:15	データ解析環境R：基礎	初・中級	実習	門田幸二(東京大学)
7月30日	10:30-18:15	データ解析環境R：Bioconductorの利用法	中級	実習	門田幸二(東京大学)
8月3日	10:30-18:15	NGS解析（基礎）	初級	実習	山口昌雄(Amelieff)
8月4日	10:30-18:15	NGS解析（ゲノムReseq、変異解析）	初級	実習	山口昌雄(Amelieff)
8月5日	10:30-14:45	NGS解析（RNA-seq前半）	初級	実習	山口昌雄(Amelieff)
	15:00-18:15	NGS解析（RNA-seq後半）	初級	実習	門田幸二(東京大学)
8月6日	10:30-18:15	NGS解析（ChIP-seq）	初級	実習	森岡勝樹(理研)
B日程					
8月26日	10:30-18:15	NGS解析（基礎）	初級	実習	山口昌雄(Amelieff)
8月27日	10:30-18:15	NGS解析（ゲノムReseq、変異解析）	初級	実習	山口昌雄(Amelieff)
8月28日	10:30-14:45	NGS解析（RNA-seq前半）	初級	実習	山口昌雄(Amelieff)
	15:00-18:15	NGS解析（RNA-seq後半）	初級	実習	門田幸二(東京大学)

表 1. 講習会概要

平成 26 年度講習会との主な違いは、下記の通りである：

- ① 講師数を減らして項目間の連携を強化（10 人 → 4 人）
- ② Python 追加（7/28）
- ③ NGS 解析項目を増加（2.5 日 → 4 日分）
- ④ 統計解析を追加（8/5 後半）
- ⑤ ハンズオンのみ（講義のみはなし。全て実習を含む講義）
- ⑥ 予備日の確保（8/26-28 に受講希望者数の多い項目を B 日程として実施）

## 2-2. 講習会実施体制

本講習会は、東京大学、および JST-NBDC の 2 つの機関による主催として実施した。また、他に 2 つの機関が共催として参画した。具体的な各機関の担当業務は表 2 の通りである。

主催機関1	東京大学・大学院農学生命科学研究科・アグリバイオインフォマティクス教育研究ユニット
担当業務	NGSハンズオン講習会の統括、講師の選定、具体的な講義内容の検討、実習用PCへのソフトウェアインストール作業、講義実施および補助（Teaching Assistant; TA）、教材に関する情報収集、ポータルサイトの構築支援
参考URL	<a href="http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngs_sokushu_2015">http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngs_sokushu_2015</a>
主催機関2	科学技術振興機構・バイオサイエンスデータベースセンター（JST-NBDC）
担当業務	NGSハンズオン講習会の統括、企画・推進、講義実施および補助（ビデオ撮影）、受講者アンケートの実施、ポータルサイトの構築、受講者への追跡調査
参考URL	<a href="http://biosciencedbc.jp/human/human-resources/workshop/h27">http://biosciencedbc.jp/human/human-resources/workshop/h27</a>
共催機関1	HPCI戦略プログラム分野1「予測する生命科学・医療および創薬基盤」人材養成プログラム（産総研・ゲノム情報研究センター）
担当業務	受講申込受付業務全般（受付ウェブシステムの開発・運用）
参考URL	<a href="https://hpci.cbrc.jp/modules/tutorial/NBDC2015.html">https://hpci.cbrc.jp/modules/tutorial/NBDC2015.html</a>
共催機関2	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）
担当業務	講義補助（TA）

表 2. 講習会実施体制

実施場所は、平成 26 年度講習会と同じ東京大学農学部 2 号館 2 階第一講義室（最大 140 名収容）とした。平成 26 年度の講習会後に改修が行われ、座席数が約 20 席増えた。また、前方のスクリーンが 2 つに増え、講義スライド画面とコマンドを打つ別の PC 画面の同時表示が可能となった。



図 1. 講習会会場（7/22 の自習風景）。講義室後方から撮影

### 2-3. 講習会受講対象、受講定員、受講条件など

受講対象は、学生・社会人、部分受講・全日程受講を問わず可能とし、共催機関1のウェブサイトで受講を受け付けた。平成27年度は、昨年度に設けたような優先順位（修士卒以上や参加可能な日程の多い方など）を設けなかった。ただし、受講対象の欄に下記情報を掲載し、受講希望者には事前準備および予習を課した（予習内容については4-5を参照）：

- NGSの概論的な講義は実施しません。
- リード、ペアエンドなどの基本的な用語やNGS原理の説明などは省略する予定です。
- 昨年度（平成26年度）の講義資料や参考図書など参照の上、受講希望項目の基礎知識を取得しておくことをお勧めします。

また、以下の5項目を受講必須条件として挙げた：

- キーボード操作等（テキスト入力、マウスによるウィンドウ操作やコピー・ペースト等）PCの基本操作をスムーズに行うことが可能な方
- 本講習会は講習会場後方よりビデオ撮影を行い、後日公開予定です。このビデオに映る可能性があることを了承していただける方
- 受講前および受講期間中に、受講に必要なソフトウェアのインストールやデータのダウンロードをお願いすることがありますが、これら事前準備を行ったうえで受講可能な方
- 受講者アンケートにご協力いただける方
- 受講者追跡調査にご協力いただける方

さらに、以下のお願い事項も示した：

- スケジュールおよび内容は、やむを得ない事情等により変更される場合があります。
- 万が一受講できなくなった場合は速やかに事務局に連絡をお願いします。当日講義開始15分後までに事前連絡なく会場にいらっしゃらなかった場合には、他の方に席を譲ることがあります。
- 周囲に困っている人がいた場合、わかる範囲でアシストをお願いいたします。

受講申込受付サイトは4月15日に公開し、ウェブサイトやメーリングリストなどで告知を行った。募集人数は、平成26年度と同様に30名程度（応募者多数の場合、最大80名程度まで）とした。受講申し込み期間は、4月20日～5月15日とし、先着順ではない旨を明記した。申込み推移は、5日ごとに26名（4/20）、56名（4/25）、60名（4/30）、70名（5/5）、92名（5/10）、122名（5/15）であった。

表3は、項目ごとの受講希望者数の内訳と実際の受講者数を示したものである。5/15の締切時点で最も受講希望が多かったのは、8/3-5開催分の3項目であった（113名、110名、113名）。これらの受講希望者の中から、予備日でもよいので受講希望と表明していた一部をB日程（8/26-28）に振分けることで、受講希望者の最大数を8/6のChIP-seqの104名まで減らした。7/22以降の講習会当日までに一部項目や全日程のキャンセルがあり、事前連絡のあったものを集計すると「受講当日までの希望者数」欄のようになった。今回は、部分受講を認めていることから、受講当日に受

講者が自分で出席簿に○をつける形式を採用した。「実際の受講者数」欄は、その出席確認表の数を集計したものである。多少漏れている可能性は否定できないが、ほぼ実数通りと思われる。

実施日	項目	5/15締切時点		受講当日までの希望者数	実際の受講者数
		振分前	振分後		
A日程					
7月22日	PC環境の構築 (Bio-Linux 8とRのインストール状況確認)	85	85	81	59
7月23日	Linux基礎	93	93	86	67
7月24日	スクリプト言語(シェルスクリプト)	86	86	80	69
7月27日	スクリプト言語(Perl)	94	94	84	67
7月28日	スクリプト言語(Python)	90	90	83	64
7月29日	データ解析環境R：基礎	96	96	80	68
7月30日	データ解析環境R：Bioconductorの利用法	98	98	85	68
8月3日	NGS解析 (基礎)	113	87	82	71
8月4日	NGS解析 (ゲノムReseq、変異解析)	110	84	77	59
8月5日	NGS解析 (RNA-seq)	113	85	78	65
8月6日	NGS解析 (ChIP-seq)	104	104	95	71
B日程					
8月26日	NGS解析 (基礎)		26	24	32*
8月27日	NGS解析 (ゲノムReseq、変異解析)		26	25	29*
8月28日	NGS解析 (RNA-seq)		28	28	33*

表 3. 受講希望者数の内訳および実際の受講者数

5/15の締切時点での受講希望者数は122名であったが、このうち26名が結果的に全日程キャンセル（一度も講習会に出席しなかった）した。26名中13名が連絡あり、残りの13名が連絡なしの不参加（平成26年度はゼロ）であった。受講希望総数が平成26年度（93名）に比べて増えていること、部分受講可能にしたことも多少影響していると思われる。今後の講習会においては「連絡なし不参加者」はゼロであることを願いたい。

本講習会では、A日程（7/22～8/6）の講習会期間における実際の受講者数およびB日程の受講希望者数を鑑み、A日程終了後の8/7～8/18にB日程追加受講申し込みを受け付けることとした。この期間の追加受講希望者数は、14名であった。表3において、B日程の「実際の受講者数」（32名、29名、33名）が「受講当日までの希望者数」（24名、25名、28名）よりも多いのは、「実際の受講者数」欄のみ、この追加受講希望者分を含めているためである。

追加受講分を含めた総受講希望者数は136名（=122+14）であった。136名の各種内訳は表4の通りである。関東在住の受講生が8割弱を占めた。それ以外は、青森、山形、群馬、石川、山梨、三重、京都、鳥取、岡山、広島、徳島、福岡、沖縄、海外、未選択が各1名となっていた。研究分野（複数選択可）は、生物学・医歯薬学・農学が多数を占めた。職種は、学生が61名、ポスドク・

研究員が40名、助教・講師が11名となっていた。これは20代・30代がほとんどを占める年齢層とも合致する。尚、受講希望者136名うち、1項目以上の実際の受講者数は110名であった。

住所	東京：68名、神奈川：13名、埼玉：10名、茨城：6名、千葉：6名、愛知：5名、静岡：3名、宮城：2、大阪：2名、兵庫：2名
研究分野（複数回答可）	生物学（総合生物学を含む）：82名、医歯薬学：42名、農学：42名、情報学：12名、理工学：5名、その他：11名
職種	学生（大学院生）：52名、ポスドク：24名、研究員：16名、講師・助教：11名、学生（学部生）：9名、教授・准教授：7名、研究支援：5名、主任研究員：4名、開発：3名、その他：4名
学歴	博士課程修了：57名、博士課程在学中37名、修士課程修了：13名、修士課程在学中：16名、学部：在学中9名・卒2名
年齢	20代：62名、30代：45名、40代：21名、50代：7名、60代：1名
所属	大学：99名（うち、東大57名）、研究所：17名、企業：17名、その他（病院関係）：3名
Linux利用経験	全く触ったことが無い：49名、使ったことはあるがよく覚えていない：31名、簡単なコマンドなら知っている：35名、時々or頻繁に使っている：21名
R利用経験	全く触ったことが無い：39名、使ったことはあるがよく覚えていない：26名、簡単なコマンドなら知っている：38名、時々or頻繁に使っている：33名
NGS利用経験	今のところ使用予定はない：4名、使ったことはないが使用予定がある：62名、いずれ使用すると思う：22名、時々or頻繁に使っている：46名
実習PC環境	Win持込：64名、Mac持込：37名、アグリバイオPC貸与希望：35名
H26年度講習会受講の有無	受講していない：131名、受講した：5名
本講習会を知った経緯（複数回答可）	友人・知人：46名、アグリバイオインフォマティクス（主催機関1）：31名、メーリングリスト：22名、NBDC（主催機関2）：19名、HPCI（共催機関1）：14、キーワード検索の結果：4名、SNS（Facebook、Twitter等）：3名、アメリエフ（講師）からの案内メール：2名、DBCLS（共催機関2）：2名

表 4. 受講希望者 136 名の各種内訳

#### 2-4. ボランティア講義補助 (Teaching Assistant; TA) 制度について

本講習会では、平成 26 年度に引き続いてボランティアの TA も募集した。結果として、TA のみでの参加希望者は 2 名 (仮名 : A 氏と B 氏) であった。平成 26 年度とは異なり、講習会への参加形態を部分受講可能にしたため、受講生が得意分野の一部項目のみ TA として参加する枠組みも提供した。結果として、受講生として申込みをした 4 名が、表 5 に示す項目に TA 参加表明をした。TA が一定数必要と思われる項目については、主催・共催機関からの応援要員の派遣で対応した。

実施日	項目	TA申込		受講生として申込			
		A氏	B氏	1	2	3	4
A日程							
7月22日	PC環境の構築 (Bio-Linux 8とRのインストール状況確認)	○					
7月23日	Linux基礎	○					
7月24日	スクリプト言語(シェルスクリプト)	○	○	○			
7月27日	スクリプト言語(Perl)	○	○				
7月28日	スクリプト言語(Python)	○	○				
7月29日	データ解析環境R : 基礎	○	○			○	○
7月30日	データ解析環境R : Bioconductorの利用法	○	○			○	○
8月3日	NGS解析 (基礎)	○			○		
8月4日	NGS解析 (ゲノムReseq、変異解析)	○	○	○	○		
8月5日	NGS解析 (RNA-seq)	○		○	○		
8月6日	NGS解析 (ChIP-seq)	○	○		○		
B日程							
8月26日	NGS解析 (基礎)	○		○			
8月27日	NGS解析 (ゲノムReseq、変異解析)	○		○			
8月28日	NGS解析 (RNA-seq)	○		○			○

表 5. TA 申込み項目の内訳

平成 26 年度と比べるとボランティアの TA 申込み数は減った。平成 26 年度は、主に日頃指導する側の立場の人を対象として講義の仕方や指導内容を学んでもらうことを理念として掲げた。しかし、これらの講義映像や資料は、統合 TV や Youtube で公開済みである。そのため、上記の目的は公開済みの講義映像を見れば事足りる。また、昨年度は試行実施であること、物珍しさなども手伝ったのかもしれない。元来、ボランティア TA は、主催側にとって好都合な仕組みである。積極的に募集をやめるものでもないという判断のもと、募集の文言を多少変更して掲載したにすぎない。それゆえ、TA 参加人数の増減に一喜一憂する類のものではなく、TA 参加者にただ感謝するのみである。いずれにせよ、アンケート結果 (後述) でもほとんどの受講生が TA への感謝や TA の必要性を認識していることから、今年度の対処方針同様、TA 要員の必要数を穴埋めすることは重要であろう。具体的には、NBDC 取りまとめのもと、NBDC (主催機関 2) および DBCLS (共催機関

2) から各日 2 名の TA が派遣された。尚、昨年度に引き続き Twitter (<https://twitter.com/hashtag/AJACS>)も活用された。TA、講師、受講生らが#AJACS のハッシュタグつきでツイートすることで情報の共有を効率的に行うことができたようである。

## 2-5. 受講希望者に課した事前準備（予習）について

ハンズオン講義は、大きく2つの形態に分けることができる。1つは主催側が用意した同一OS、同一バージョンの統一的なPC環境で実習を行うものであり、もう1つは受講生が普段利用しているPCを持ち込んでヘテロなPC環境で行うものである。本講習会は後者のスタンスに立ち、受講生のレベルを一定以上に揃えるべく、平成26年度試行実施のときよりも厳しい事前準備を課した。Linux環境構築については、研究代表者が日本乳酸菌学会誌に連載中のNGS関連記事の連載第1-3回、および第4回のウェブ資料W6-5までを予習の必須とした。また、Rについても必要なパッケージのインストールから基本的な利用法までの理解を必須とした。これらの内容の多くは、平成26年度受講生の感想・要望・提案を反映させたものである。具体的には、下記URLから得られる情報の予習を要求した：

- 乳酸菌学会誌のNGS連載記事およびウェブ資料（Bio-Linux 8環境構築とその利用）  
[http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html#about\\_book\\_JSLAB](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#about_book_JSLAB)
- Rのインストール  
[http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html#about\\_install](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#about_install)
- Rの基本的な利用法  
[http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html#users\\_guide](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#users_guide)

受講生にはまた、平成26年度の報告書およびプレゼン資料、および平成27年度開催概要のプレゼン資料に目を通すよう促し、注意事項や心構えを説いた。

- 平成26年度試行実施報告書  
[http://biosciencedbc.jp/gadget/human/h26\\_ngs\\_report.pdf](http://biosciencedbc.jp/gadget/human/h26_ngs_report.pdf)
- 平成26年度試行実施のプレゼン資料  
[http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo\\_ngs\\_sokushu\\_2015/20150126\\_kadota.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2015/20150126_kadota.pdf)
- 平成27年度開催概要プレゼン資料（注意事項や心構え）  
[http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo\\_ngs\\_sokushu\\_2015/20150722\\_kadota.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2015/20150722_kadota.pdf)

講習会までに学んでおく具体的な到達目標として受講生に示したチェックリストは、下記の通りである。

- Bio-Linux 8 関連（Linux 環境構築）
  1. （第4回だけでなく第2-3回のウェブ資料PDFにも目を通している）
  2. ゲストOS、ホストOSなどと言われてうろたえない(第1回)
  3. Linux環境では、一般にダブルクリックでソフトウェアのインストールはしない(第1回)
  4. Perl, Ruby, Pythonはプログラミング言語である(第1回)
  5. 「円(yen)マーク」と「バックスラッシュ」の違いは見栄えの問題であり、文字コードとしては同じなので気にしない(第2回)
  6. コマンドとオプションの間には「スペース」が入るものだ(第2回)

7. . (どっと) から始まるファイル名のもの (の多く) は、環境設定ファイルである(第 2 回)
8. 全角文字をコマンドとして打ち込むのは非常識である(第 2 回)
9. ファイルの拡張子は常識的に使う。例えば FASTQ 形式ファイルの拡張子を .fasta や .fa などにしたない(第 2 回)
10. 「Tab キーによる補完」、「上下左右の矢印キー」を駆使して効率的にコマンドを打てるようになっている(第 2 回)
11. リダイレクトやパイプを含むコマンドを眺めて、何をやっているものかが大体わかるようになっている(第 2 回)
12. Bio-Linux の左側の引き出しみたいなアイコンまたはターミナル上で、ホスト <--> ゲスト間のファイルのやり取りが自在にできる(第 3 回)
13. ホストからゲストへはドラッグ&ドロップ可能だが、ゲストからホストへはドラッグ&ドロップ不可能[W8-1](第 3 回)
14. ターミナルの画面サイズや不具合は自分で変更および対処可能[W9-2-4](第 3 回)
15. 共有フォルダの概念を理解している[W9-3](第 3 回)
16. rm コマンドは「ゴミ箱」への移動ではなく本当に消える[W10-4](第 3 回)
17. ダウンロードしたいファイルの URL 情報取得ができる[W11-1](第 3 回)
18. ゲスト OS のネットワーク接続が不調の場合は、つながるまで待つのではなく、ホスト OS 上で必要なファイルをダウンロードして共有フォルダ 経由でゲスト OS 上の適切な場所に置くことができる[W11-1](第 3 回)
19. ダウンロードしたいファイルの URL 情報取得ができる[W11-1](第 3 回)
20. 圧縮・解凍コマンド(zip, unzip, gzip, gunzip, bzip2, and bunzip2)の使い方をマスターしている[W13](第 3 回)
21. head や tail コマンドでファイル中の最初と最後の部分を確認することができる[W14](第 3 回)
22. cat, more, less コマンドの基本的な使用法(脱出したいときは CTRL + C など)を理解している[W14-4 から W14-6](第 3 回)
23. 多くの Linux コマンドには様々なオプションがあり、--h や--help などを利用して可能なオプションを眺めることができる[W15-2](第 3 回)
24. PDF ファイルや PowerPoint ファイル中のコマンドをコピーするときは気をつける[W16-1](第 3 回)
25. コマンドオプション利用時の「ハイフン」と「ダッシュ」の気をつける[W16-3](第 3 回)
26. 初心者向けテキストエディタ gedit の基本的な利用法は知っている[W17](第 3 回)
27. 仮想環境全体のディスク使用状況把握(df コマンド)や、カレントディレクトリのディスク使用量(du コマンド)を把握し、適切に不要なファイルを削除できる[W18-2 から W18-4](第 3 回)

28. 「.(どっと)」はカレントディレクトリを表し、「..(どっとどっと)」は1つ上の階層のディレクトリを意味する[W18-3](第3回)
29. 複数のコマンドを組み合わせたパイプ(|)利用の基本を知っている[W19](第3回)
30. 連載原稿中では乳酸菌 RNA-seq データ([SRR616268](#))を例に解説しているが、NGS ハンズオン講習会(または予習)では、(自己責任で)自分が興味あるデータで平行して解析してもよいことを知っている[W20以降](第3回)
31. bzip2 は gzip に比べて圧縮率は高く、圧縮に要する時間も早い、解凍は遅い[W22-5 から W22-6](第3回)
32. ファイルサイズが大きいものは、一つ一つの処理に要する時間がとてつもなく長くなることを知っている[W23](第3回)
33. コマンドオプション、パイプ、リダイレクトを駆使することでNGSデータのサブセットを抽出できることを知っている[W25-2](第3回)
34. アスタリスク(\*)などのワイルドカードというものがあり、それらの挙動がある程度わかる[W1-2 から W1-3](第4回)
35. ゲスト OS がフリーズ状態に遭遇しても、自力で再起動するなりして復旧できる[W2-1](第4回)
36. Permission 問題は chmod で自力で解決できる[W3-1](第4回)
37. 同じ目的を達成する上で様々なやり方があることを知っている[W3-2](第4回)
38. ファイルのダウンロード時に拡張子がオリジナルと変わっていることがある[W4-4](第4回)
39. 自動マウント(オートマウント)の設定ができている[W4-5; W4-7](第4回)
40. 絶対パス、相対パスの概念が理解できている[W4-6](第4回)
41. 文字コード(改行コード)を od や file コマンドで確認し、perl で変換できる[W5-2; W5-3](第4回)

● R 関連

1. [インストール | について](#)をよく読み、ここに書いてある手順に従って 2015 年 4 月 4 日以降にインストールを行った
2. [インストール | について](#)で書いてある内容は Bio-Linux8(ゲスト OS)とは無関係であり、Windows や Macintosh(つまりホスト OS)上で行う作業である
3. ファイルの拡張子(.txt や.docx など)はちゃんと表示されている
4. R の起動と終了ができる。終了時に表示されるメッセージにうろたえない
5. R 本体だけでなく R パッケージ群のインストールもちゃんと行った
6. R パッケージ群のインストール確認も行い、エラーが出ないことを確認した
7. library 関数を用いた R パッケージのロード中に、別のパッケージがないことに起因するエラーメッセージが出ることもあるが、必要なパッケージを[個別](#)にインストールするやり方を知っている
8. [基本的な利用法](#)をよく読み、予習を行った

9. 作業ディレクトリの変更ができる
10. 例題ファイルのダウンロード時に、拡張子が勝手に変わることがあるので注意する
11. 慣れないうちは、`getwd()`と`list.files()`を打ち込むことで、作業ディレクトリと入力ファイルの存在確認を行う
12. エラーに遭遇した際、「ありがちなミス 1-4」に当てはまっていないかどうか自分で確認

## 2-6. 講習内容

講習会初日（7月22日）は自習、それ以外は全て実習（ハンズオン講義）である。カリキュラムの全体像、講義映像（統合TV）および講義資料は、NBDCのサイト（<http://biosciencedbc.jp/human/human-resources/workshop/h27>）より公開している。講義時間は10:30-18:15としたが、多くの項目で18:00前には終了した。

【日時】 2015年7月22日10:30-18:15

【題目】 Bio-Linux 8とRのインストール状況確認

【講師】 門田幸二（東京大学）、寺田透（東京大学）

【内容】 自習形式。講習会会場を解放し、事前準備として課した予習内容を各自で確認してもらった。また、持ち込みPCの無線LAN設定や名札などの各種手続きが行われた。受講生滞在時間を概ね2時間単位で区切り、あらかじめどの時間枠に来場するかを事前に問い合わせた。受講生の来場時間を分散化させることで、スタッフがゆとりをもって個別対応できた。

予習内容は、Bio-Linux8のisoファイルをスタート地点として、各自のPCスペックに応じた設定を行わせるものであった。7/22の来場段階でその設定ができていない、または環境設定に失敗している受講生に対しては、予め作成しておいた一通りの環境設定済みのovaファイルをUSBメモリで渡し、7/23以降の講習を受講可能なPC環境を構築してもらった。詳細な手順書([http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo\\_ngo\\_sokushu\\_2015/20150722\\_BioLinux8\\_ova\\_win.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2015/20150722_BioLinux8_ova_win.pdf))を配布し、資料を見ながら自力で行うのを基本とした。Rについては、全くインストール作業を行ってこなかった場合には、数GBにおよぶパッケージ群のダウンロードを講義室の無線LANで当日中に行うのは実質的に不可能である。この場合を想定し、アグリバイオインフォマティクスの事務局にR用のサーバを立てて、事務局にて寺田講師が有線LANで個別対応する体制を敷いた。結果として、持ち込みPCでそのような作業を行うことになった受講生は一人もいなかった。環境構築済みのノートPC貸与を選択したためかは定かではない。全体として、自習形式の受講生が殺到しない環境下で行うやり方は、実習本番中のトラブル極小化によるスタッフの負担軽減効果だけでなく、事前チェックができてよかった（付録1）など受講生からの評判もよかったように思われる。



【日時】 2015年7月23日10:30-18:15

【題目】 Linux基礎。LinuxコマンドなどUNIXの基礎の理解

【講師】 門田幸二（東京大学）

【内容】 7/24以降で用いる解析用データの取得。Bio-Linux8およびLinuxコマンドの復習。pwd, ls, cd, wget, clear, rmコマンドおよび各種オプション。ドラッグ&ドロップ、共有フォルダ経由でのゲストOS（Linux環境）へのデータコピー。mv, unzipや相対パスの復習。amelieffディレクトリと中身の確認。

8/3-5で用いるIntegrative Genomics Viewer (IGV)のインストール。7/23参加者（ドラッグ&ドロップ）と不参加者（wget経由）に分けたzipファイル取得の説明。unzip、パスの概念説明、「sudo ln -s ...」の実行。igv.shの実行と起動確認。

データ解析の全体像と、クオリティコントロール(QC)の位置づけを説明。QC用プログラムFastQC (ver. 0.10.1)の実行。ヘルプなどいくつかのオプション付きで実行。「fastqc -h | less」で、パイプの復習。「ls -d」、「rm -f」、「rm -rf」などのTips。シェルスクリプトの基本、chmodでの権限変更。「cp -r」。FastQC (ver. 0.10.1)実行結果の概観。最新版(ver.0.11.3)の存在認識、ダウンロード、解凍、およびインストール。パスの概念説明、正規表現、およびパスを通す作業。「unzip -q」でオプション利用の幅を広げる。一連のインストール作業をFastQC中のインストールマニュアルと対比させ、マニュアルを読み解くノウハウを伝授。2つのバージョンを実行し、バージョンによってデフォルトのオプションが異なりうることを学ぶ。

8/4に用いるGenome Analysis Toolkit (GATK ver. 1.6-13)の取得。/usr/local/src上で解凍すべく、sudoを使って一時的にrootになる意味を学ぶ。権限や所有者、chmod。到達目標は、①Linuxコマンドを自在に使えることの確認、②一般的なプログラムのインストールに必要なノウハウを身に着ける、③バージョンが異なると挙動が異なることを知る。



【日時】 2015年7月24日10:30-18:15

【題目】 スクリプト言語。シェルスクリプト

【講師】 服部恵美 (アメリエフ株式会社)

【内容】 シェルスクリプトのメリットは、①効率的に解析できること、②バイオインフォは日進月歩だがLinuxコマンドやシェルスクリプトの文法は大きく変わっていないので一度身につければ長く使える。シェルとは、ユーザが入力したコマンドをコンピュータに伝えるプログラムであり、**bash**、**zsh**などいくつかの種類がある。講習会では、Bio-Linuxのデフォルトである**zsh**をベースとした記述になっている。シェルスクリプトの作成と実行。**vi**や**gedit**などのテキストエディタで実行内容をファイルに書いて保存したのち、**sh**コマンドで実行。主に使ったのは、**gedit**のほう。テストデータは、**FASTA**形式ファイル。到達目標は、「FASTAファイルから指定した遺伝子の配列だけを取り出す」。

最初は「**head -n 4** ファイル名」のような簡単な記述内容のファイル(**gedit test1.sh**)からスタートし、**sh test1.sh**で実行。最初の8行分を表示するには?などの問いから徐々に難易度を上げていった。同じファイル名のものを何度も書くのは間違いのもとであり、面倒なので、変数を使って値を一元管理するやり方。「変数名=値」と書いたのち、**\$変数名**として変数に入っている値を呼び出すことができる。スクリプト内に何度も登場するゲノム配列のファイル名などは変数にしておいたほうが便利。対象ファイルが変わるたびにスクリプトファイル内のファイル名の値を変更しないとイケない。その場合は、「引数」を使って実行するファイル名を外から与える。引数は、実行時にスクリプト名以降に入力された値 (空白区切りで複数入力可)。引数の値は、専用の変数に入る。変数**\$1**に1番目の値、**\$2**に2番目の値、など。条件付き処理や、繰り返し処理、標準出力と標準エラー出力。シバン、**cdtools**、最終課題へと進んだ。



【日時】 2015年7月27日10:30-18:15

【題目】 スクリプト言語。Perl

【講師】 服部恵美 (アメリエフ株式会社)

【内容】 Perlは、ファイル形式の変換などに利用される。高速な処理には向かないが、比較的手軽に書いてテキスト処理が得意なことから、バイオインフォマティクス業界ではよく使われている。Perlはゆるい。同じ処理を様々な書き方で書ける言語である。シェルスクリプトとの主な違いは、Perlのほうが、正規表現・複雑な計算や処理に向いている。ファイル形式の変換 (BAM → BED) や、異なるプログラムの出力結果をマージするなど独自解析ができるのがPerlのいいところ。到達目標は、FASTAファイルのID行を変更したり、アミノ酸の出現頻度を数えたり...

Perlの文法。printを例として、行の末尾に;をつけるとか、全角記号・全角空白は使ってはいけないなど。シェルスクリプトのechoと違って、改行する場合には明示的に¥nを書く必要がある。Perlでも変数を使うことができる。基本形は「my \$変数名=値;」であり「\$変数」と書くと、変数に入っている値を呼び出すことができる。myに関連して、スコープの話。値が沢山あるときは、「配列」を使ってまとめて取り扱うことができる。配列とは、複数の値を1つの名前ですべてまとめたもの。「my @配列名=(値,値,...);」のように書く。値を取り出すときの「添字」や、添字が0から始まることなど。pushを使った値の追加、配列の要素数。配列を結合して文字列にする(join)、文字列を分割して配列にする(split)。コマンドライン引数@ARGVの説明。値が沢山あって各データに名前を付けたいときは、「ハッシュ」を使えばよい。ハッシュへのキーと値の追加。Perlでも条件つき処理や繰り返し処理が可能。Perlの比較演算子。whileやforの具体例。繰り返し処理中での次の処理へのスキップ(next)や処理の中止(last)。{}で囲んだ範囲をスコープと呼ぶ。スコープ内で定義した変数は、そのスコープ内でのみ有効 (ローカル変数と呼ぶ)。ファイル入出力、コメント、シバン、正規表現、メタ文字。



【日時】 2015年7月28日10:30-18:15

【題目】 スクリプト言語。Python

【講師】 服部恵美 (アメリエフ株式会社)

【内容】 PythonはPerlより「文法にうるさい」 → 誰が書いても似たスクリプトになる → 他人のスクリプトでも「理解しやすい」という利点がある。Perlとの使い分けの提案としては、自分しか使わない or 1回しか使わないスクリプトはPerl、他人に渡すスクリプトはPython。Pythonには複数のバージョンがあり、Pythonの2と3は大きく仕様がかわっている。資料はなるべくPython2でも3でも動くような記述にしている。2と3で仕様が変わっているところには説明を入れている。到達目標は、「BioPythonを用いてFASTA形式ファイルを操作できる」。行末に;がつくのがPerl、つかないのがPython。変数名に\$, @, %がつくのがPerl、つかないのがPython。ブロック構造を{}で囲むのがPerl、「:」とインデントで示すのがPython。データ型。「type」関数でデータ型を確認、「str」関数で数値を文字列に変換。Pythonの主な演算子。「+」は数値だけでなく、文字列にも使える。条件付き処理、繰り返し処理。Perlの配列にあたるものが、Pythonのリストとタプル。「リスト」は値の変更や追加が可能だが、「タプル」は値の変更や追加ができない。そのため、1回定義したら値が変わらないものや、辞書のキーにするものはタプルにするとうまい。リストの基本と操作（要素の追加や削除）、タプル。リスト・タプルの共通操作。要素数はlen関数、最小と最大はmaxとmin。リスト・タプルの相互変換。行頭に「#」をつけると、その行はコメントとなる。Perlのハッシュにあたるものが、Pythonの辞書（ディクショナリ）。キーと値をセットで格納。geditで楽にインデントするTips。値の整形として、formatを使うとprintの中に値を埋め込んだり、数値の桁数を揃えたりできる。Pythonのお作法、PEP8の説明。1行の文字数を長くしすぎない。PEP8に準拠しているかチェック・修正。ライブラリ、シバン、BioPython、ファイル入出力、最終課題。オブジェクト指向。時間が余ったのでvi講座。



【日時】 2015年7月29日10:30-18:15

【題目】 R : 基礎

【講師】 門田幸二 (東京大学)

【内容】 予習内容のおさらいとして、ウェブサイト「(Rで)塩基配列解析」の基本的な利用法をMacとWin両方で説明。「イントロ | 一般 | 任意のキーワードを含む行を抽出(基礎)」項目のコード内部の説明。dim関数で行数と列数を表示、行列の要素へのアクセス、関数とオプション、タブ補完、table、sort、is.element関数など。二重クォーテーション問題、入力ファイルのフォーマットの違いへの対処。サンプルデータはUSBメモリ経由で配布。

リアルRNA-seqカウントデータを数値行列データの実例として利用。列の並べ替えの一環として、length、cbind関数を説明。colSums, range, apply, min, max, colMeans, rowMeans, summary関数を用いてデータの全体像を概観。TCCパッケージ中のclusterSample関数を用いたサンプル間クラスタリング。代表的なパッケージの利用法として、Biostringsパッケージ中の関数を用いて「イントロ | NGS | 読み込み | FASTA形式 | 基本情報を取得」項目を実行。RでもFASTA形式ファイルを読み込んで、総配列長、コンティグ数、N50の計算などができることを示した。getwdとlist.files関数を用いて、作業ディレクトリの変更が正しく行えたか、入力ファイルが存在するか確認するのは最初のうちは特に重要。multi-FASTAファイル読み込み後のsubsetting、条件判定の基本の説明。ヒトゲノムファイルでも任意の染色体配列の取り出し、任意の領域の切り出しが可能。「?関数名」で関数の使用方法を知ることができる。基本テクニックの組み合わせで、FastQC実行で得られるhtmlファイル中のOverrepresented sequences項目と同じ結果が得られることを示した。GC含量計算部分の詳細な説明を通じて、塩基配列情報の読み込みから数値解析に切り替わる感覚をつかむ。multi-FASTAファイル中の全データのGC含量計算の応用編として、コンティグ (染色体) ごとのGC含量を計算可能であることを示した。



【日時】 2015年7月30日10:30-18:15

【題目】 R : Bioconductorの利用法

【講師】 門田幸二 (東京大学)

【内容】 CRANとBioconductor。R本体とパッケージの関係、パッケージの概念をlibrary関数の意味とともに説明。R本体やBioconductorは定期的にバージョンされており、定期的にアップデートしたほうがよい (5月と11月)。予習で行ったパッケージのインストール部分の説明。Bioconductorの概観。library関数で読み込んだときのメッセージと、内部的に利用する他のパッケージとの依存関係を説明。ゲノム配列情報を含むBSgenomeパッケージの説明。R本体のバージョンの違いで、利用可能なBSgenomeパッケージ数に違いがあることを2つのスクリーンで示す。BSgenomeの具体例として、ヒトゲノムの最新版パッケージBSgenome.Hsapiens.NCBI.GRCh38で読み込み後のオブジェクトを説明。subsettingや染色体ごとの配列長の概観がRだと便利であることなど。アノテーションパッケージTxDbの説明。ゲノムとアノテーションのバージョン確認の重要性と絡めてsessionInfoで自分の環境を知る。プロモータ配列取得。R ver. 3.2.0でうまくいったが、ver. 3.1.3でうまくいかなかった実例とその対処法 (個別パッケージのインストールおさらい)。

データの型というのは、見栄えに相当するものである。型は、R本体のバージョンを上げることで意識しなくてもいい局面もある。「library(help=パッケージ名)」でそのパッケージが提供する関数群を概観できる。FASTQファイルの解析。FastQC実行結果とQuasRパッケージの実行結果を比較し、異なる結果になった原因究明を通じてLinux上で動かすFastQCプログラムのオプションを調査する動機づけとなる。FastQCの--nogroupオプションをつけた結果とQuasRの結果が同じになることを見出し、Rでの検証やFastQCのKmer Content項目の挙動を学ぶ。末端塩基のトリムやRプログラミングの例外処理、様々なスキルを相補的に利用し、知識の幅を広げる。



【日時】 2015年8月3日10:30-18:15、8月26日

【題目】 NGS解析：基礎

【講師】 山口昌雄（アメリエフ株式会社）

【内容】 NGS解析でよく使われるファイル形式として、fastq, bam/sam, vcf, bed, fastaを説明。FASTQのクオリティは「記号のASCIIコード - 33」と対応する。例えば、クオリティ値が%のときは $37 - 33 = 4$ 。sam/bamは、リードをマッピングしたアラインメント情報。samはテキストデータ。bamは圧縮したsamであり、コンピュータが扱いやすいバイナリデータである。相互変換には、主にsamtoolsというソフトを用いる。「samtools view -Sb sam > bam」と「samtools view -h bam > sam」。samファイルは、@から始まるヘッダ行と、1行に1リードの情報がタブ区切りで記載されているデータ行からなる。最初の11列は必須。

vcf形式は変異情報を含む。#で始まるヘッダ行と、1行に1つの変異の情報がタブ区切りで記載されているデータ行からなる。bed形式は、ゲノム上の領域の情報を含む。エクソームシーケンスなどのターゲットシーケンスで解析範囲を指定するために用いられるほか、ChIP-seqで検出されたピークを示すのに用いる。「bamToBed -i bam > bed」。fasta形式は、NGS解析以外にもよく用いられる、塩基配列やアミノ酸配列の情報。拡張子は統一されておらず、.fa, .fasta, .fna, .fasなどが使われていることがあるが、中身は同じ。データの可視化として、Integrative Genomics Viewer (IGV)の基本的な使い方を説明。インストール自体は7/23のLinux基礎で行っている。「igv.sh」でIGVを起動。NGS解析では、サイズの大きなファイルを高速に扱うべく、インデックス（目次）ファイルが別に必要なことが多い。データのクオリティチェックとしてFastQCがよく用いられる。FastQC結果ファイル中の各項目の説明。マッピング、アセンブルの概要説明。15:30ごろに当初予定内容が終了したため、休憩をはさんでゲノムアセンブルプログラムvelvetのマニュアルを眺めながら、アドリブでアセンブルを実行しコンティグを得た。



【日時】 2015年8月4日10:30-18:15、8月27日

【題目】 NGS解析：ゲノムReseq、変異解析

【講師】 山口昌雄（アメリエフ株式会社）

【内容】 大まかな流れは、公開データ取得 → QC → マッピング → 変異検出。ショートリードでも、SNVやInDelなど様々な変異を検出可能である。検出アルゴリズムとソフトウェアをいくつか紹介。解析パイプラインとは、「あるソフトの出力結果が、次のソフトの入力ファイルとなる」連続した解析処理の流れのことである。データ取得として酵母のリファレンス取得。IlluminaのiGenomesは、リファレンスのfastaのみではなくマッピングソフトのインデックスファイルや遺伝子情報ファイルも一緒に圧縮して公開されているので使い勝手が良い。

lessコマンドでゲノムファイルやインデックスファイルの中身を確認。ヒトリファレンスの話としてデコイ配列を説明。Reseq解析は、リファレンスに対して変異検出するので、リファレンス自体がどの程度確かなのが非常に大切である。マッピングプログラムBWA ver. 0.7のインデックスファイルを作成。「ln -s」でシンボリックリンクを作成。データ取得例として、DDBJ SRAにてアクセス番号で検索し、データをダウンロードする基本形を説明。FastQC実行。QCの一環として、クオリティ30以上の塩基が90%未満のリードを削除すべく、fastq\_quality\_filterを実行。マッピング結果のbamファイルを入力としてawkで列の合計を計算し、マップされたリード数を調べるやり方。

GATK UnifiedGenotyperコマンドの使い方を確認し、SNV/Indel検出を実行。このあたりから1つのコマンドが長くなるので、コピペ用のファイルをgeditで開いてコピペでコマンドを実行した。得られたvcfファイルをlessで眺めてジェノタイプやカバレッジを眺めた。vcfファイルを入力としてawkコマンドで検出したSNV/Indelの数を確認。リアラインメントの重要性。IGVを起動して、リファレンス配列とvcfファイルを読み込み、結果を概観。snpEffの説明。



【日時】 2015年8月5日10:30-14:45、8月28日10:30-14:45

【題目】 NGS解析：RNA-seq前半

【講師】 山口昌雄（アメリエフ株式会社）

【内容】 RNA-seqの概要。リファレンスがある場合とない場合。講習会は前者のほうを行った。Illumina iGenomesで酵母ゲノムのリファレンス取得。DDBJでSRR518891を検索。一般には複数サンプルで解析するが、講習会では1サンプルのみで実行。FastQC実行とhtmlレポートを確認。fastqcコマンド実行時に「-t 2」オプションをつけると、CPU2個を使って処理するので早くなる。オリジナルはpaired-endだが、講習会では片側のみ使用。FastQC実行結果のOverrepresented sequences項目で、polyA/T tailが存在することを確認し、RNA-seqの原理から説明。fastx\_trimmerのxは、FASTQ/FASTAどちらでも対応可能という意味。資料では、最初の1塩基と書いているが、よく見ると最初の7塩基をトリムするほうがよさそうなので、8番目の塩基から使うという意味で「fastx\_trimmer -f 8 ...」とした。fastx\_clipperを用いてpolyA/T tailの除去を行うが、grepで調査した結果と明らかに異なる結果であり、全体としてFASTX\_toolkitで提供されているプログラムはバグが多いのではという議論になった。同様なプログラムとしては、fastx\_clipperのほかにtagcleanerやcutadaptがある。tagcleanerは捨てるすぎる、fastx\_clipperはあまり捨ててくれない、cutadaptはちょうどよかった、という経験談。fastq\_quality\_trimmerを用いてクオリティの低いリードを除去。一連のQC実行後に、もう一度FastQCを実行して、クオリティスコア分布を比較。アメリエフさんが普段使っているのはprinseq、最近ではtrimmomatic。マッピングプログラムとして、tophatを利用。サンプルファイルはリード数が少ないのでマッピング基準を緩めて実行。IGVで可視化。「遺伝子の発現量 $\propto$ 遺伝子上にマップされたリード数 ( $\propto$ は比例の意味)」の話から、RPKMやFPKMの基本的な考え方と数式を説明。Cufflinksで発現量を計算。応用編としてcuffmerge, cuffdiffを簡単に説明。



【日時】 2015年8月5日15:00-18:15、8月28日15:00-18:15

【題目】 NGS解析：RNA-seq後半

【講師】 門田幸二（東京大学）

【内容】 カウントデータ取得以降の統計解析。7/29のR基礎で簡単に触れておいた3群間比較用リアルRNA-seqカウントデータのみを用いて、おさらい（MacとWinそれぞれの作業ディレクトリの変更など）を含めながらクラスタリングと発現変動解析について述べた。サンプル間クラスタリングでは、TCCパッケージ中のclusterSample関数のオプションの意味や、Spearman相関係数( $r$ )から $1-r$ として距離を定義することで、数値が小さいほど類似度が高いことを表現できる数式の意味を述べた。最も類似度が高い2つのサンプルに着目し、サンプル間クラスタリング結果の縦軸の高さと、独立に $r$ から計算した $1-r$ 結果が同じであることを示した。クラスタリング結果の解釈として、群間類似度と発現変動解析結果の相関についても述べた。群間類似度が低いほど2群間比較で得られる発現変動遺伝子(DEG)数が多いことが予想され、実際にそうである(Tang et al., *BMC Bioinformatics*, 2015)ことを述べた。

上記の検証を兼ねた様々な2群間比較。最も類似度が低いヒト vs. アカゲザルを最初の例として、サブセット抽出のおさらい、DEG検出結果の解釈上重要なfalse discovery rate (FDR)の説明、緩めのFDR閾値を用いたDEG数の見積もりについて述べた。M-A plotの説明とDEG検出結果上位のM-A plot上での位置、FDR閾値を変化させたときのM-A plot上でのDEGの分布を説明し、統計的手法の説明のイントロを行った。FDR 5%条件下で、ヒト vs. アカゲザルの結果は2,488 DEGs、類似度が近いヒト vs. チンパンジーの結果は578 DEGsであり、クラスタリング結果（全体的な類似度）と同じであることを示した。non-DEG分布の意味を説明すべく、DEGがないことが予想される同一群（ヒト）の反復データの統計解析結果がDEGがわずらかし検出されないこと、2倍以上の発現変動など倍率変化でDEG検出を行う危険性、3群間比較への展開を述べた。



【日時】 2015年8月6日10:30-18:15

【題目】 NGS解析：ChIP-seq

【講師】 森岡勝樹（理化学研究所）

【内容】 ChIP-seq解析の流れを概要として掴み、自ら解析するときの足がかりとなることを目指した。講習会初めにはうまく機能していた共有フォルダ設定に軒並み不具合が生じ、ゲストOS上に数百MBに達するテストデータ群のコピーがうまくいかず、無線LAN経由で長時間かけてダウンロードする以外の選択肢がない状況になった。それゆえ、午前中はダウンロードの個別対応に多くの時間を費やした。

ChIP-seqの原理と目的（DNA結合性タンパク質のゲノム上の局在をみる）を述べた。得られるデータの良い図の例としてTP53の遺伝子領域を挙げ、抗体の特性によって様々な「山」が得られることを述べた。データ解析の流れは、QC → マッピング → ピークコール → 遺伝子注釈やモチーフ解析である。ピークコールは、enrichされた領域を山の頂点としてみたいが、ピークの形状によって用いる手法が異なる。Broad peakの場合はSICERやBroadPeakプログラム、narrow and sharp peak with low noiseの場合はMACSやF-SeqやSWAsプログラム、S/N is non-good or low peakの場合は、F-SeqやZIMBAがおすすめである。ピークコール後の解析は、HOMERが一通りの解析パイプラインが揃っているのでおすすめである。NGS plot、Galaxy、ChIP-seqデータをとにかく可視化するSratailer (Oki et al., *Genes Cells*, 2014)についても触れた。実際の解析データは、ヒトリンパ芽球様細胞の細胞株、CTCF抗体のENCODEプロジェクトのものであった。



## 2-7. アンケート結果

講習会終了時の受講生の課題、講義内容に関する課題、講習会運営に対する改善点等を明らかにし、次回の講習会に活かすことを目的に、NBDC 担当者がアンケート調査を実施した。受講希望者 136 名うち、1 項目以上の実際の受講者数は 110 名であった。これらの 110 名の受講生をアンケート対象とした。調査対象者に対して、アンケート用ウェブページの URL を認証 ID、パスワードとともに配信し、アンケートへの回答依頼を行った。調査期間は、当初の 2015 年 7 月 22 日～9 月 11 日であった。9/14 時点で未回答であった 16 名について、個人宛メールによる督促を行い、最終的に 104 名（回収率 94.5%；未回答者 6 名）から回答を得た。最終的な調査期間は 2015 年 7 月 22 日～9 月 18 日である。

講義全体についての結果を表 6 に示す。計 4 つの設問に対し、①そう思う、②ややそう思う、③どちらともいえない、④あまりそう思わない、⑤そう思わないの 5 つの選択肢を用意した。また、上記 4 設問について自由記述欄（アンケートの設問 16 に相当）も設けた。寄せられた計 88 件の生の記載内容は、付録 1 に示した。

設問	①	②	③	④	⑤
1. 講義内容の分量は適切であった	69	22	9	4	0
2. 講義のレベルは適切であった	61	34	7	2	0
3. 講義で取り上げられた事柄は興味ある内容であった	87	15	2	0	0
4. この講義で学んだことは今後役に立つと思った	94	9	1	0	0

表 6. 講義全体に対するアンケート結果

講義方法についての結果を表 7 に示す。計 6 つの設問に対し、上記と同じく 5 つの選択肢を用意した。また、「分かり易かったところ・分かりにくかったところについて、講義項目や内容など具体的にご記入ください」という自由記述欄（アンケートの設問 21 に相当）も設けた。寄せられた計 82 件の生の記載内容は、付録 2 に示した。

設問	①	②	③	④	⑤
1. 説明の仕方はわかり易かった	72	28	4	0	0
2. 講義の進度や時間配分は適切であった	55	34	12	3	0
3. 配布資料などの教材は適切であった	82	19	2	1	0
4. 講義に TA は必要である	92	9	3	0	0
5. 講義の内容は全体的に理解できた	58	40	5	1	0
6. 総合的に評価して、あなたはこの講義に満足しましたか	83	20	1	0	0

表 7. 講義方法に対するアンケート結果

各講義（項目）についての講義内容の理解度に関する結果を表 8 に示す。「講義内容を十分理解できたか？」という質問に対し、①そう思う、②ややそう思う、③どちらともいえない、④あまりそう思わない、⑤そう思わない、⑥受講していないの 6 つの選択肢を用意した。

実施日	項目	①	②	③	④	⑤	⑥
7月22日	PC環境の構築（Bio-Linux 8とRのインストール状	46	7	4	1	1	45

	況確認)						
7月23日	Linux基礎	58	11	1	1	0	35
7月24日	スクリプト言語(シェルスクリプト)	44	20	2	1	0	37
7月27日	スクリプト言語(Perl)	30	28	5	4	0	37
7月28日	スクリプト言語(Python)	32	29	3	1	0	39
7月29日	データ解析環境R：基礎	42	18	5	2	0	37
7月30日	データ解析環境R：Bioconductorの利用法	42	18	5	4	0	35
8月3日	NGS解析（基礎）	62	32	3	1	0	6
8月4日	NGS解析（ゲノムReseq、変異解析）	43	36	7	3	2	13
8月5日	NGS解析（RNA-seq）	48	35	8	1	0	12
8月6日	NGS解析（ChIP-seq）	13	23	17	13	2	35

表 8. 各講義の講義内容の理解度に対するアンケート結果

上記以外に、感想、要望、意見、あるいは改善のための提案として、自由記述方式の5つの設問を行った。設問内容と回答件数、および付録番号との対応付けを表9に示す。

設問	回答件数	付録番号
1. 講義内容に取り入れて欲しいものについてできるだけ具体的にご記入下さい	60	付録3
2. 本講習会を受講後、次のステップとして必要または身につけたいと思う解析スキルについて、できるだけ具体的にご記入下さい	48	付録4
3. NGS解析以外で、開催して欲しいと思う講習会がございましたらご記入下さい	38	付録5
4. 設備、講義室、受講人数、TA、その他、講義全体に関する要望・コメントなどがございましたらご記入下さい	35	付録6
5. その他、ウェブで公開・提供して欲しいバイオインフォマティクス教材に関するご要望などがございましたらご記入下さい	28	付録7

表 9. 自由記述方式の5つの設問、回答件数、および付録番号との対応関係

## 2-8. 総評

### 謝辞

本講習会は、平成 26 年度に計 10 日間かけて施行実施した NGS「速習」コースの結果を踏まえ、平成 26 年度受講生のコメントや要望をできるだけ反映させた構成にした（表 1）。昨年度の報告書中で述べた主な課題は、項目間の連携強化であった。その具体的な対策として、できる限り少人数の講師陣でハンズオン講義を行うことを目指した。当初予定の A 日程（7/22～8/6 の計 11 日間）のうち、5.5 日分を快く引き受けてくださったアメリエフ株式会社の山口昌雄 講師および服部恵美 講師のおかげで、講習会の全体的な統一感が大幅に向上した。アメリエフ株式会社からはまた、2 名の TA（K 氏、O 氏）を派遣していただいた。特に K 氏は、7/28 の Python 講義終了後に、受講生から希望のあった vi 講座の講師を急遽、担当した。本講習会の中核として多大な貢献をいただいたアメリエフ株式会社の諸氏に深く感謝する。また、ChIP-seq (8/6)の講師担当となった理化学研究所の森岡勝樹 講師にも深く感謝したい。諸事情により限られた準備期間での登板となり、また所属先変更などかなり厳しい状況且つ体調不良の中、よりよい講義資料作成にギリギリまで最大限尽力いただいたことも記しておきたい。

本講習会では、昨年度に引き続きボランティア TA（講義補助）を募集した。TA 枠での申込者 2 名、および一部項目で TA として尽力いただいた 4 名の受講生にも感謝したい。世話人（門田）に全くゆとりがなく、ねぎらいの言葉をかけられなかった方もいるが、この場を借りてお礼申し上げる。特に早稲田大学大学院の N 氏には、Mac のテキストエディタに関する注意点など全般的に助けていただいた。東京大学大学院の S 氏、T 氏、A 氏にも、実質的なボランティア TA として幾度となくタイムリーなサポートをしていただいた。B 日程を含め、計 4 週間に渡って講習会会場を確保できたのは、東京大学大学院の清水謙多郎 先生のサポートによるものである。また、ネットワーク環境などインフラ関係は、寺田透 先生が全面的に担当した。貸与 PC の環境構築や PC 管理などは三浦文 氏が担当した。講習会の裏方として尽力いただいた諸氏に感謝したい。

開催時期は、昨年度の 9 月から一か月以上前倒しとなった。これは東京大学の学期制度が平成 27 年度から変わり、7-8 月の時期でしか講習会会場を確保できなかったためである。結果的に昨年度受講生の一部の要望を反映した形になったが、それに伴い、ある程度予想はしていたが世話人（門田）が全くゆとりのない状況に陥った。アグリバイオインフォマティクス（主催機関 1）は、今年度 1 名スタッフが減った状況下で、昨年度よりも多くの大学院講義受講生への対応を行っていたためである。世話人は、年度初めから 7 月初旬まで本務の大学院講義が入っており、7 月下旬の NGS ハンズオン期間に入ってもまだ後半の講習会資料が完成していないという精神状態で講習会を迎えた。この状況との因果関係は不明であるが、結果的に本講習会は NBDC（主催機関 2）に多数の人員を派遣していただいた。NBDC は、TA、ビデオ撮影、講義映像編集作業、講習会前日の OA タップ配置や講習会後の片づけ、その他事務関連全般を担当した。誰が何を担当したかは完全にお任せだったため把握しきれないが、実務担当の諸氏（白鳥、佐藤、舘澤、眞後、三橋、大波、川嶋、宮崎、櫛田、信定、佐久間、森、高田）に感謝したい。特に白鳥亜希子 氏は、講習会全体の調整役として大車輪の活躍をしていただいた。本報告書のアンケート結果は、白鳥氏がまとめたものに基づいて作成した。当初の想定以上の労力になったと思われるが、相当数の人員派遣にご協力いただいた高木利久 NBDC センター長、星潤一 企画運営室長にも感謝したい。また、NGS 用カリ

キュラムの取りまとめ役、およびNGS講習会のアドバイザーとしてご尽力いただいた藤博幸 先生の名前も記しておきたい。

本講習会の受講申込み業務は、昨年度に引き続き HPCI 人材養成プログラム（共催機関 1）の寺田朋子 氏が担当した。新年度開始直後からのウェブサイト作成であったが、受講申込みサイトの公開を 4/15 に設定できたのは、NBDC との緊密な連携のもとで迅速に仕事をこなした寺田氏のおかげである。また、DBCLS（共催機関 2）の小野浩雅 氏、大田達郎 氏は TA としてだけでなく、Twitter や 8/6 の ChIP-seq でのデータの速やかなダウンロード作業周辺でも尽力いただいた。また、受講生として参加しつつも、他の受講生へのアドバイスや、速やかな問題解決に協力いただいた諸氏にも感謝したい。

## 運営全般について

本講習会では、全般的に昨年度好評であった部分を残しつつ、不満や要望のあった部分の改善を行った。アンケート結果の解釈と本質的に同じであるが、講習会実施内容・方針・運営については問題なかったと思われる。TA 申込者数同様、項目あたりの受講人数が昨年度に比べて減少したのは、おそらく①全日程受講者のみから部分受講可にしたこと、および②年度講習会の講義映像・資料を含めて全て無条件公開にしていることに起因すると思われる。同様な講習会は他所でも行われているものの、講義映像や資料まで無条件公開しているところは多くない。特に、世話人は講習会にわざわざ足を運ばなくても自習で十分に理解可能な講義資料の無条件公開をポリシーとしている。A 日程の計 11 日間のうち 4.5 日分を担当した世話人の講義資料は、このポリシーに則って作成し、講習前日までに公開した。5.5 日分を担当したアメリエフの講義資料も一週間以上前から公開した。もし自分が勉強する側の立場であり、上記情報を Twitter や SNS など知っていたら、受講申込みをしない。それゆえ、136 名の受講申込み者のうち、実際の受講者が 110 名であったことに対して特に驚きはない。

世話人の本講習会運営に対する基本的な考え方やポリシーの理解は、必須事項として掲げた予習の範囲内に含め、講習会期間中も何度か再確認した。これは、全てのヒトを満足させる講習会は事実上不可能であり、世話人の方針を理解・納得した受講生が多数派となることを目指したためである。そのうえで、きっちり予習を行ってきた受講生の満足度をできる限り高めることを本講習会の主要な目的として掲げた。強権的・独善的・独裁的と批判する余地はあるかもしれないが、世話人に対する異論や不満がほとんどなかったことから、狙い通り運営方針が多くの受講生に認知されていたといえるだろう。運営方針に対する批判票が入りこむ余地を与えなかったともいえるかもしれないが、それでもアグリバイオ大学院講義と同じく日本最大の受講人数規模を維持している。費用対効果の点からも、多数派の要望や需要を満たす効率的な講習会を今後も目指していきたい。もちろん、大規模な講習会であるがゆえに実施が困難なウェブツール系（DDBJ pipeline や Galaxy）は、乳酸菌学会誌の NGS 連載や講習会中の簡単なデモなどを通じて、できる限り効率的に伝授していきたいと考えている。

## 2-9. 来年度実施方針について

研究代表者は、NGS 解析の実務から遠く離れた理論屋である。基本的な生物学の知識に乏しいた

め話についていけないこと、下記3理由により興味をもってヒトの話を聞く精神的ゆとりが全くないことから、個別の研究相談にも基本的に応じていない：①少人数で運営するアグリバイオインフォマティクスの活動、②本 NGS 講習会活動、③乳酸菌学会誌の NGS 連載。①は、研究代表者の本務であり R を中心として教えている。③は、①の枠組みでは困難な Linux 環境での NGS 解析のノウハウを豊富なウェブ資料とともに述べており、研究代表者個人の基本理念（講習会にわざわざ足を運ばなくても自習で十分に理解可能な講義資料の無条件公開）に完全合致したアウトリーチ活動である。これまでの NGS 連載記事の内容は、②の平成 26 年度受講生の意見を反映させたものである。また、平成 27 年度受講生の要望は、今後の連載記事の内容に反映させる予定である。さらに平成 27 年度分の講習会映像および資料も、全て公開している。

来年度実施については、研究代表者個人としては、(平成 27 年度実施もそうであったが) 消極的である。負担が大きいことが一番の理由であるが、全体方針を定めた世話人自身が多く項目を担当し、少人数の講師陣で全体の連携を強化した講習会であったことが、受講生の高い満足度につながったものと思われる。少なくとも研究代表者はそう判断しているため、この基本方針からのずれが大きいものほど受講生の満足度は低下するであろう。また、講義内容が今年度と同レベルのものにとどまるのであれば、来年度も継続実施する意義はそれほど大きくはない。したがって、もし引き続いて来年度も世話人を務めるのであれば、講義室の確保状況にも大きく依存するが大まかに下記方針で行う：

### 1. 部分受講可能、少人数の講師陣

平成 27 年度と同じく、受講生本位の講習会として行う。講師もできるだけ減らし、さらなる連携強化を図る。講習会後半部分の項目は、前半部分や予習内容を理解済みであるという前提で行う。多数派の意見を取り入れて、ある程度は現実的な対応（多少復習を含める）を行う。予習してこなかった受講生に対してはサポートをしないなどの措置はとる。完全に予習をきっちり行ってきた超優秀な受講生は outlier になるため、逆の意味（優秀すぎるので講習会出席の必要なし）で配慮はしない。

### 2. NGS 解析部分を中核として実施

情報のアップデートと1日あたりの分量増加を基本とする。今年度アドリブで行ったゲノムアセンブリ部分や、関連プログラムのインストールを含めて行う。また、メモを取るのに労力を割かなくて済むよう、講義資料に必要な情報を含める。トランスクリプトームアセンブリ、アノテーション、ウェブツールの一例として DDBJ pipeline の簡単なデモも可能な限り含めたい。IGV の基本的な利用法についても、もう少し丁寧かつ詳細に説明したほうがよいと思っている。

### 3. プログラミング言語はなし

昨年度・今年度と実施し、統合 TV で公開されているのでそれを見てもらう方向。受講生アンケートで複数人が書いていた「最終課題とそれまでの部分の乖離」については、誰がやっても難しいところ。あまり丁寧に手順を踏んで説明しても間延びする可能性があること、全体の進捗を揃えながら行うことは難しいことなどが挙げられる。プログラミング言語（シェルスクリプト、Perl、Python）のどれか1つに絞り、それを2日間程度かけてじっくりやるという選択肢も無きにしても非ずだが、現実問題として、2の NGS 解析部分の充実化により

多くのエフォートを割きたい。もちろん、シェルスクリプト化は NGS 解析部分のより実践的な内容（解析パイプラインの作成と実行に相当）として必須事項だと思われるため、このあたりは NGS 解析の一部として含める予定である。

#### 4. 統計解析を強化

今年度の RNA-seq 後半部分の拡充で対応。0.5 日分を 1 or 2 日分に増加。データ解析環境 R の基本的な利用法を理解済みという前提になるため、今年度実施した 2 日分の R の講義をなくすかどうかは悩ましいところ。NGS 解析の一通りの講義が終わったあとに、統計解析を中心とした講義を手段として R を用いるような内容で行うのが現実的なのところであろう。基本的には、世話人の得意分野を中心に、実験デザインごとに複数の解析例の紹介や、アルゴリズムごとの出力結果の特徴を紹介するような内容を想定。また、作図のリクエストにも応えたい。Pathway 解析や GO 解析は、行うとしたらこの枠組みでの説明になる。

#### 5. Linux 環境構築や予習は継続実施

乳酸菌学会誌の NGS 連載を用いた予習、7/22 の解析環境の自習、7/23 の Linux 基礎の枠組みでの実習用データダウンロードやプログラムのインストール講義は好評だったように思われる。特に、7/22 初日の自習は、受講生が一時に殺到しないため運営側としても心穏やかであった。2 の NGS 解析の中核部分へとつなげる上で重要なイントロ部分であるため、これらについては、需要はそれほど多くはないかもしれないが来年度も行う必要があると考えている。1.でも述べた優秀すぎる受講生への無配慮とも関連するが、現実問題としては復習の意味を込めて Linux コマンドの説明をある程度含める必要があり、来年度以降もそうする予定である。もちろん、この点については来年度実施時の基本理念説明の部分で明示しておく必要はあるだろう。

#### 6. B 日程はなし

受講日程の選択肢を広げるという意味で受講生には好評であったが、実施側の負担が大きいためである。A 日程（～8/6）から間が空いたのは、講義室をまとまって確保できる日程が 8/26-28 しかなかったためである。追加受講を受け付けたことも影響しているが、B 日程のみの参加者で Linux や R の事前準備を行ってこなかったものが少数ながら存在したため、それらの受講生への対応で疲弊したことも大きい。また、A 日程の枠組みでは一連の流れで説明できたが、B 日程までの期間が空いたため、Linux や R の項目内容を思い出させながら説明するのに苦労した（8/28 の最後を担当した門田の感想）。また、実際の受講人数的にも、今年度実績で追加受講申込みを受け付けなければ 100 名弱で収まる。

#### 7. 初級・中級などのレベル設定はなし

常識の範囲はヒトそれぞれとも関連するが、受講生に予断を与えないためにもレベルに関する情報は、来年度以降は提供しないほうがいいかもしれない。

### 2-10. よりよいハンズオン講義のために

昨年度の報告書中の記載内容と大枠で同じであるが、備忘録的に多少アップデートしつつ一般論として記載しておく。受講生は、自身の研究を遂行する時間を削ってハンズオン講義に参加している。ハンズオン講義実施主体は、予習用のよりよい教材を整備するのはもちろんのこと、やるべき

予習をきっちり行ってきた受講生の多数派を満足させられるよう、受講生側の気持ちになって最大限の努力をすべきであろう。受講生側も権利の主張ばかりではなく、しかるべき義務（予習やアンケートへの回答など）を果たすべきであろう。もちろん建設的な提案はぜひ寄せていただきたい。

- **主催側：受講者数の 1/3 程度以上の USB メモリを用意**

ネットワーク（有線、無線 LAN）の不具合に対応するため。当日ウェブ経由でダウンロード予定のファイルを USB メモリに置いておき、万が一ネットワークが使えなくてもハンズオン講義を継続できる体制にしておく。平成 27 年度は、研究代表者が担当したいくつかの項目で数百 MB レベルのデータを USB メモリ経由でコピーしてもらったので、より現実に近い規模のデータ解析を行うことができた。

- **主催、受講生側：質問関連**

実は配布資料が間違っていただけだったとしても、受講生は自分だけが間違っているのではという不安感をもつ。受講生の質問も、全体で共有可能な講師への質問と、動作確認程度の（TA への）も質問に切り分けると効率的である。本講習会では、「講師側に手を挙げる全体質問」と「後ろ向きで手を挙げる個別質問」に切り分けた。本会場は、設備的な問題によりデフォルトはマイクが 2 本のみである。平成 26 年度は、講師と世話人用のみで、質問時は世話人が走ってマイクを渡した。受講生から改善要望があったため、平成 27 年度は拡声器を 2 つ用意し、受講生の質問時に用いたが、接触不良やハウリングなどの問題や質問者へ拡声器を向けるまでのタイムラグがあるなど、それほど効果的ではなかった。平成 27 年度受講生の提案を反映して、質問者が後方においてある拡声器に自ら移動するのを徹底させることで、タイムラグや TA の負担軽減を図るのがよいだろう。

- **主催、講師側：配布資料やスライド**

18 pt 以上の大きさにスライドを作成すべし。「A4 横、4 アップ、両面印刷」の印刷物を配布した。印刷に伴うインクの消費量を減らすため、スライドの背景を白とするのも基本である。詳細かつ丁寧な講義資料は、復習にも役立つほか実習時においていかれた場合でも自力での巻き返しを可能にするため、**実習の成否を握ると**いっても過言ではない。受講生がノートをとらなくてもいいように、Tips を含めて全て記載しておいたほうがよい。コマンド実行結果も含めてスクリーンショットをとったものを講義資料に載せておくとよい。これは、初心者は自分の打ったコマンドが正しく動作しているかどうかの確認ができないため、「このような画面になっていれば正解」というものがあると安心するからである。

- **主催、講師、受講生側：スクリーン下部に注意を払う**

これは、講師の演壇が低い位置にあつて、受講生が講師を見下ろすような講義室の場合には特に気にする必要はない。しかし、本会場のような平面の講義室の場合には、スライドを投影するスクリーンの下部は、前方の受講生の頭に隠れてしまい、後方の受講生には見えないことが多い。本講習会では、Linux コマンド実行画面を見せる場合に、ターミナル画面下部がスクリーン画面の上から 2/3 程度の位置になるように配慮した。講師が無意識に画面を広げる場合があるので、ターミナル画面下部がなるべく講義室前方の受講生の頭より上になるように気をつけたほうがよい。同様な理由から、ターミナル画面に限らず、**講師はスライド中の記載情報はなるべく上のほうにシフトさせるように徹底する**（つまり後方の受講生が見づらい位置に重要

な情報を載せない)。主催側で徹底しきれない部分もあるため、受講生側も見えない場合や次のスライドにいてもらいたくない場合にはその場で意見表明すべき。

- **主催側：予備の（ノート）PC を可能な限り確保**

アンケートにも書かれていたが、Mac のほうが Win に比べて不具合遭遇率が高い。そのため、Mac ユーザにはあらかじめその旨を伝えておき、環境設定済みのノート PC も貸与するのがいいかもしれない。

- **主催側：バックアップ用の環境構築済みイメージファイルを複数種類準備**

平成 27 年度講習会では、7/22 時点の ova ファイル、7/23 の Linux 基礎終了時点までの ova ファイルなど、計 4 通りのイメージファイルを予め作成しておいた。各イメージファイルが 5GB 程度に達するため、USB メモリも 32GB 以上のものが数十個必要である。ファイルのコピーに十数分程度かかるため、解析環境の再構築に計 20 分程度を要する。しかし、それでも一から環境構築をやり直す手間に比べれば、持ち込み PC に対しても圧倒的に短時間で対応可能である。実際にこれで救えた受講生が少なくとも 10 人程度はいたように思われるため、可能な限り用意しておいたほうがいいだろう。

- **主催、講師側：OS の違いに気をつける**

上記の代替 PC とも関連するが、講師の PC と受講生持込 PC の OS の違いによる不具合も存在する。動作確認は、可能な限り Win と Mac の両方で主催側が行っておき、講師にエラー事例報告するという形がよいだろう。特に R は GUI 画面が OS 間で若干異なるため、できれば両方の OS 下での操作説明資料を用意しておくことよい。平成 27 年度は、講義室の 2 つのスクリーン有効活用例として、研究代表者による R の講義(7/29, 30)で Win ユーザの受講生は Mac の GUI 画面を、そして Mac ユーザの受講生は Win の GUI 画面を見るように促し、どちらの OS で講師がしゃべっても受講生側が順応できるようにした。

- **講師側：講義資料送付などの締切を厳守**

多少認識のずれがあった場合でも、主催側のチェックや修正依頼に対応する余裕が生まれる。

- **講師側：受講生への事前ソフトウェアインストール指示は早めに**

ソフトウェアのバージョンなども忘れずに。主催側は、受講生からの各種質問メールに忙殺されます。できるだけ丁寧なインストール手順書を作成し、想定されるエラーへの対処法なども示しておくとなおよい。

- **講師側：解析データの準備**

計算時間が数分～十数分程度で終わるくらいのデータ量の解析データを用意すべし。少なすぎて一瞬で終わるようなデータ量だと、現実との乖離がありすぎる。概ね数十 MB 程度が目安だろう。この程度であればアセンブリも現実的な時間で終わられると思われる。当日のネットワーク不具合を想定して、早めに主催側にデータ送付すべし。

- **講師側：コピペ用コード集も用意**

例えば NGS データのマッピングなどは、一連のコマンドが数行にも及ぶことがある。講師が手入力で打ち込むというポリシーだったとしても、受講生自身が遅れを挽回しやすいように、実習で用いる一連のコマンド集を別に用意しておくことよい。また、パワーポイントなどで作成したコードは、オートコレクト機能のため、オプションの「ハイフン」が「ダッシュ」に変更

されていて、コピペでエラーが出る場合が多い。これでハマる受講生も一定の割合で存在するため、注意喚起する必要がある。平成 27 年度は、予習の枠組みでオートコレクトや二重クォーテーション問題について注意喚起していたため、このような問題は起きなかったが、引き続き前知識として与えておくべきであろう。

- **講師側：「手を動かしながら聞いてください」は極力避ける。**

受講生はコマンド入力に慣れていないことに気を配るべきである。講師はコマンド入力にかかる時間を十分に与え、コマンド実行後の説明は、なるべく足並みが揃ったことを確認してから進めるべきであろう。同様に、コマンド実行結果の画面はなるべくいじらないのが鉄則である。入力ミスなどで遅れがちな受講生は一定の割合で存在する。エラーに遭遇した場合、受講生は自身の PC 画面と前で示しているスクリーン画面との違いを確認しようとする。しかし、講師が実行後の時間を十分に取らずに別のことを実行すると、画面がスクロールし、正しい入力かわからない、正しい動作結果がわからないといったことがおきる。いくら TA がついていたとしても、TA ごと置きざりにされてしまうこともある。実行直後に結果の説明をしても、エラーに遭遇した受講生は、エラーの原因究明で精いっぱいのため、エラー関連のノートをとっている間に説明を聞きそびれる場合が多い。それは結局、正しいコマンドの打ち込みだけに追われて、肝心な中身の理解にまでたどり着けないという不幸な結果を生む。尚、丁寧な講義資料作成は、この種の不満を緩和する効果がある。

- **講師側：「これ」や「あれ」のような指示語は使わない。**

指示語は説明時に前の画面を見ていないと何のことを言っているのかわからなくなることがある。反射的に使ってしまう傾向にある場合は、手を動かすのを一旦やめさせて受講生の視線をスクリーンに向けさせてから話すように心がけるとよい。

- **講師、受講生側：受講生の PC 環境は多様だということを認識する。**

これは OS の違いに気をつけることと本質的には同じである。同じ Bio-Linux 環境で同じインストールコマンドを打ち込んでも `nkf` コマンドのインストールに失敗する受講生が現実存在する。同じ一連のコードを打ち込んでも実行結果が微妙に違うことはしばしばある。講師がスクリーン上で示す結果のみが、唯一無二の結果とは限らない。「スクリーンと同じ結果になっていないヒト？」と確認しながら進めていくとよいだろう。特にざわついているときは何か起きていたと思ったほうがよい。受講生も積極的に TA に確認し、周りがざわついているときは勇気をもって講師に確認しよう。

## 付録 1

**設問内容：**講義内容の分量およびレベルは適切であったか、講義で取り上げられた事柄は興味ある内容であったか、この講義で学んだことは今後役に立つと思うか、について講義名も含め、内容を具体的にご記入下さい（アンケートの設問 16 に相当）。

計 88 件の生の記載内容を以下に示す。主催側の主観で、32 件の改善・検討の余地ありのコメントと、56 件のポジティブなコメントに分けた。前者のほうが重要であるため、特に今後の検討課題と思われる部分をハイライトさせた。

**改善・検討の余地あり 1：**Python は、もう少しレベルを上げた方が良いと感じました。NGS 解析基礎は、より実践で使用できるレベルの内容にした方が良いと感じました。ゲノム Reseq.変異解析、RNA-seq 前半は、より実践的な内容にした方が良かったと思いました。

**改善・検討の余地あり 2：**トラブル対応のため時間に余裕を持たせておくことは必要だと思うが、私としてはもっと 1 日あたりの分量を多くしてもいいのではないかと思いました。講義の内容自体は今後役に立つと思いましたが、講義で紹介しただけの方法 (Fseq や Galaxy) のほうが使えそうで、そちらをもっと説明して欲しいと思う場面がありました。

**改善・検討の余地あり 3：**perl python は難しかった（仕方ないとは思いますが）ChIPseq の講習は大変面白かった。時間が足りなくて最後まで行けなかったのがとても残念です。2 日あってもよいのかもしれない。

**改善・検討の余地あり 4：**シェルスクリプト、Perl、Python の講義では、個別のコマンドについてはよくわかったが、組み合わせた長いスクリプトの理解が難しく感じた。そのため個別のコマンドを用いたクイズは簡単だったが、最終課題で急に難しくなる印象だった。スクリプトが読めることが目的の一つかと思うので、複数のコマンドを使った長いスクリプトから、アウトプットを答えるようなクイズをもう少し多くやってもらえたらいいように感じた。

**改善・検討の余地あり 5：**大変有意義な講義内容であったため、前半のスピード・内容量で最後まで実施して頂ければより実りのある集中講義になると思った。

**改善・検討の余地あり 6：**R の基礎の説明に割く時間をもう少し短時間にして（予習必須とのことですので）、個々の解析法について解説する時間や、今回取り上げた以外の解析法について説明する時間を多くとっても良い気がしました。

**改善・検討の余地あり 7：**perl, python の講義は基礎でしたが、最初で教えてくれる内容と、最終課題のレベルに差がありすぎた。また、省略できるところなどが、説明なく急に出てきたりして、わかりにくかった。

改善・検討の余地あり 8：150727Perl 入門：スライド 9 5 演習 9 の訂正内容 9，10 行目をウェブにアップしていただけないでしょうか。翌日の Python 入門と合わせて、イントロに対応させ、後半部に各言語を使ってどのような処理ができるかの具体例、例えば、最終課題のようなもの（目的、使用コマンド、結果）をリストしていただければと思います。

改善・検討の余地あり 9：タイプされたスクリプトの文字がスクリーン上で小さくて見づらかったです。OHP などでも構文をお示しいただくなどしていただけると、文字も大きく、文字がスクロールされず、見やすいかもしれません。

改善・検討の余地あり 10：興味ある内容であったかについて、世間一般では動物細胞の NGS による解析が盛んなのかもしれないが、私自身はバクテリアゲノムの解析例などを取り上げてほしかった。

改善・検討の余地あり 11：アメリエフと理研の先生方の講義は講義資料外での内容が多く、余った時間もアドリブでの講義があったため非常に面白かったが、その分当日・前日の内容を理解していないとフォローするのが大変だった。特に Python の講義日に行われた Vim に関するレクチャーは個人的に非常に興味のある内容だったが、パッケージのインストールなどを同時に進行している関係でフォローするのが難儀だった。コマンドライン上でのエディター作業は重要なので、そちらの講義をコマとして(Shell の日などに)取っていたらとても有用だったと思う。全体としては非常に有益なセミナーだった。個人的に欲しかった講義として、最終日に Shell(上級)などの枠で、実際に Shell の日に習った内容とそれ以降の実際の作業を基にどのように現場で Shell を組んで仕事を飛ばしているのか、実例を基に教えてもらえると、ハンズオンとしての全体の復習になり、かつ現場での実例を見れて良かったのではないかと思う。

改善・検討の余地あり 12：24 日のシェルスクリプトと 25 日の Perl を 1 日にまとめ、26 日の Python はもう少し深く教えていただければ嬉しかったです。

改善・検討の余地あり 13：自分が知らないことも多々あり、勉強になりました。ただ、日によって難易度の違いが”ものすごく”あり、もう少し平均化できたらいいのではないかと思いました

改善・検討の余地あり 14：スクリプト言語の講義の時間配分や分量があまり適切でないと感じました。もう少しスムーズに授業を進めて色々な事を細かく教えて欲しかったです。

改善・検討の余地あり 15：スクリプト言語とゲノム Reseq、変異解析、及び RNA-seq では資料をもう少し細かく作って頂けると助かります。考えていたり、メモしている最中に講義の話が進んでしまい、結構重要な情報を聞き逃してしまうことが多々ありました。資料に書いていない情報で有用なものが多かったです。

改善・検討の余地あり 16：スクリプト言語 (Perl) とスクリプト言語 (Python) の講義、中級と銘打たれていた割には基礎的な内容であり、期待していたものとのズレを感じた

改善・検討の余地あり 17：設問の 14 に関してですが、講義で出された練習問題がもう少し実用的なものの方が良いのではないかと思います。例えば perl 入門での最終課題では「複数の遺伝子全体でのアミノ酸の組成を求める」ということが課題になっていましたが、実際には個々の遺伝子についてアミノ酸組成を求めることはあっても複数の遺伝子全体のアミノ酸の組成を求めることはないと思います。

改善・検討の余地あり 18：最終課題でかなり飛躍があったので、最後がやや消化不良だった (プログラミング)

改善・検討の余地あり 19：シェルスクリプト, Perl, Python。講義そのものはとてもためになったのですが、1日完結ということもありなかなか定着は難しい。後半の解析の実践編でもそれらを使った実習が組み入れられるとよかったと思います。

改善・検討の余地あり 20：全般に Linux のコマンドなどは予習前提の講義だったはずなのですが、結局コマンドライン操作に講義の時間を取られてしまった感があります。もう少しすっぽかしてもよかったのでは。

改善・検討の余地あり 21：ゲノム Reseq、変異解析と RNA-seq はこの解析を目指して受講する人も多いと思うので、それぞれ2日づつにしてもっと深くやってもらいたい。Perl や Python も大事だがこれらは NGS 解析に直結しない部分もあるため、他の講習に譲り、上記を拡充しても良いかともう。

改善・検討の余地あり 22：スクリプト言語の講義はもっとレベルが高い内容がほしいです。そして、どうやってスクリプト言語を NGS で使える例があれば嬉しいです。

改善・検討の余地あり 23：私は貸与 PC を使ったので特に迷うこともなく講義についていけました。ただ今回の講義でアメリカ様の用意したサンプルデータが1000リードと少なかつたため実際にソフトを動かしてマッピングする際に正確にできなかったのかなと思いました。

改善・検討の余地あり 24：ハンドアウトがあったため、メモが取りやすく、後日復習する際に大変役に立った。「ChIP-seq」の午後の講義内容の資料がなかったのが残念だった。Perl, Python について、最終問題で突然難易度が上がったように感じた。もう1ステップ、簡単な問題を用意していただけたら、より理解が深まると感じた。「ChIP-seq」の講義で課題が出されたが、後日知識の習熟

を確認する際に役に立った。ぜひ他の講義についても、実際の利用場面を想定した課題を提供していただきたい。

改善・検討の余地あり 25：ChIP-seq の講義は量が多く、消化不良であった。興味深い内容でもあるので、もう少し時間があると（例えば2日間とか）よいかもしいない

改善・検討の余地あり 26：スクリプト言語で、レジメを丁寧にまとめていただいたのでしっかり理解出来、基本は知ることができましたが、レジメを見ればわかることがほとんどでした。もう少し早く進めてもっと多くのことが出来たのではないかと思いました。逆に R やゲノム解析はやったことがなかったので、専門用語が理解出来ず、操作は追いつけるのですが何をやっているのかしっかり理解出来ませんでした。

改善・検討の余地あり 27：今回は、メタゲノムも加えてもらえると嬉しい。

改善・検討の余地あり 28：初心者にも解り易い内容で、講義内容は十分理解できた。perl の最後の課題が少し難しかったです。

改善・検討の余地あり 29：講義よりも実習が多く、他の授業にない刺激をいただきました。先生の人数が少なかつたおかげで、内容の繰り返しや初歩的知識の再確認が少なくうれしかったです。全体的なつながりを示していただければさらに嬉しいです。服部先生の Unix とスクリプトの講義は他に比べ優しく感じました。講義は `fasta` を読み込める程度のところまで終わっていますが、もしできることならその後どう進めばよいのか（教科書、ゲノム解析にあたっての勉強法、見るべきホームページや学術雑誌など）を示していただければ、さらに嬉しく存じます。また他の講義とのつながり（どうすれば他の講義のレベルまで行けるか、解析例に役立てられるか）を示していただければ、我々の勉強への励みになると思いました。森岡先生の講義は尻切れになったのが少し悲しかったです。どこまで勉強し、どこまで解析すれば、どのような論文執筆に到達できるのかについて、もう少しお話しいただけると嬉しいです。また、門田先生、山口先生のお話の中に、ゲノムの `de novo` 解析や `annotation` 解析もあると嬉しく存じます（これらについての授業が来年あるなら、受講したいです）。

改善・検討の余地あり 30：講義内容は NGS の基礎から実践へと結びつける上で非常に有意義なものであった。しかしその一方で、1日の講義時間がしっかり確保されているにも関わらず、生データを実際に扱った練習が不足気味に感じた。これは8月3日から8月6日の講義すべてを通して感じたことである。

改善・検討の余地あり 31：スクリプト言語の講義では、分量は適切であったと思うが、講義の時間配分が悪いと感じた。後半にもう少しゆとりのある時間配分が望ましいと感じた。

改善・検討の余地あり 32 : 全体的に、コマンドのオプションを変えて「結果が変わる」ところを見るまでほとんど至らなかったことに少し物足りなさを覚えた。

ポジティブなコメント1: NGS解析をする上での最初のハードルは、解析環境の設定であると思う。その意味で、7/22, 23 の講義は、BioLinux の環境設定を中心に説明されており大変有意義であった。おかげで以降の講義が非常にスムーズに進めた。

ポジティブなコメント2: R の基礎、Bioconductor の利用法については配布資料、Web 資料が非常に良くまとめられており、講義中に理解できない部分があっても、改めて学習できる点が非常に素晴らしいと感じた。NGS 解析基礎、ゲノム Reseq、変異解析、RNA-seq についても、実地の解析に直ぐに応用できるような講義内容であった。またさまざまな有用なオープンソースソフトも紹介していただき、個人的には非常に有益であった。

ポジティブなコメント3: 極めて懇切丁寧で、とても充実した内容でした。積極的に質問する参加者の方もいらっしゃったのですが、そのお陰で理解が深まりました。参加者で、ある程度経験のある方や、率直に質問する方は本当に大切だと思いました。門田先生の R の講義は、圧巻でした。ここ掘れワンワンで、手順通りやった後で結果の解釈ができるようにと説明いただき大変分りやすかったです。もっともっと、講義を聞きたいと思いました。

ポジティブなコメント4: 全ゲノムシーケンスによる変異解析と RNA-seq は今後ほぼ確実に行うので、その解析手法の講義は絶対に役に立つものであり、実際に試してみても実感したが、解析のコマンドは時間がかかるものもあるため、シェルスクリプト等の利用は解析にほぼ必須であると感じたので、スクリプト言語の講義は受けて良かったと思っている。終盤の実際の解析は難解な部分も増えたと感じたが、スクリプト言語の講義は全体を通して分かりやすかった。

ポジティブなコメント5: Bio-Linux は使用したことがなかったが、非常に丁寧な資料のおかげで簡単に設定することができた。perl や python などの言語の基礎的な事を扱ってくれたので、ソフトウェアを使う以外にも応用できると思った。できれば R 言語も言語としての基礎的なところを扱って欲しかった。(もちろん今回扱った実践的な内容は今後役にたつと感じている)。実際の解析フローは実際に扱っている人の話だったので、普段している解析で気になっている事や気にしていなかったところ等が聞く事ができ、勉強になった。

ポジティブなコメント6: 全体的に分量は適切だったと思います。NGS 解析基礎は、もう少し分量が多くても良かったと思います。

ポジティブなコメント7: 全体的に興味のある内容でした。

ポジティブなコメント8: 特に門田先生の R に関する授業は、本当によく練り上げられて、すばら

しい限りです。すでに役立っていますし、今後もかなり役に立つと思います。

**ポジティブなコメント 9**：これから NGS 解析に取り組もうとしている身にとっては、NGS 解析の流れをつかむことができ非常に有難い講義内容でした。講義分量・レベルは、あれが限界でした。実データを使った取り組みは、非常に不安が残ります…。

**ポジティブなコメント 10**：非常に詳しい講義（特に門田先生の講義）で、不明点もなく、非常にわかりやすかった。

**ポジティブなコメント 11**：私が参加した講義はどれも大変有用でした。諸先生方に改めて感謝いたします。

**ポジティブなコメント 12**：150803Reseq 変異解析（山口先生）、150805NGS 解析基礎（山口先生）、Try マークをうまく利用され、重要なところにフォーカスが当てられていて、初心者への配慮があり、助かりました。

**ポジティブなコメント 13**：門田先生は、お疲れのところ、会場の電源コードの配線、服部先生、山口先生へのサポートなど、とてもご配慮いただき、ありがとうございました。

**ポジティブなコメント 14**：Bio-Linux8 と R のインストール状況の確認は、その後の講義を受講する上では大変役立つイントロでした。準備が大変だったとは思いますが、大変感謝しています。Linux 基礎もどのように、講義スライドがリッチに作成してあり、独学でもかなり理解できそうだとおもいました。スクリプト言語（シェルスクリプト）はもう少し難しい課題（現場で直面する問題）も含めていただいても良かったと感じました。Bioconductor の利用法も、コピペでどんどん進める点は良いと思いました。ただ、スクリプトの中身を完全に理解できてはいないので、バグがあればお手上げです。今後のさらなる自己研鑽が必要だと思いました。RNA-seq と ChIP-seq は、後半のスピードが速く、消化不良になりました。これから実際の課題を解いて行く中で独学を進めたいと思います。独学を進める上でのきっかけとしては有り難い講義でした。

**ポジティブなコメント 15**：日本乳酸菌学会誌の予習も含めて事前準備について具体的な指示があったのはとてもよかったです。

- ・Bio-Linux8 と R のインストール状況の確認：確認の機会があり、とてもよかったです。
- ・Linux 基礎：ウェブからの DL、パスを通す、インストールのための情報をどう読めばよいか、など丁寧な解説はわかりやすかったです。
- ・スクリプト言語（シェルスクリプト）：門田先生との連動もよく受講者が既に何を知っているかを前提にした講義がとてもよかったです。
- ・スクリプト言語（Perl）：\$\_（ドルアンダーバー）の説明がとてもよかったです。普段から R などを利用しているので for 文や if 文などはわかっていたのですが、\$\_のような Perl 独特な用法が理解

しづらく、Perl 本を読んでも体系的に書かれておらず気づくことさえできなかったのも、勉強になりました。

- ・ゲノム Reseq、変異解析：マッピングのところも含めて実演できたのはとてもよかったです。
- ・ChIP-seq：森岡先生が苦心されていることはとてもよくわかったのですが、少し講義の流れがわかりにくかったです。

**ポジティブなコメント 16**：RNAseq の内容に興味がありました。わかりやすく、これからの NGS 解析の勉強になりました。

**ポジティブなコメント 17**：気持ち的にはもっと受講したいが、受講する側も講義をして下さる先生方、運営スタッフの方々の労力を考えればこれ以上の長期間は難しいと思うので適切と思う。レベルはどの講義も初心者を対象としていただき難しすぎることは無かった。講義内容はどれも興味があった。Chip-seq は自分ではやらないかもしれないが興味をもって聞けた。この講義で学んだことを役立てるのは自分の努力量にかかっている。必ず役立てたいと思う。

**ポジティブなコメント 18**：Linux と R を用いた解析の方法について、初心者からでも分かりやすかったと思う。

**ポジティブなコメント 19**：事前に予習項目や講義資料が提示されていたので予習することが出来たので、解析初心者の私でもついていける内容であったのは良かったと思います。

**ポジティブなコメント 20**：全くの初心者で参加させていただいたが、非常に丁寧な説明(スライドも含め)でわかりやすい講義だった。特に門田先生のスライドは後からそのまま読み返して復習しやすいです。また、アメリカの先生の例題もあとで使うことができるものが多いのではないかと考えています。もう少し解析に使える事例も加えていただけたら、ありがたいと思いましたがこれは欲張りすぎ意見だと自覚して申し上げます。Biolinix を共通のものでインストールさせていただいたのは、とても良いと思いました。個人で動かしてトラブった場合、初心者はどう対応していいかわからないので・・・。

**ポジティブなコメント 21**：今、最先端の NGS のデータ解析の方法について、全体的に興味をかき立てられる内容であったと思います。

**ポジティブなコメント 22**：講義資料はいずれも丁寧で、講師の方々が準備にかなり時間を割いていただいたことが伺える。後で復習に使えるそう (全般)

**ポジティブなコメント 23**：RNA-Seq。fastx\_clipper など、公開されているオープンソースのソフトの中にも善し悪しがあるということは頭ではわかっていましたが、いざ目の当たりにしてみると衝撃でした。

**ポジティブなコメント 24:** 次世代シーケンサの解析など大規模なデータ解析において、市販のソフトウェアやプログラムをそのまま何も知らずに出てきた結果だけを見るだけというのは、自分が本当にやりたいことができているという確証が持てない。よって今回の Linux の基礎やスクリプト言語、R、NGS 解析基礎などをとりあえず、まんべんなく学べたのが今後の解析や NGS の理解をすることに良かったと思う。次に繋げる足がかりとなったことがよかったと思う。

**ポジティブなコメント 25:** アメリエフの山口先生の講義では、NGS 解析だけではなく、unix 環境で作業する上で注意する基本的なことを教えてくれたのがよかった (ディレクトリ名は数字始まりを避ける、ファイルをコピーしたらサイズを確認して壊れていないか確認など)。またゲノム変異解析の講義で、山口さんが velvet を初めて使う様子を見て、解析のプロの人が実際に作業するとき

に気にする情報や気にしない情報の感覚を知ることができたのがよかった。

**ポジティブなコメント 26:** RNA-seq の講義において、fPKM 値と DEG による 2 群間比較の使い分けが特に今後役に立つと感じた。

**ポジティブなコメント 27:** 現在トランスクリプトーム解析を主に行っており、当面はこれを行っていく予定であるため、その上で非常に勉強になった。Reseq や ChIP-seq は現在のところ行う予定がなく、自身の基礎知識が少ないこともあり一部理解に苦労したが、今後解析を行うことがあれば、参考になると思われる。

**ポジティブなコメント 28:** 自分も含め、日常的に linux を扱っていない講習生にとっても、とても分かり易いレベルの講義内容を準備して頂きました。毎日 18 時過ぎまでの講義を覚悟していたので、26 日「NGS 解析基礎」および 27 日「ゲノム Reseq、変異解析」の両日の講義が早く終わったのは、嬉しい反面、少し物足りないと感じました。

**ポジティブなコメント 29:** 「ChIP-seq」では実際の研究にそった形での講義だったので、非常に良かった。

**ポジティブなコメント 30:** 講義の Linux の基礎は、これまで Linux を使ったことがなかったので、事前学習をしっかりとっていたので、講義はよくわかり分量、内容ともに適切でした。また、講義では十分な時間をとっていたので、わからなかったり、コマンドを打って動かないときなどは、TAの方が親切に指導してくれたので、安心して講義を受けることができました。スクリプト言語も時間をかけてひとつひとつ丁寧に教えていただいたので、まだ自分で自由に使いこなせるということではないですが、これから勉強するよいきっかけになりました。R も事前学習をしましたので、なんとなくついていけました。この講義で学んだ内容はとても一人で始めから勉強するにはハードルが高いものばかりだと思いましたが、基礎的な知識を身につけられたので、これから自分で勉強するきっかけになりました。全体的にはこれから NGS を取り組もうという人にはとても役に立つ講

義だったと思います。

**ポジティブなコメント 31** : 8/26 NGS の基礎は、実習としては既に知っている内容でしたが、他の人に説明するときはどう伝えれば良いか参考になりました。8/27 変異解析は、時間と機会があれば取り組んでみたい内容でした。8/28 RNA-seq は、受講の主な目的でした。特にカウントデータ取得以降の統計解析はごく短時間でしたが、参加して良かったと思う内容でした。

**ポジティブなコメント 32** : 自分で得たサンプルで RNA シークエンスをやったことはないが、公開データを使って学んだことを実践してみようと思った

**ポジティブなコメント 33** : もともと NCBI の SRA などの NGS データを扱ったことはある程度で、基礎から学びたいと思っていました。自分で理解していないところを学ぶことができたりと、意義のある授業だったと思います。

**ポジティブなコメント 34** : 受講者の解析経験がバラバラであったので、初心者にとっては少し難しく、経験者にとっては少し物足りないと感じたかもしれないが、全体的を通して適度な難易度だったと思う。私自身は、少し物足りないと感じたところもあったが、未経験だった Chip-Seq などは少し難しく感じたところもあり、経験したことのある範囲についても、受講者の方々による質問や先生方の補足的説明によってとても今後の役に立つ知識は多かった。

**ポジティブなコメント 35** : NGS や linux について初めてさわったので、どの講義に対しても十分理解できたとは思えませんが、今、復習をしていて、少しずつ理解できるところが増えてきています。シェルスクリプト、パール、及びパイソンの講義に関しては内容の分量が少なく体力的精神的に余裕がもてました。NGS や linux をはじめてさわるのにはこのコース全体はとてよいきっかけになります。一人で勉強するとつまずいたところからなかなか抜け出せなくてくじけることが多いのですが、つまずいているところを助けてくれる方々がまわりにいると続けて取り組みます。

**ポジティブなコメント 36** : 8/26,27,28 の受講です。1 ヶ月ほど NGS を使った実験の準備をしているのですがデータのクオリティーチェック(NGS 解析基礎)、クオリティーコントロール(ゲノム Reseq、変異解析,RNA-seq)、フィルタリング(RNA-seq) に特に不安を感じていました。今回の講義でこれらに対応することが出来るようになったと思います。また発現変動についてもそれが有意か否かの判別に不安が有りましたが今回の統計解析についての講義(RNA-seq)で手がかりを得たと感じています。

**ポジティブなコメント 37** : NGS 解析基礎、ゲノム Reseq、変異解析ともに、予習をしておいたので講義内容は簡単に理解できた。

**ポジティブなコメント 38** : NGS 解析基礎 : NGS 解析によく使用されるデータファイルについて、

それぞれのファイルの特徴を説明して頂き、勉強になりました。

**ポジティブなコメント 39:** トランスクリプトーム解析の中でデータの整形方法から複数状態のサンプルへの応用例まで教えていただけたことで実際の解析研究に非常に役立つ知識を得ることができました。

**ポジティブなコメント 40:** 私自身がほぼ無知の素人でしたので、全体を通して丁寧すぎる内容は分かりやすく大変助かりました。まずは実際のデータを触って、しっかり学んだことを身につけていきたいと思います。

**ポジティブなコメント 41:** B 日程の 3 講義を受講しました。私はバイオインフォマティクス初学者ですが、丁寧な説明のおかげで十分理解しながら授業を聞くことが出来ました。下記の設問 18 とも関連しますが、もう少し分量が増えてスピードが上がっても付いていけると感じました。ただ、トラブル対応の時間的余裕を確保するという意味では現状で最適なのかもしれません。

**ポジティブなコメント 42:** Linux と R を同時に、習得するのは難しかったが、バイオインフォマティクスに関する講義があることは、とても有難い機械でした。メタゲノムに興味があったので、実際に自分が行っていることとは異なったが、基礎的な共通の部分は、参考になりました。

**ポジティブなコメント 43:** 現在、データ解析を行っているが、データの処理の仕方について適切な内容であり、今後役立つと感じた。

**ポジティブなコメント 44:** どの先生もとても丁寧に解説いただきました

**ポジティブなコメント 45:** amelieff さんや門田先生の講義は非常にわかりやすかった。今回の講義は初めの一歩としてはとても良かった。今後研究していく中で徐々に解析スキルを伸ばしていこうと思います。

**ポジティブなコメント 46:** NGS 解析の中身について詳しく理解でき、また R に関する講義も今度の解析にとっても有用だった。

**ポジティブなコメント 47:** RNA-seq 解析の講義では、生データを取得するところから教えていただけたので、実際に解析を行うときにも迷子にならずに済むだろうという安心感があった。大規模データの解析に馴染みがなく、初めて NGS 解析をするレベルの受講者にとって、このレベルから話を始めるのが適切だと思う。

**ポジティブなコメント 48:** どの講義も事前に資料がウェブ上にあげてあることで予習を行い、効果的に講義を受講することが出来ました。門田先生の乳酸菌の連載及びそのウェブ資料には大変助け

られました。

**ポジティブなコメント 49** : どの講義回も、一通りの流れを掴むには十分な講義ボリュームだった。R、Linux の扱い方が雰囲気ですり分かった程度で途中から参加したものの、講義についていけたので個人的にはレベルはマッチしていたと感じた。RNA-Seq も ChIP-Seq もおこなっているの、解析の中身を知ることができ、今後自前でも解析をやってみようと思えた。

**ポジティブなコメント 50** : 非常に用意された講習で、大変助かりました。十分な予習をしたたので (Web にスライド掲載して頂いていたので) 事前に質問したいことまで準備できて、非常に有意義でした

**ポジティブなコメント 51** : linux や R については必須であった予習がを行っていたのでわかりやすく理解も深まりました。

**ポジティブなコメント 52** : トランスクリプトーム解析について特に勉強になりました。

**ポジティブなコメント 53** : 8月26日 : NGS 解析基礎を受講させていただきました。自分の準備不足もあり TA の方には大変お世話になりました。講義に加えてですが、自分で試行錯誤をしながらやってみないと、バイオインフォマティクスに慣れることはないかと反省しました。誠にありがとうございました。

**ポジティブなコメント 54** : 自宅の WINDOWS 環境でゲノム解析ができることがわかったので、手始めにゲノムの小さいウイルスまたは細菌を対象に比較ゲノム解析で習得したものの実践しスキルの定着を図りたい。

**ポジティブなコメント 55** : Chip-Seq の講義での原理の説明など、Wet の研究とも結びつく実践的で興味深いものでした。

**ポジティブなコメント 56** : In my research I basically use licenced software to do genomic analysis derived from whole genome shotgun sequencing. It was very useful to learn about the many free software alternatives to do routine analysis like quality check of reads and de novo assembly. I also found very interesting the use of different programming languages to write scripts capable of extracting the relevant information from files containing sequences or information regarding their quality and annotation.

## 付録 2

設問内容：分かり易かったところ・分かりにくかったところについて、講義項目や内容など具体的にご記入ください（アンケートの設問 21 に相当）。

計 82 件の生の記載内容を以下に示す。主催側の主観で、45 件の改善・検討の余地ありのコメントと、37 件のポジティブなコメントに分けた。前者のほうが重要であるため、特に今後の検討課題と思われる部分をハイライトさせた。

改善・検討の余地あり 1：8/3, 8/4（山口さんのところ）の説明がやや駆け足気味で、講義資料にな  
いところもあったりして、少しついていくのに労を要した。

改善・検討の余地あり 2：全体を通して、丁寧に講義していただき非常にわかりやすかった。R や  
Bioconductor は個人的な予習が不足していたために、講義に追いつくのが大変であったが、充分予  
習しておけば、もっと有意義に学習できたと思われた。講習前に、最低限身に付ける必要のある知  
識を持ち合わせる事が重要であることを痛感した。

改善・検討の余地あり 3：ChIP-seq 解析ではマッピングまでの話は前日までに十分やっていると思  
いますので、クレンジング・ピークコール・アノテーション付けの話をもっと詳しく聞きたかった  
です。

改善・検討の余地あり 4：特にどの部分が"分かり易かった・分かりにくかった"は感じませんでした。

改善・検討の余地あり 5：NGS 解析基礎、Reseq、RNAseq 共通。重要な講師コメントをメモしな  
がら、コマンドを打つのが大変でした。配布資料にある程度記載していただけると大変助かります。

改善・検討の余地あり 6：R での解析は、具体例を使った解析で、わかりやすかった。統計解析に  
ついてなど、もう少し時間をとって説明してもらいたかった。TA は親切で、いろいろと教えてく  
ださり、必要であった。欲を言えば、mac 担当、win 担当それぞれが数人いてくださるといいか  
と思った。

改善・検討の余地あり 7：森岡先生の話は興味深かったのですが、最後まで行き着かなかったのがや  
や残念でした。

改善・検討の余地あり 8：ChIP-seq の講義について、途中から受講生がおいていかれて、説明のみ  
が進んでいくような感じを受けました。

改善・検討の余地あり 9：パイプラインとして全体の処理の流れが書いてある講義は分かり易かつ

た。一方資料中のコマンドを順に打っても、処理が続行できないケースがあった。ちゃんとパイプラインとして機能するよう資料を作してほしい。RNA-seq の授業では前半の講義終わりで cuffmerge、cuffdiff などの説明が流され、よく理解できなかった。非常に重要な部分であると思うのでちゃんと説明してほしい。現在、手元のデータの処理がこの部分でとまっている。また後半の授業では前半に作ったファイルがそのままつかえるほうがよい。

改善・検討の余地あり 10 :

- ・ スクリプト言語 (Python) : ライブラリの説明をもっと詳しくしてほしいです。代表的なライブラリは何か、それはどこでどのようにしらべればよいのか、など。
- ・ NGS 解析基礎 : igv の説明では自分のもっているゲノム情報や gff ファイルの使い方も詳しく解説してほしいです。
- ・ RNA-seq : tophat-cufflinks-cuffmerge-cuffdiff-cummeRbund といった、cufflinks ベースのマッピングから DEG 解析も解説があるとよかったです。

改善・検討の余地あり 11 : 課題や作業の前に、どのようなデータをどのような形に変換するのか、を明示していただくとより分かり易くなると思いました(例. 7月 27 日の Perl 入門の最終課題)。

改善・検討の余地あり 12 : 8月 3?7日はとてもわかりやすく、満足しました。8日は講義内容はわかりやすかったのですか、パソコンでの実習ではついていけなくなりました。

改善・検討の余地あり 13 : シェルスクリプトや Perl に比べ、Python の講義はわかりやすかったです。シェルスクリプトや Perl の講義は、少し各論に入りすぎたのではないかと思います。シェルスクリプトや Perl も Python の講義程度の分量だと、もう少しわかりやすかったのではないかと思います。

改善・検討の余地あり 14 : 門田先生の配布資料がわかりやすく助かりました。

改善・検討の余地あり 15 : スクリプト言語は自身の勉強不足もありますが、考え方に慣れるまで時間がかかりました。問題の回答コマンド一つ一つに、何をやっているのか解説が記載してあると理解が早かったかと思えます。

改善・検討の余地あり 16 : プログラミング言語の講義は前半の内容はついて行けたのですが、最後の課題の難易度が急に上がったのでついていけませんでした。perl の最終課題などはいきなり省略形だけではなく、full で書いたスクリプトも載せていただくとよかったですかなあと思いました。

改善・検討の余地あり 17 : わかりにくかったところ : 途中でだいぶ慣れたが、2画面のスクリーンでどちらを見るべきかわかりにくいときがあった (全般)

改善・検討の余地あり 18: まずコメントを説明してからコピペするほうがいいんじゃないですか。

改善・検討の余地あり 19: 講義のために字を大きく表示しなければならないということはわかるのですが、コマンドを打ったあとの結果でログがウィンドウ上から流れてしまって、講師の先生がどういうコマンドを打ったのかがわかりにくかった。

改善・検討の余地あり 20: Bio Linux において: 形式ファイルが複数出てくるため、現在の処理が何を行なっているか分からなくなりやすかった。

改善・検討の余地あり 21: 前半の講義で扱った各種プログラミング言語に関して、後半の講義で活用しきれていなかったようにも感じる。同様な機能の各種プログラム、アルゴリズム (例えばマッピングアルゴリズム) について、それぞれの特徴、長所短所等についてももう少し詳しく聞きたかった。

改善・検討の余地あり 22: RNAseq の講義において、クオリティコントロール (アダプター配列の除去) に関して、いくつかのツールを紹介して頂いた。これらのソフトを同条件で run した際に、得られる結果がどの程度異なるのか、実際どの部分の配列をカットしているのか等、スライド上でも結構ですのでソフトごとに紹介して頂けるとスッキリしたと思います。

改善・検討の余地あり 23: 講義資料を探すのが大変だった。東京大学からのリンクだけでなく、NBDC からも資料へのリンクを作成してほしい。

改善・検討の余地あり 24: 基本的には前半部分の Linux やプログラミング言語については、自分でもよく理解できたと思っています。後半の NGS 解析については、これまでほとんど知識がなく、解析によって様々な形式のファイルが出てきて、いったいそれぞれのファイルが何を意味しているのか、ぼんやりと聞いていると分からなくなってしまうようになりましたが、復習をしてある程度分かるようになりました。NGS 解析については、講義であった RNAseq や Chip-seq などもう一度聞いてみたいと思いました。

改善・検討の余地あり 25: アメリエフさんの講義の中で、「通常はこちらのソフトを使っていますが、今回は違うソフトでやります」という場面がいくつかあったが、ぜひ実際に使用されているソフトを使っていたら良かった。フリーソフトであれば、予習による事前準備、インストールは可能と思われる。

改善・検討の余地あり 26: 予習前提というスタンスはよいと思いましたが、山口先生の講義の資料は内容が少なく予習すべき項目がよくわからず実際の講義も説明が丁寧ではなかった為わかりにくかった。

改善・検討の余地あり 27 : Biolinux に入っていないソフトウェアで、有用なもののインストールの仕方などを別資料で構わないのももらえると思えた。

改善・検討の余地あり 28 : IGV を起動し可視化までは出来るのですが、どの項目や印が何を示しているのかや可視化することでどういった情報が得られるのか具体的にわかりませんでした。 IGV でここにデリベーションがありますね。といわれても確かにあるけどだから何なのか、どうやってデリベーションがあるのか探し出したのかわかりませんでした。これは基礎的な質問で、考えたり調べたりすればわかることだと思うのですが講義中は疑問点でいっぱいでした。

改善・検討の余地あり 29 : ゲノム Reseq、変異解析、RNAseq

・講義始めに解析全体の流れがまとめられており、わかりやすかった。一般的な解析の流れについて、参考サイトの情報も併せて説明があり、良かったです。同じような機能を有するソフトが複数あり、それぞれの長所や短所など口頭で説明があったが、あらかじめ表など一覧にした資料があるとわかりやすくて良いと思います。

・RNAseq は扱うデータによりカウントデータ取得後の統計解析の方法が変わると思うので、代表的な解析例をいくつか挙げて紹介して頂けるともっとわかりやすいと思いました。

改善・検討の余地あり 30 : どの講義でもツールのインストールからしっかり教えていただけたことと、オプションの意味や用例の説明がわかりやすかったです。Reseq の中でツールの動作原理に関する説明は難易度が高く、講義の中で理解することはできませんでした (bwa, cufflinks などマッピングツールの違いについて)。

改善・検討の余地あり 31 : 全体的に、講義資料の PDF が細かい説明のためにいくつもページが分かれていたのが逆にわかりづらいつと感じた。

改善・検討の余地あり 32 : Linux 基礎、NGS 解析基礎、R の基礎など、門田先生が担当されていたところは、実用的な考え、(情報処理のベースがなくても) で説明されていたので、データ処理するうえで、必要な部分がわかりやすかった。一方、スクリプト、perl、python については、最初の簡単な説明は、ついていけたものの、練習問題が、急に高度になり、理解できていないと思う。実用性の前に、もっと、基礎的なことを、反復練習をしたほうが、自分としては、理解できたと思う。

改善・検討の余地あり 33 : 部分受講が可能であり、途中から受講すると (予習をしても) ついていけない部分があった。

改善・検討の余地あり 34 : 事前の予習シリーズが大変分りやすく、有難いと感じました。門田先生がここまでの準備をなさっていただくのにかかるご苦勞は、並大抵ではないのではないかと思います。一つ困ったのは、予習の中で Biolinux やゲノムのダウンロードについてでした。仕事先ではなく家で繋いでいたため、特にゲノムダウンロードでの段階で詰まってしまう予習が最後まで出

来ませんでした。

**改善・検討の余地あり 35：**ChIP-seq の講義は無駄話が多いわ円滑に進まないわわかりにくいわ途中で終わるわでとても残念でした。ざっくりとでいいんで ChIP-seq のデータの解析を一通り通して欲しかったです。結局マッピングまでで終わってしまってオイオイと思いました。他の講義についてはとてもわかりやすかったと思います。

**改善・検討の余地あり 36：**Perl や Python の思想や文法を混同するのは連日の講義なので仕方ないことだが、最後の演習問題のレベルが飛躍的に難しくなったことで、少し混乱してしまった。もちろん、このレベルのことができなければ有意義な解析ができないことは理解できるので、このこと自体に不満はない。ただ、時間配分の工夫の余地があれば、段階をもう少し踏んでいくと、より分かりやすくなるのかなとは思う。質問で、コードを一行一行さらっていくことどちらが効率的なのかは分かりませんが。

**改善・検討の余地あり 37：**全体を通して分かりにくい点というのは特になく、簡潔に説明されていたように思う。先にも述べたように、実際の生データを使ったアウトプットの項目を増やして欲しいと思った。

**改善・検討の余地あり 38：**「8月6日：ChIP-seq」の講義は、少し脱線気味になったからか他の講義回に比べて駆け足で終わったため、もう少し理解を深めながら進めたら良かった。また、後半の実践編では門田先生や山口先生のような流れに沿ってコマンド等が記してある講義資料にしていただけだと、持ち帰って復習する際にも助かります。(作っていただくのは大変手間だと思いますが。。。)

**改善・検討の余地あり 39：**講義ページ [http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html) はやはりわかりづらい。せめて1段階だけでもよいので階層化して欲しい。

**改善・検討の余地あり 40：**コマンドライン操作や、NGS データに触れたことのない状態で参加させていただきました。それでも十分に理解していける講義になっていたと思います。懇切丁寧なスライドを作成していただき、ありがとうございます。確かに、RNA-seq の統計学の部分は、とても理解しやすくもっと講義をしていただきたいと感じました。ぜひご検討ください。

**改善・検討の余地あり 41：**R や linux については必須であった予習をしっかりしていたのでわかりやすく感じました。スクリプト言語 (perl,python) は、途中までは簡単すぎましたが、最後に急に難しくなりついていけませんでした。難しいところをゆっくり説明いただければありがたいです。

**改善・検討の余地あり 42：**R 言語の豊富な例題を提供していただき、自分でゲノム解析する場合もそれらの例題をそのまま使えば大半は事足りると思いますが、R 言語のプログラミングの基本枠

組みのようなものも知りたかった。

**改善・検討の余地あり 43** : For me probably the most difficult part to understand was the several explanations given about the usage of commands in the different programming languages, especially the order in which they have to be written in order to make them work.

**改善・検討の余地あり 44** : 森岡勝樹先生の講義は期待はずれだった。実は私は、森岡先生が医科歯科大学の博士後期課程でいらして田中博先生のプログラムのTAされていたときにいろいろと教わりました。森岡先生が大変分かりやすく教えてくださることを経験していたので、今回の講義は事前の打ち合わせ不足で分かりにくくなっていたことがわかるだけにもったいない講義でした。

**改善・検討の余地あり 45** : 暑い所講習会してくださって本当にありがとうございました。自分の知識レベルの問題で毎講義の後ろの部分難しい部分を分かりにくいでした。やはり予習する必要があると思われました。具体的に記入するの私によって今の時点で難しいですが、勉強が進みます。申し上げないです。

**ポジティブなコメント 1** : 全体を通じて、説明は非常にわかりやすく、とくに 8/4, 8/5 については、コマンドもコピー&ペーストすれば使えるように用意されており、助かった。TA さんのサポートも素晴らしかった。事前予習の資料が非常によく出来ており、大変助かった。7月22日 : Bio-Linux8 と R のインストール状況の確認と 7月23日 : Linux 基礎は、大変役立つ内容で、このままで問題ないように思います。

**ポジティブなコメント 2** : TAの方がとても親切で、大変にありがたかったです。聞きやすい雰囲気、とても親切に接して頂いて感謝感謝です。TAの先生方によろしくお伝え頂ければ幸いです。本当に有難うございました！

**ポジティブなコメント 3** : スクリプト言語の講義は全体的に分かりやすかった。Rでの解析はなかなか中身の理解まで頭が追いつかなかったが、実際に自分でいじらないと理解できないと感じた。

**ポジティブなコメント 4** : Linux 基礎は Linux を十分使えるレベルまで教えてもらえたので非常にためになった。Shell や perl、python は簡単なレベルであったが、一通りの事をやってもらえたので今まで書けなかったことが書けるようになった。R はスライドがわかりやすかったので、今後も必要ときに参照にしたい。

**ポジティブなコメント 5** : 門田先生の R に関する授業は、資料が本当によく練り上げられていて、かつ時間配分、休憩の取り方等考慮されているため、非常に分かり易かったです。ゲノム Reseq. 変異解析、RNA-seq 前半、ChIP-seq は、もう少し講義資料を作り込んでいただけると、メモを取ることが軽減されるため、理解が深まると思います。これこそが、ハンズオン講習会に参加する意義

だと思います。

**ポジティブなコメント 6** : TAの方が困ったら全部解決してくれたのでとても良かったです

**ポジティブなコメント 7** : プログラミングは全くの初心者ですが、服部先生の3講演は、特に非常に分かりやすく有意義でした。先生方がやってもうまくいかないことはたくさんある、ということ全体を通じて把握できたことから、専門外だからと簡単に挫折せず努力していけるのではないかと思います。

**ポジティブなコメント 8** : 判らない点があったときすぐに TA が駆けつけて大変助かった。

**ポジティブなコメント 9** : 昨年度も受講しました。時間も余裕が有るように設定されていた点もあり2回目でしたが山口先生の講義は大変有意義でした。門田先生の講義は個人的には興味深かったですが、実用性は初心者には欠けるようには思いました。ただ門田先生が用意された Biolinux の設定などの資料は極めて有用な資料でした。分かり易い工夫として、講義のメモを門田先生が web に載せて頂いている記載や、講義中に適宜質問をして頂いた内容はいつも参考になりました。

**ポジティブなコメント 10** : TAの方からサポートをいただける点は大変有り難かったです。このような仕組みは大変すばらしいと思いました。

**ポジティブなコメント 11** : 講義は、入門者にとってはとっかかりとして十分な機能を果たしていると思います。Webで見直すこともできるので非常に助かります。

**ポジティブなコメント 12** : 門田先生担当分の講義資料 (7月23日、29日、30日、8月5日) はスライドのキャプチャがほぼ全て載せてあり、理解しやすかったです。

**ポジティブなコメント 13** : 門田先生の講義は全て資料が完璧なのでわかりにくいことはなかった。服部先生も同様です。ハンズオンがたくさんあってよかった。山口先生の講義はハンズオンが少し少なめなのと少し専門用語がわからず逡巡することがあった。森岡先生の講義ではDLで時間がかかったりしたが「アドリブ」でやるとどうなるかがわかりトラブルの対処法の勉強になった。

**ポジティブなコメント 14** : スクリプト言語については、もう少し色々やれたらいいのかとも思いましたが初心者の私としては分かりやすかったです。

**ポジティブなコメント 15** : 自分の予習不足を棚上げしている意見で恐縮ですが、いささかスピードについていくのが大変な項目もありました(山口先生の初日)。以降は細かい補足説明はわかりやすかったです。些細な入力ミスに本人は気がつかないで焦るばかりのときに、なんども TAの方に助けていただきました。問題形式で実施していただいた服部先生の講義は自分の理解を図るうえでと

でも良かったです。できれば回答の入ったスライドを講習会終了後に門田先生の HP に Up してほしいです。

**ポジティブなコメント 16:** お話を聴き、資料を読み、実習を行うことで、全体的に理解できる内容で、わかりにくかったことはほとんどありません。

**ポジティブなコメント 17:** わかり易かったところ: これまできちんとイメージできていなかった配列とかハッシュなどの説明がわかり易かった (プログラミング)。Python は初めてだったが、Perl と両方やっていただいたことで理解し易かった (プログラミング)。予習を義務づけられたことで、自身の理解は深まった、講義中のトラブルも比較的少なかった (門田先生)

**ポジティブなコメント 18:** 全体的に基礎の基礎をやっていただけたので、初心者にはわかり易く、とてもありがたかったです。

**ポジティブなコメント 19:** この講習会に参加するために NGS 解析について少し勉強した程度の知識しかありませんでしたが、講習会の資料がとても丁寧でわかりやすかったです。また、予習不足で TA さんに大変お世話になったのですが、丁寧に対応していただきとても助かりました。

**ポジティブなコメント 20:** 特にありません。わかりやすかったです。

**ポジティブなコメント 21:** アメリエフの山口さんが担当された講義に関しては、内容が難しくなかったこともあり大変分りやすかった。

**ポジティブなコメント 22:** Linux 基礎から、またスクリプト言語についても全くの素人でもなんとかついていけることができたのはよかったです。

**ポジティブなコメント 23:** 自分にとっては、全体的にわかりやすく、理解もしやすかったと思う。

**ポジティブなコメント 24:** 基本的な Linux コマンドから説明するなど、受講者全体のペースに合わせて講義が進行したことで、Linux に慣れていない人でもついていける内容だったと思います。

**ポジティブなコメント 25:** NGS 解析で、ソフトの使い方がわからないとき、それをどのようにして理解していくかを、講師の方が実地に示していただけたので、とても参考になった。

**ポジティブなコメント 26:** アメリエフ社の配布資料はそれにそってハンズオンで実地で訓練するのにシンプルでよかったです。その他の配布資料は詳しく説明があり、講義後に自分で復習するのによかったです。自分が linux 等になれていないのでハンズオンの実地で訓練するには読むのにもついていけないところがありました。TA はハンズオンの訓練には必要でとてもいいシステムだと

思います。講義についていけないところが沢山ありましたが何度も助けていただきました。

**ポジティブなコメント 27:** RNA-seq の講義で 使用ツールの種類と順番について説明して頂いた部分が特に解りやすかったです。ネットなどで大まかに把握していたつもりでしたが適当に動かせば何らかのデータが出てしまうので不安でした。QC やフィルタリングの部分でネットや自習では不足していた部分が得られたと思います。

**ポジティブなコメント 28:** 山口さんの蘊蓄がマニアックでよかった

**ポジティブなコメント 29:** 全体を通して、分かりやすい丁寧な講義資料を公開して頂いていたので予習復習に大変役立ちましたし、今後も活用したいと思います。講師の門田先生が「資料を見て自習すれば良いのでは？」と仰っておりましたが、ドライの解析のずぶの素人としては、解析の実際の雰囲気を知るのに、その現場のプロの人から直接教をを請うという体験は貴重だと思っております。

**ポジティブなコメント 30:** 配布資料が非常に分かりやすく、流れに多少遅れても資料を読んですぐに追いつけました。講義資料を載せたページも見やすく、先生のコメントまで掲載して下さっていたので、復習も捗りました。また、TA の方がいて下さったのは非常に助かりました。助けていただき、ありがとうございました。

**ポジティブなコメント 31:** TA の方におしえていただき、講義についていくことができました。

**ポジティブなコメント 32:** それぞれにスクリプトが用意され、とてもついていきやすかった。

**ポジティブなコメント 33:** Chip-seq は扱ったことがなく予習不足で内容を理解するので精一杯だったが、内容はよかったと思います。R の Bioconductor の講義は次回からは必要ないと思いました。(門田先生の過去の講義資料でも十分自習可能なため)

**ポジティブなコメント 34:** 私はエンドユーザーですので、計算の中身の詳細は兎も角、パイプラインの使い方を理解出来ました。実際にはつまるところあると思いますが、その解決する方法も示して頂いていたので、自分でも進めそうです。

**ポジティブなコメント 35:** 資料に従ってコマンドを動かせば大体なんとかなるのがよかったです

**ポジティブなコメント 36:** 事前に資料が配られていたので、予習もできて大変助かりました。

**ポジティブなコメント 37:** 実は講義をリアルタイムに理解するのは困難でした。教材がきめ細やかでしたので、しっかり理解できました。

## 付録 3

**設問内容：**感想、要望、意見、あるいは改善のための提案 1 として、「講義内容に取り入れて欲しいものについてできるだけ具体的にご記入下さい」。

計 60 件の生の記載内容を以下に示す。今後の検討課題と思われる部分は、ハイライトさせた。

**コメント 1：**本講習では、各ステップを確認しながらすすめていくような形式でした。今後は、各受講日に宿題という形で課題を与え、実際に自分で学んだステップをひと通り使って解析し、翌日に解説をしてもらえると、より理解が深まり、自信がつくようになると思いました。small RNA 解析についても、情報をいただけるとよかったです。また、使うシーケンサーによって解析に注意が必要な点等を教えていただけるとよかったですかな。

**コメント 2：**主にエクソーム解析による遺伝性疾患の原因遺伝子変異の検出を行っており、候補遺伝子の絞り方などの骨などを教えていただきたい。

**コメント 3：**参加者の目的や利用したい内容などがわかれば、大変有り難いな、と思いました。個人的に仲良くなった人はおりましたが、お互いに知り合う事で、情報交換もできるかな、と思いましたが、懇親会、なども有れば参加したいと思いましたが、盛り沢山かつ、分かり良い内容でしたので、贅沢なお願いかもしれませんね。

**コメント 4：**出芽酵母やヒト以外（データベースが充実していない生物）の遺伝子のアノテーション付加についてももう少し詳しく取り扱ってほしい

**コメント 5：**統計学。RNA-seq、ChIP-seq のデータからどういった検定を使ってデータを出すのか知りたかった。

**コメント 6：**2 点の講義内容の取り入れを希望します。

1. 計算環境構築について。具体的には、どの解析でどのくらいの計算機のスペックが必要であるとか、ワークステーション等の計算環境構築の方法（ソフトのインストール方法、順番など）。
2. パイプラインの作成と実行。多くのラボで同じような解析を何度も実施すると思いますが、その場合に、シェルスクリプト、Perl などを利用した解析の半自動化の方法について。

**コメント 7：**講義中に紹介された Galaxy を講義後に独学で扱ってみたところ、今回の講義に参加するような人たちにとってはそちらを使ったほうが解析に役立つのではないかと思いました。受講者全員が Galaxy にアクセスするとサーバを圧迫すると思われるので、実習には不向きかと思いますが、基本的な使い方と容量の大きいファイルのアップロード方法やワークフローの作製といった躓きやすそうな部分だけでも教えていただければ、すぐに実際に使うことができそうだと感じました。

**コメント 8 :** アンプリコンシーケンスに関するQCの具体的な考え方や手法について。テーマの範囲をある程度絞ったうえで受講者が課題を持ち寄り、全員で実施し解決案を模索していくという講

**コメント 9 :** miRNA のターゲット解析について取り入れて欲しいです。具体的には、RNA-seq.によって得られた mRNA 情報データと、miRNA-seq.によって得られた miRNA 情報データを組み合わせること手法等について講義内容に取り入れて頂けたらなと思っています。

**コメント 10 :** Perl, Python の分量をもう少し増やしていただけると助かります。

**コメント 11 :** R での解析でもっと統計などについても解析をしてもらいたかった。また、論文でよくみるようなヒートマップやその他図の作成法・その時犯すやすい間違えなども説明があるとよかったと思う。全体を通して、結局 perl, python はいつ使うのかよく分からないこともあった。(自分では使っているので違和感なかったが・・・)

**コメント 12 :** 遺伝研などの協力が必要かと思いますが、スーパーコンピュータにログインして、UGE にジョブを送って検討する手順 (遺伝研以外でも大体同様かと思いますが)、pipeline 化された解析法の利用について、などがあると実際的かと思いますが。Linux の有る程度のスペックの機器が直ぐに準備できない場合はこの形が必要になるかと思いますがので有用に思います。

**コメント 13 :** bisulfite を用いたエピゲノム解析の講義を取り入れてください。

**コメント 14 :** 欲を言えば、受講者が持ち込む課題 (実際の研究データ解析) について講師と受講生と一緒に問題解決するような講義があれば素晴らしいと思います。もしかすると、このような形式は「講義」ではなく、「共同研究」かもしれませんが。

**コメント 15 :** 何らかのプログラムを使って処理する際に、

1. 「必要なもの (用意しておくファイルなど)」
2. 「この処理を行う意味 (何がどうなるのかの説明)」
3. 「出力されるもの (出力ファイル)」

この3つを必ず説明していただきたいと思う (口頭だと時間を食うので資料に印刷してあるとよいと思います)。よくわからないままコマンドを打つということが多かったように思う。

**コメント 16 :** トランスクリプトーム de novo アセンブリングをぜひ取り入れてほしいです。ゲノムのわかっていない非モデル動物での RNA-seq をしたいので、そのために必要なパイプラインを理解するために必要な講義をぜひ開講してほしいです。

**コメント 17 :** 突発的ではあったが vi の講習は非常に有用であった。vi に限らなくてもよいがテキ

ストエディタの使い方の講習があると効率的なスクリプト入力あるいはテキスト編集に役立つと考えられる

**コメント 18:** 8月4日か5日に予定していた講義に加えてアドリブで山口先生がマッピングの異なる方法について説明してくださっていましたが、その内容をはじめから組み込んでおいていただけると幸いです。

**コメント 19:** RNA解析のクラスタリングのところでは少しやっていますが、RやCircosなどを使って、得られた結果を図示化させる講義がもっとあればいいような気がします。

**コメント 20:** 26年度にあった分子生命科学のスライドが楽しかったので、講義も聞いてみたかったと思いました。

**コメント 21:** 難しいとは思いますが、仮想環境ではない実際の解析環境の構築について座学でよいので教えてほしいです。今そこで困っております。独学あるいはお金で解決しかないとは思いますが。

**コメント 22:** NGS解析について、演習を多めにしてもらえたらさらにいいのかと思いました。

**コメント 23:** 講義につきましては、大変満足しております。丁寧に作ってくださった資料など準備のことも含めて先生方のご苦勞に大変感謝申し上げます。今後ぜひ加えていただけるのであれば、メタゲノム解析を入れていただきたいです。

**コメント 24:** 課題の答となるスクリプトが見えてしまう講義資料になっているので、最終ページにまとめるなどしてほしい（服部先生）。データサイズを大きくするのは実行速度や処理能力の関係上難しいかもしれないが、もう少し実際の解析に近い過程が体験できるようなサンプルデータセットを使用してほしい（もう少し多めのリード数を扱うとか）（山口先生）。RNA-Seqに関して、リードカウント取得後の正規化手法について詳しく取り上げてほしい（門田先生）

**コメント 25:** ゲノム Reseq、変異解析と RNA-seq はこの解析を目指して受講する人も多いと思うので、それぞれ2日ずつにしてもっと深くやってもらいたい。特にゲノムの変異解析は実際に複数のシーケンズデータから mapping して変異を較べられるくらいの内容にして欲しい。

**コメント 26:** 実際の解析に必要なコンピュータのスペックについて、講義の序盤でお話し頂けるとより良いと思いました（終わりに近くなってから、手持ちの環境では到底、解析は不可能と知ってショックを受けてしまいました）。

**コメント 27:** 世代シーケンズ解析に使えるような、統計についての講義をもう少し受けてみたいで

す。

**コメント 28 :** 次世代シーケンサで得られたリードのマッピングまでのパイプラインを学べたが、それぞれのソフトを通して得られた結果の確からしさを判定するような実践的な判定法を知りたい。

**コメント 29 :** 他の解析ソフトとの比較をより詳細に知りたかった。

**コメント 30 :** 非モデル生物の NGS 解析についても、少しでもいいので触れてもらえるとありがたい。

**コメント 31 :** データ量が大きくなると、自前のパソコンでは解析が難しいと思われます。その場合に遺伝研のスパコンや AWS などをつかった解析も内容に入れて頂けると嬉しいです。

**コメント 32 :** 今回は 1 個体に由来するリードのマッピングを行い、多型の検出を行ったが、複数の個体に由来するリードを一度にマッピングし、複数個体間の SNP の検出を行ってみたい。

**コメント 33 :** NGS 解析で利用できる、R のパッケージ、コマンドについてももっとたくさん紹介していただけると良いと思います。

**コメント 34 :** 実習の課題を増やして頂きたいです。変異解析の実際や RNA-seq から融合遺伝子をどのように見つけてくるのか。という実践的な内容の実習（宿題でよいと思います）があるとよいと思いました。

**コメント 35 :** B 日程にも Chip-seq の講義を入れて欲しかった。

**コメント 36 :** 例えば、前講義として、サンプルデータのダウンロード、Biolonux に入っていないが、使うべきソフトのインストールを解説するのはどうでしょう？

**コメント 37 :** RNAseq が発現解析までだったので BLAST や KEGG などのもっと下流の解析も取り入れて欲しいです。

**コメント 38 :** Python や Perl などの講義は受けていませんが、実際にそれらを使って、NGS 解析をどのようにして行うかなど紹介してもらえると良いと思いました

**コメント 39 :** 時間的にも量的にも難しいとは思いますが、発表された論文を題材に、そこでのデータ解析を学べるといいかもしれない。

**コメント 40 :** スクリプト言語に関しては、そのスクリプトに触れたことのない人を対象にしている

ため、少し難しいと思うが、実際の解析でどのように用いるか（たとえば、pipeline の繰り返し処理やとあるツールを用いるために必要なファイル加工など）などの内容があれば、より実践的になるのではないかと思う。

**コメント 41 :** 個別ツールの出力の使い方についてはまだ良くわからないところがあるので、その解説を増やしていただくと勉強になります。統計解析についての計算内容についても同様です。ただし自分は地方からの参加で、これ以上の日程は取りづらいのでこれらについては自習とし講義は現在の内容で適切です。

**コメント 42 :** GATK を利用したリアライメントについて、もう少し詳しく聞きたかった

**コメント 43 :** 各シーケンサのデータの特徴、解析する際の注意事項など

**コメント 44 :** スーパーコンピュータ（遺伝研）、AWS などリモートアクセス端末の運用方法。

**コメント 45 :** 今回はデータの一部を用いて解析の手法を知るということでしたが、実際の研究の具体例のようなものも見たいように思います。論文を読んでそれに従えば良いとは思いますが、やはり現場の研究者が、何を考えて、どう工夫して解析を始めて、終わらせるのか伺ってみたいです。

**コメント 46 :** シーケンサから出力されるデータを処理した後の統計解析について、特に規格化や発現変動解析時のアルゴリズムの選び方、使用するパッケージの選び方を先生方のご経験を聞きながら学びたい。

**コメント 47 :** オプションを変えて計算結果を変えてみるなどの応用部分の充実。

**コメント 48 :** 現在、NGS の解析のなかでも、QIIME に関する、要望や問い合わせが非常に多い。要望を受けて、イルミナが独自に、セミナーを実施してくれていたが、1 日未満の時間しかなく、時間的に不足していた。

**コメント 49 :** R での解析方法を多々紹介頂けると良かったかもしれません。

**コメント 50 :** de novo アセンブリ（特に PacBio を使ったバクテリアゲノムについて）、アノテーションの詳しいやり方などです。

**コメント 51 :** pathway 解析など

**コメント 52 :** 非モデル生物関連について。非モデル生物を実際に扱っている方からどのような点で苦労されているのかなどをお聞きしたいです。

コメント 53 : NGS で利用する統計処理の本質、扱いに関する講義

コメント 54 : シーケンシングの手法は年々進化していますが、そこもフォローしきれていません。NGS 解析がメインで、このボリュームなので難しいと思いますが、解析手前の最新の動向、新しい手法について軽く聞けると良いかと思いました。

コメント 55 : データ解析後の kegg、go の解析

コメント 56 : 本講習会の範囲外になってしまうと思うが、NGS のケミストリーや出てくるデータについての最低限の基礎知識についての講座はあった方が良いでしょうに思えた。

コメント 57 : ChIP, RNA-seq など講習会では実際に使用している解析ソフトと違うものを使用していたり、一部が省略されていたので、明日から自分で解析を行うために”実際に”使用する解析ソフトを用いた一連の解析を学べる機会があればありがたいです (→講習会後に佐藤さまよりご回答・アドバイスいただきました。さらに

- ・ Peak call 後の一連の代表的な解析 (motif 解析、Heat map 解析など)
- ・ (コマンドラインを使用しない) galaxy を用いた NGS 解析など
- ・ 4C-seq などさらにアドバンストな解析

コメント 58 : 通常のパソコンで実際の NGS データを解析するのは難しいということだったが、クラウド等を使って実際のデータを手持ちのパソコンで解析できるスキルを学びたい

コメント 59: トーゴデータベースにアップロードされたスクリーン映像はコマンドラインが見えにくい。森岡勝樹先生に聞いて、東京医科歯科大学の田中博先生のところで、オミックス情報学の振興分野人材養成プログラムの講義の録画で使用していたのと同じシステムを使うようにすれば見にくいということはなくなるはず。東大の内部にもどこかが所有していて借りられるのではないのでしょうか。コンピュータのデスクトップ画面と話し手の両方が動画内に配置されて大変見やすい講義映像になります。今回の NGS ハンズオンはよちよちレベルに手とり足とりおしえるがコンセプトと分かっていたら森岡先生はターゲットを絞って分かりやすく話してくれたと思います。今回の森岡先生は受講者のレベル範囲をひろく捉えすぎていて上級者への気配りにいらぬ労力をかけすぎてしまっていました。今回の失敗を受講者が森岡先生の講義力と受け取ってしまわないことを願うのですが、森岡先生にしてはかなり失敗でした。森岡先生のレクチャーは、いいし深いし、味があるということ。今回受講した皆さんがあらためて知ることになるようなチャンスが今後あるといいと思います。がんばれ森岡先生。

コメント 60 : I think it would be a great idea to have a comparative genomics module within the 講習会.

## 付録 4

**設問内容：**感想、要望、意見、あるいは改善のための提案 2 として、「本講習会を受講後、次のステップとして必要または身につけたいと思う解析スキルについて、できるだけ具体的にご記入下さい」。計 48 件の記載内容を以下に示す。今後の検討課題と思われる部分は、ハイライトさせた。

**コメント 1：**まずは今回の講習の復習をして、自分で公共データを落としてきて、解析を試みようと思います。遺伝子発現・ネットワーク解析に関心があるので、RNA-Seq でデータを算出後の解析を身につけたいと思います。具体的には、cytoscape や WGCNA といったものが利用できるようなことです。

**コメント 2：**VCF ファイルのハンドリングなどに必要な、実際の、具体的なシェルスクリプトや perl のスキル

**コメント 3：**メチル化の解析法など。また、講習内容を理解することができたならば、次に自学自習できる領域などがわかれば大変有り難いです。HP を見ればわかりますでしょうか？

**コメント 4：**NGS の結果から作れるデータを増やす。NGS 関連のソフトウェア開発

**コメント 5：**解析スキルに関しては、ゲノム Reseq.変異解析、RNA-seq、ChIP-seq 全体に言えることですが、解析の流れのさわりではなく、今後は実際に解析する場合の一連の操作や考え方を身につける必要があると考えています。

**コメント 6：**メチル化 seq の解析。CNV の解析 (exome でも array でも) が個人的に興味があるので、できればよろしく願いいたします

**コメント 7：**python や perl でのより実践的なプログラミングスキル。

**コメント 8：**プログラムを作成できるように、少なくとも perl 言語をある程度習得できるよう勉強を始めました。

**コメント 9：**miRNA のターゲット解析

**コメント 10：**NGS データベース、Galaxy などについて知りたいです。AJACS の講習会などで対応できるとお聞きしました。補完していきたいと思います。

**コメント 11：**自分でおこなったマイクロアレイ、RNA-seq、ChIP-seq のデータがあるので、それについてまずは解析を試みようと思う。次のステップでは、ぜひ、まだアノテーションされていな

い遺伝子（例えば eRNA など）をヒストンのメチル化を指標に座標を抽出する、プロモータ上流 2 kb を全遺伝子についてとってくる、それにマッピングするなどをしてみたい。

**コメント 12:** 新規解析スキルではありませんが、遺伝研のスーパーコンピュータに自分のデータを置いて受託の解析結果を自身で検討してみようを予定しています。講義のお陰でこの 1 週間で作業が進みました。作業は同じとしても mammalian のデータで実際に動かしてデータ取得する流れをくみたいと思っています。人がいないので wet で得た結果を自分で有る程度解析して、必要に応じてバイオインフォマティクンに相談する流れを作りたいと思っています。

**コメント 13:** メタゲノム解析の基礎について

**コメント 14:** R を使った各種統計解析が学びたいと思います。特に、論文を作成するために必要な統計解析について学びたいと思います。

**コメント 15:** Galaxy による処理を学習したいと思っています。

**コメント 16:** 異なる種間で遺伝子のオーソログやパラログなどの解析をしたい。この作業はアノテーションとも関連すると思うので、アノテーションの仕方（提案 3 参照）とも連動して行いたい。RNA-seq は 1 種での遺伝子発現変動を調べるだけでなく、複数種での発現変動の類似点と相違点を調べたい。そのためには、種間でどの遺伝子がオーソログなのか、重複遺伝子でどの遺伝子どうしがパラログなのかを知る必要がある。また、これらを調べることで、どの種で新規に重複した遺伝子なのかなども知ることができる。

**コメント 17:** RNA-seq のより細かい解析、今回は 2 群間がメインだったと思うが、例えば 3 群以上の解析や時系列解析などを身につけていきたい。

**コメント 18:** 解析後のデータをよりグラフィカルに示す方法を身につけたい。

**コメント 19:** パスウェイ解析

**コメント 20:** 統計的な解析スキルを身につけたいです。

**コメント 21:** 個人としては、変異解析がメインなので、変異の絞り込みと、CNV,SV などの解析です。

**コメント 22:** linux が使えるようになったので、タンパク解析にも使っていきたいと思っています。

**コメント 23:** 実際にワークステーションに Linux を入れて、非モデル生物（マダイ、カンパチ）を

つかった RNA-seq をやりたいです。非モデル生物（マダイ、カンパチ）のゲノムデータの構築、整備もやりたい。

コメント 24 : 解析結果の比較などを行う方法

コメント 25 : CLIP など、RNA-seq の応用編の解析の仕方を学びたいです

コメント 26 : Biolinux によるスクリプトあるいは R をもう少し使えるようになりたいと思っています。

コメント 27 : ChIP-Seq が都合で受講できなかったので、これを学びたい。

コメント 28 : 今回 Best practice を初めて知ったので、参考にしてパイプラインの各ステップのより深い理解と、それを実現するための環境構築

コメント 29 : 変異解析の具体的手法。IGV で眺めるだけで無く、意味のありそうな変異箇所をみつけてくるまでのステップを身に付けたい。また統計の基礎の基礎が出来ていないので徐々に学習していきたい。

コメント 30 : 適切なプログラムやパッケージを選定し、自身でトラブルシューティングを行いつつ使用するスキル。同じ目的に対し複数のパッケージがリリースされているので、自分の目的に合わせてどれが最適なものか判断できることが必要だと感じました。

コメント 31 : 解析ソフトウェアの細かな設定について自分で変更できるくらいのスキルを身につけたいです。

コメント 32 : 主成分分析などの多変量解析や、機械学習を用いたビックデータからのデータマイニングを身につけたい。

コメント 33 : マッピングの後に見たいデータを抜き出して解析する方法。

コメント 34 : 解析を行うにあたり、Linux や R についての勉強不足を感じた。

コメント 35 : 非モデル生物での NGS 解析（de novo アセンブリ、トランスクリプトーム解析）

コメント 36 : perl や python など、基本を教えて頂いたのですが、もう少しスキルアップしたいと思います。

コメント 37 : 連続処理の方法を身につけたいです。

コメント 38 : 統計解析

コメント 39 : 実際の論文等で使用されている解析手法や、講習会で使用されなかったその他の解析ソフトの使い方

コメント 40 : 本講習会で得た知識をもとに、自身の研究で得た生データから解析可能なデータに変換するまでの一連の流れを実際に自分自身で行ってみることが次に行うべき大切なステップであると考えている。そこで試行錯誤することによって初めて自身のなかから有意義な質問や考えが浮かんでくるように思う。さらに、アウトプットを繰り返すことによって本講習会で得たものが自身のスキルとして着実に身についていくように思う。

コメント 41 : Publicdata を用いたインシリコンスクリーニング。特にパスウェイ解析を中心に。

コメント 42 : ある程度複雑なスクリプトを組めるようになりたい。

コメント 43 : さらなる R 操作習得

コメント 44 : 提案 1 と同じになりますが、まずは実践あるのとのことで、自分の ChIP-seq データを解析してみたいと思います。

- ・ Peak call 後の一連の代表的な解析 (motif 解析、Heat map 解析など)
- ・ (コマンドラインをあまり使用しない) galaxy を用いた NGS 解析など
- ・ 4C-seq などさらにアドバンストな解析の講習会も開催されればぜひ参加させていただきたいです。

コメント 45 : 通常のパソコンで実際の NGS データを解析するのは難しいということだったが、クラウド等を使って実際のデータを手持ちのパソコンで解析できるスキルを学びたい

コメント 46 : PERL の正規表現は配列解析に便利な道具となると思いますが、とても自分で考えて応用できる自信がありません。身につける事例豊富な資料がありましたらご教示ください。

コメント 47 : Basically following the same idea of the previous paragraph, I would like to learn more about tools available to do comparative genomics.

コメント 48 : 転写因子のネットワーク解析についてです

## 付録 5

**設問内容：**感想、要望、意見、あるいは改善のための提案 3 として、「NGS 解析以外で、開催して欲しいと思う講習会がございましたらご記入下さい」。計 38 件の記載内容を以下に示す。

**コメント 1：**CRISPR 関連の講習会があれば参加したいです

**コメント 2：**統計の講習。論文で頻用される検定などの解説。計算方法などではなく、どのような解析にはどのような検定が適当であるなどについて。

**コメント 3：**より進んだスキルや解析方法など、実際に論文で出されているバイオインフォマティクスの利用・応用の仕方の好例などをサーベイできるような講義があると大変有り難いと思いました。

**コメント 4：**電子顕微鏡の取り扱い、試料作成

**コメント 5：**統計の講習会

**コメント 6：**メチル化アレイ・SNP アレイの解析、GWAS など

**コメント 7：**MGS にこだわらず、3 次解析に必要な多変量解析について

**コメント 8：**統計解析の基礎、論文で行う統計のやりがちなミスなどがわかる講義が受けたい。

**コメント 9：**統計解析の講習会をお願いします。

**コメント 10：**10 人程度の少人数制で、データ持込みによる実際の解析の手ほどきがあったら非常に助かります。

**コメント 11：**RNA-seq をシアセンブルが終わった後の、アノテーションの仕方についての講習会を開いてほしい。どのような基準で、どのような手法で、どのような手順でアノテーションを進めるのかを知りたい。

**コメント 12：**Vim の講習会があればとても興味がある

**コメント 13：**Image J を使った画像解析。 Cytoscape の使い方講習

**コメント 14：**パスウェイ解析の基礎についての講習会

コメント 15 : 統計解析に関する講習会

コメント 16 : Linux での解析方法

コメント 17 : スパコンの講習会

コメント 18 : 統計解析

コメント 19 : システム生物学

コメント 20 : 自然科学実験に必要な基礎的な統計学について R も使いながら学べる講習会がある  
といい。

コメント 21 : NGS (シクンシングをかける段階) の講習会がほしいです。

コメント 22 : NGS などの大規模解析に役立つような統計学。

コメント 23 : オミクスデータ解析に必要となる統計学の基礎 (主成分分析、回帰など) や、周波数  
応答解析に必要な知識 (微分方程式) など、生物系研究者が必修で学んでこなかった数学的知識を  
扱う講習があるとうれしいです。

コメント 24 : 統計学や機械学習の講習会

コメント 25 : データベースの構築と管理、研究成果をウェブで公開するためのテクニックのレク  
チャーなど情報系の講習があると良いと思います。

コメント 26 : マイクロアレイ解析 (遺伝子、miRNA)

コメント 27 : 今さら感はありますが、マイクロアレイでのデータ解析と様々ある統計手法の違いの  
解説講義

コメント 28 : 解析ではなく、NGS の wet のほうで講習会があると良いなと思います。

コメント 29 : 大量情報データからの統計解析

コメント 30 : プログラミング (Python など) に特化した講習会を希望します。

コメント 31 : ・機能性 RNA 解析、・構造解析、・プロテオーム解析

コメント 32 : 広く浅くよりも、NGS 関連で、メタゲノムはじめ、必要なことは、まだまだありますので、NGS ニーズも広がりつつあるので、NGS 中心に、項目を広く、詳細な内容で開催してほしい。

コメント 33 : 遺伝研スパコン関連。もう少し細かいチュートリアルてきな講習会（またはウェブ資料）を開催して頂きたいです。

コメント 34 : 同講義を続けて欲しいです。

コメント 35 : 生物統計学（とくに遺伝統計学）の実習講義

コメント 36 : 統計学に関する講義

コメント 37 : NGS 解析になりますが

- ・ Peak call 後の一連の代表的な解析（motif 解析、Heat map 解析など）の講習会
- ・（コマンドラインをあまり使用しない） galaxy の講習会
- ・ 4C-seq などさらにアドバンストな解析の講習会も開催されれば

NGS 解析以外では、

- ・ 臨床研究を行う上での統計的解析法の講習会

コメント 38 : Something related to genome evolution.

## 付録 6

**設問内容：**感想、要望、意見、あるいは改善のための提案 4 として、「設備、講義室、受講人数、TA、その他、講義全体に関する要望・コメントなどがございましたらご記入下さい。」。計 35 件の記載内容を以下に示す。

**コメント 1：**設備、講義室、TA、講義全体について、大変すばらしく、要望はありません。この受講人数で、よく開催できたなど逆に大変感心しております。講師の先生方やスタッフの方々の準備や労力を思うと、これ以上のものは望めないように思います。こういったセミナーでは、参加者が事前準備の課題へしっかり取り組んでくれることが最も大事だと感じました。

**コメント 2：**貸し出しパソコンもあり、素晴らしい、と思いました。唯一、スライドの下端が前の人で隠れてしまうので、もう少し上目に映写されていたら見やすかったと感じました。特に打ち込み中のコマンドが見えずに次のスライドに行くことが多かったので、

**コメント 3：**全体に渡って、非常に満足しています。運営されている方々に、感謝申し上げます。提案としては、ボランティアで参加されている TA さんに、感謝の気持ちを込めて、講義の最後に拍手を送ったりすると良いかもしれません。私自身、TA さんに何度も助けて頂きました。ありがとうございました。

**コメント 4：**満足しています

**コメント 5：**Bio-Linux に入っているツールを中心に使っていたので、講義内で使うものと講師の先生が普段解析で使うツールが異なる感じだったことが残念だった。先生方がいつも使っている解析ツールを使ってやってもらえたら、より実践的で講義後すぐに同じようにやってみれるのではと思った。

**コメント 6：**乳酸菌学会誌の連載を予習として勉強したが、これは講義を実のある者にするために非常に有用であった。部分講習が可能であったため、急な業務対応の際には退出することが出来、社会人としては本当に助かった。

**コメント 7：**設備等につきましては、不都合全くなかったです。

**コメント 8：**設備・スタッフ共、非常に充実しており助かりました。ありがとうございました。

**コメント 9：**非常に詳しい講義ありがとうございました。

**コメント 10：**大変判りやすく丁寧に教えて頂き、今後も続けて行って頂きたいと思います。

**コメント 11 :**

- 1) TAの方が不具合を解決出来ず、どんどん置いて行かれそうになりました。別のTAの方を呼んで対応してもらいました。TAも大変かと思いますが、自分で解決出来ないときの対応策も一応お伝えされると良いかと思いました。
- 2) 一般の方で全く素人風の年配の方が頻回にTAを呼んでおりTAの方が気の毒でした。専門性が高い講義故に、受講者は、経歴、所属、目的等で多少は絞った方がよいのではと思いました。
- 3) NGSについて8月上旬、下旬に2回の機会をつくって頂いたことは助かりました。

**コメント 12 :** 設備や教室の大きさ、受講人数、TAの数などはバランスが良かったと思う。一方、これはコントロールしにくい部分だとは思いますが、特定の人が何回も何回も繰り返し質問する傾向にあり、それがさもすると「皆の理解に役立つ、講義中にすべき質問」ではなく、「非常に自分の環境特異的な、特殊な質問」になる傾向があったように思う。その点、注意事項として運営側から注意喚起したほうが有意義に時間を過ごせると思う。

**コメント 13 :** 講義室は特に寒くも暑くもなくよい空調でした。TAの方はとても熱心にかつ丁寧に教えていただけてとてもよかったです。ありがとうございました。

**コメント 14 :** 自身がTAとしても参加したため感じたが、やはりTA人数が受講生に対して少ない+受講生が講義内容外でつまづく(コマンドライン操作やフォルダ共有)が多いため、自身で問題解決する自信の無い場合は持参した端末でなく、用意された端末を利用させたほうが進み方としてはスムーズかと思う(もしくは基本的に貸し出し端末で作業させ、余裕がある人だけ持参した端末で同時作業させるなど)。

**コメント 15 :**

- ・体調の悪い方が出てきた場合を除き、冷房の強さを統一していただけたとなお良いと思いました。
- ・私の見落としかもしれませんが、Macの人はトラブルが発生しやすいことを事前に伝えておいたほうがよいと思いました。
- ・その他につきましては非常に洗練された講習会であったと思います。門田先生をはじめとしたスタッフの皆様方の尽力に深く御礼申し上げます。

**コメント 16 :** とても快適な環境でした。お菓子が嬉しかったです。

**コメント 17:** スクリプトの基礎から実践的な解析手法まで幅広く学ぶことができる講習会はなかなかないと思います。しかもハンズオン。大変有意義な講習会でした。初心者向けとはありましたが、結構ハイレベルな受講生の方が多く、質問の意味自体よく分からないという状態でしたが、TAの方のフォローでなんとか乗り切れました。

**コメント 18 :** 少ない TA という制約がありながらも、非常に上手く運営がなされていたように感じます。設備含めた運営体制には満足しています、ありがとうございました。

**コメント 19 :** できれば人数少なめの方がよい。質問している人が限られていたと思う。少人数だと躊躇や遠慮が減るように思います。

**コメント 20 :** 今回の内容で十分と思います。

**コメント 21 :** 受講人数などは適切だと思いました。

**コメント 22 :** 見せたくないかもしれない失敗の事例なども教えていただき、ありがたく思います。初心者のため何度も TA の方に救っていただきました。あらかじめ USB 等で大きなファイルを配布していただいたのは良かった。一方で、みんなで Web ページを開いて見に行きながら講義を受けるのは、とても今風で良いと思います。無茶ぶりにも即座に対応してくださる講師陣もすごいと思いました。スクリーンが複数あって、片方にターミナル画面、片方に講義資料の表示など使い分けていただいたのもとても分かりやすく良かったです。

**コメント 23 :** 講義室、受講人数、TA、ちょうどよかったと思います。B 日程もあったので、都合を合わせることができ大変助かりました。講義のご準備大変であったと思いますが、NGS データ解析の一通りを学ぶことができとても有意義でした。後は、自分自身でもっているデータの解析で実践したいです。

**コメント 24 :** お菓子がおいであったのはうれしかった

**コメント 25 :** 特にないです。おおむね満足しています。

**コメント 26 :** これだけ充実した講義をインセンティブも無くやっただきありがとうございます。講師の先生方や運営に携わった方にお礼申し上げます。この講義を無駄にしないように成果につながるようにしたいと思います。

**コメント 27 :** 先生方は勿論ですが、TA の方々にも非常にお世話になりました。心より感謝申し上げます。

**コメント 28 :** 初心者でもついていけるようにサポートしていただきありがとうございました。

**コメント 29 :** 研究や仕事の合間を縫って準備してくださった講師・TA のみなさまに感謝しております。

**コメント 30:** 今回の講義では質問の時に TA が拡声器を持って質問者のところへ行っていたところを講義室の一番前の席にマイクを3本くらいおいて質問者にマイクのところへ行ってもらったほうが質問者の数など把握しやすかったのではないかと思った。

**コメント 31:** 先生や TA の方が丁寧でとても助かりました。

**コメント 32:** TA さんが迅速に対応していただき、非常に助かった。今後も TA さんのいる講義が望ましいと思いました。

**コメント 33:** やや空調が効きすぎだったと思います。

**コメント 34:** 門田先生がこれだけ、NGSハンズオンを醸成してくださったのだから、これからはこの講義手法をまるごと専任スタッフで行った方がよいです。つまり、門田先生を初めとした研究者は監修者となって補助的に関わり、実際の講義は専任スタッフが行うようにすべきです。それは正論だと誰もが思っていることだと思いますが、バイオインフォマティクスの職員を正職員として雇うように大学は雇用体制を改革すべきです。あるいは J S T なりがそうしたバイオインフォマティクスの職員を正職員で採用して正職員のチームを組ませて全国を巡回させるべきです。

**コメント 35:** While the computers given to us by the organizers were really fast, they had a lot of bloatware. Although I assume that it must be really difficult to get rid of it in all computers, it would be nice to only have the necessary background processes running.

## 付録 7

**設問内容：**感想、要望、意見、あるいは改善のための提案 5 として、「その他、ウェブで公開・提供して欲しいバイオインフォマティクス教材に関するご要望などがございましたらご記入下さい」。計 28 件の記載内容を以下に示す。

**コメント 1：**今回の内容に加えて、中級の内容の講義資料や動画が有れば、大変有り難いです。その、学ぶ順番などのフローチャートがあると大変有り難いです、また、先にも述べましたが、応用の好例など、どのような形で利用できるかという点などもシェア頂ければ幸いです（これこそ、自分たちで考えるべきことですが・・・）講師の先生方には、本当に有難うございました。全日程参加できず大変申し訳ありませんでしたが、動画と資料を使ってできるだけ受講するように頑張りたいと思います。

**コメント 2：**受講者の情報共有掲示板の作成。なかなかプロの方に質問しづらいので、初学者どうして知識の共有等をしていくとスキルの向上につながるかもしれないと思う。

**コメント 3：**各解析ソフトの更新情報が一元的に分かると助かると思います。研究者が個々に、その都度、解析ソフトの公開情報（ホームページ）を確認する必要がなくなるため、全体として研究者の時間節約にもつながるはずです。

**コメント 4：**NGS 解析に使うソフトに入力するコマンド例を提供して頂けると有用です（そのままコピーできる形で）。オプションが多数ありますが、この条件に設定すれば（データーの場所などの情報と共に）、確実に動くと言った代表的な例があればこれを修正しつつ、自分のニーズに合った形に適宜直していけるかと思えます。まだそうした書籍もいまは限られているかと思えます。

**コメント 5：**現在の「R で～」の資料は大変役に立っております。Biolinux についても同様な資料を公開いただけましたら非常に役立つと思えます。

**コメント 6：**クラウド保管ツールの使い方。解析データは年々サイズが大きくなるばかりでシェアするときにはクラウド保管ツールを使うことも増えてきているように感じる。私自身はデータを受け取ることが多いの今のところ困っていないが今後受け渡すとき、どのようなクラウド保管ツールがあり、それを使ってどうやってデータをシェアしていくか教材があると助かります。

**コメント 7：**今回のそれぞれの解析の講習では、解析全体を行ったのではなく、解析のワークフローの一部分を抜き出しているの、ウェブで公開していただけるのであれば、解析の手順を追っていけるような資料を公開していただきたいです。

**コメント 8：**様々なケースワークを知りたいです。「ゲノム未知の生物種解析方法」「腸内細菌叢」

など。

**コメント 9:** 現状でとりあえず満足しています。勉強しきれないほどたくさんあることを知りました。

**コメント 10:** メタゲノム解析

**コメント 11:** 特にありません。解析のためのコマンドが多く掲載されているので、それを学んでからだと思います。

**コメント 12:** NGS 関連解析ツールの英語マニュアルの日本語化テキスト（探せばあるのかもしれませんが）ファイル整形の事例とサンプル入力ファイルのファイル構造の解説、整形用のサンプルスクリプト、スクリプトの解説をひとまとめにしたもの

**コメント 13:** 先生、TA さんそして事務の方たちありがとうございました。外国人でも分かれる講習会で、本当にありがたいです。

**コメント 14:** oo 年度版新着教材や新着ウェブサイトがわかる教材（？）があれば嬉しいです。

**コメント 15:** バイオインフォマティクスで使われるフリーのソフトの具体的な使い方についてオプションなどの説明が日本語で表示されているものがあればありがたい。

**コメント 16:** 統合 TV のような、動画形式でのウェブ教材をより充実させてほしい。

**コメント 17:** 今回の講習で得られた知識をもとに、実際の利用場面を想定した、問題解決型の講習会、もしくは web セミナーがあればぜひ受講したいと感じます。

**コメント 18:** ウェブ上で、セミナーや講習の内容を公開していただけるのは、とても役に立っています。

**コメント 19:** 門田先生のようにステップバイステップで解析過程がわかるような資料をアメリエフの講義内容でもアップして頂きたいです。

**コメント 20:** ウェブが少々見づらいので、どこに何があるか全て把握できませんでした。仕方がないことはわかっていますが、もう少し見やすくしてもらえると助かります。

**コメント 21:** 様々な無償ツールの使い方、エラー対処法などを分かりやすくまとめたサイトがあったらいいと思う。（質問サイトなどはありますが…）

**コメント 22 :** SNP 解析について。GWAS の解析手法の教材があれば助かります。PLINK の使い方など。

**コメント 23 :** 門田先生の R 解説のように、詳細なブログがあると、勉強しやすいです。また、ここにわからないことを、訊けるサイト(独自の Q&A)があると、心強いです。内容が自分独自のものと、訊ける人を探すのが難しいのが現状です。

**コメント 24:** 当方はがん関連の仕事をしているので、領域別に資料等があるとよいかと思いました。

**コメント 25 :** 門田先生のホームページについて、この後も持続的に最新の研究のアップデートをしていただけると嬉しいです。また、このページの中における de novo アセンブリなどの情報について、さらに詳しく説明していただければ嬉しいです。そのような学術的な仕組み（門田先生だけではあまりに大変そうなので）はできないものなのでしょうか。

**コメント 26 :** 今のところ先に上げたスパコン関連以外は満足しております。とくに門田先生のサイトには大変お世話になっており、これからも更新し続けてほしいと思っております

**コメント 27 :** 統計学に関する門田先生なりの解釈を含めた資料

**コメント 28 :** トーゴデータベースの動画を、小分けにしてアップロードしてもらいたい。必ずしも職場や自宅の PC のスペックが高いわけではないので、ロードする時間がかかりすぎてみたくてもみられないことが続きました。10 分か 15 分ずつに小分けにしてアップロードしてもらえると助かります。今夏の講義についてもこれからでも小分けバージョンを追加していただけないでしょうか。