

国立研究開発法人科学技術振興機構

ライフサイエンスデータベース統合推進事業

## 平成28年度共同研究実施報告書

代表者 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、平成 25 年度に 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」に分かれている。平成 26 年度には「速習」コースが試行的に実施され、平成 27 年度には試行実施結果を踏まえた改善を行い、ハンズオンに特化したより効果的な NGS ハンズオン講習会が実施された。

本共同研究では、これまでの実施結果を踏まえ、NBDC と連携して統計解析および高度な NGS 解析の内容を多く取り入れた NGS ハンズオン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	H28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(7/19~22, 25~28, 8/1~4)	
2) 教材の整備 ・ポータルサイトの構築支援 ・ウェブページ「(Rで)塩基配列解析」 ( <a href="http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html">http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html</a> ) 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、これまでの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）
7月19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	主催・共催機関
第1部 統計解析					
7月20日	10:30-18:15	ゲノム解析、塩基配列解析	（中級）	実習	門田幸二（東京大学）
7月21日	10:30-18:15	トランスクリプトーム解析1	（中級）	実習	門田幸二（東京大学）
7月22日	10:30-18:15	トランスクリプトーム解析2	（中級）	実習	門田幸二（東京大学）
第2部 NGS解析（初～中級）					
7月25日	10:30-18:15	NGS解析基礎	初・中級	講義 実習	山口昌雄（Amelieff） 窪川美雪（同）
7月26日	10:30-18:15	ゲノムReseq, 変異解析	初・中級	講義 実習	山口昌雄（Amelieff） 三澤拓真（同）
7月27日	10:30-18:15	RNA-seq	初・中級	講義 実習	山口昌雄（Amelieff） 尾上広祐（同）
7月28日	10:30-18:15	ChIP-seq	初・中級	講義 実習	山口昌雄（Amelieff） 久保竜一（同）
第3部 NGS解析（中～上級）					
8月1日	10:30-18:15	Linux環境でのデータ解析： JavaやRの利用法	中・上級	実習	門田幸二（東京大学）
8月2日	10:30-18:15	Linux環境でのデータ解析：マッピング、トリミング、アセンブリ	中・上級	実習	門田幸二（東京大学）
8月3日	10:30-18:15	クラウド環境との連携、ロングリードデータの解析	中・上級	実習	門田幸二（東京大学）
8月4日	10:30-18:15	トランスクリプトームアセンブリ、発現量推定	中・上級	実習	門田幸二（東京大学）

表1. 講習会概要。

平成27年度講習会との主な違いは、下記の通りである：

- ① 3部構成として実施
- ② 第1部として、統計解析を大幅に増加（0.5 → 3日）して実施
- ③ 第2部のNGS解析（初～中級）を本講習会の中核として位置づけ、アメリエフ株式会社の講師陣のみで内容を適度にアップデートして実施
- ④ 第3部のNGS解析（中～上級）として、よりハイレベルな中～上級者向けの内容を新規実施

## 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_ngo_sokushu_2016">http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngo_sokushu_2016</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、受講申込受付業務全般、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/h28">http://biosciencedbc.jp/human/human-resources/workshop/h28</a>
共催機関1	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）、統合TV
共催機関2	明治薬科大学
担当業務	講義内容事前確認（特にMacintosh PC）、講義補助（TA）

表 2. 講習会実施体制。

実施場所は、平成 26, 27 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 158 名収容）とした。



図 1. 講習会会場（7/19 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

受講対象は学部生以上、受講定員は約 30 名、受講条件は平成 27 年度と概ね同じとした。具体的には、開催概要や受講対象の欄に下記内容を掲載し、受講希望者には事前準備および予習を課すことを明記した（予習内容については 4-5 を参照）。申込みは、主催機関 2 のウェブサイト (<http://biosciencedbc.jp/human/human-resources/workshop/h28>) 上で受け付けた：

- これまでの本講習会の枠組みでは実施できなかった内容を多く含める予定ですので、第 1 部および 3 部は既に一定のスキルをもつ受講生を対象としておりますが、新規に受講される方でも、平成 27 年度の講習会資料を中心に指定された予習を実施すれば、問題なく理解していただければと思います。
- 昨年度との違い、対応関係、想定受講者、および予習事項については「こちら」をご覧ください。  
<http://biosciencedbc.jp/gadget/human/h28-h27contents.pdf>
- 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します。
- 「昨年度（平成 27 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。

<http://biosciencedbc.jp/human/human-resources/workshop/h27>

また、以下の 6 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。
- 本プログラム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。
- 講師指定の事前予習を実施していただける方。

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 4 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, 27 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 4 日 14 時～5 月 31 日 12 時とし、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（4/10）、69 名（4/15）、81 名（4/20）、85 名（4/25）、94 名（4/30）、

96名(5/5)、104名(5/10)、111名(5/15)、117名(5/20)、128名(5/25)、164名(5/31)であった。このうち2名が締切日までに全日程キャンセルしたため、5/31の締切時点での受講希望者数は162名となった。

**表 3-1**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。「5/31の希望者数」は、5/31の申込み締切時点での結果である。最も希望者が多かったのは第1部初日(7/20)の143名であり、平均して129.5名の希望者数であった。「7/19の希望者数」は、7/19の講習会初日時点の結果である。この日までに5-10名の事前キャンセルの連絡があり、最多が133名、平均で121.4名の希望者数となった。部分受講を認めていることから、平成27年度と同様に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。「出席確認表の出席者数」は、その出席確認表の数を集計したものである。この数値が公式記録ということになるが、現実には出席確認表に○をつけない受講者が少なからず存在するため、実際の受講人数よりも少なくなるのは致し方ない。

「アンケート結果の数」は、4-7. アンケート結果の**表 8**(理解度に関する感想;後述)において、①そう思う～⑤そう思わないの合計数を示したものである(⑥受講していない、および⑦未選択を除く)。「出席確認表の出席者数(平均で78.4名)」と「アンケート結果の数(平均で82.1名)」が一致するのが理想ではあるが、7/19の自習日を除く7/20～8/4の期間内でも、最大で7名の違い(8/2)が生じていることがわかる。7/19の自習日を除いた理由は、理解できたとして出席しなかった受講者が一定数存在する可能性を考慮したためである。

実施日	項目	5/31の希望者数	7/19の希望者数	①出席確認表の出席者数	②アンケート結果の数	
7月19日	PC環境の構築(Bio-Linux 8とRのインストール状況確認)	127	118	83	95	
第1部	7月20日	ゲノム解析、塩基配列解析	143	133	98	97
	7月21日	トランスクリプトーム解析1	139	130	84	87
	7月22日	トランスクリプトーム解析2	139	130	92	93
第2部	7月25日	NGS解析基礎	138	129	93	95
	7月26日	ゲノムReseq、変異解析	132	123	91	93
	7月27日	RNA-seq	137	128	89	92
	7月28日	ChIP-seq	123	117	79	83
第3部	8月1日	Linux環境でのデータ解析:JavaやRの利用法	119	114	67	68
	8月2日	Linux環境でのデータ解析:マッピング、トリミング、アセンブリ	122	115	60	67
	8月3日	クラウド環境との連携、ロングリードデータの解析	119	112	53	58
	8月4日	トランスクリプトームアセンブリ、発現量推定	116	108	51	57

表 3-1. 受講希望者数の内訳、および出席確認表とアンケート結果から得られた受講者数。

**表 3-1** の右側 2 列の違いについて詳細に分析する。まず、「①出席確認表の出席者数」列の数値は、1 日でも出席したと自己申告した（出席確認表に○をつけた）計 126 名の受講者に関するものである。そしてアンケート対象者は、1 日でも出席したと自己申告した（出席確認表に○をつけた）この 126 名の受講者である。「②アンケート結果の数」は、対象者 126 名中 119 名の回答者の結果に基づいている。①の結果は 7 名のアンケート未回答者分を含んでいるため、119 名のアンケート回答者に限定して①を再集計したものが③である。②と③共通の出席者数が④であり、③の結果とほぼ同じであることがわかる。**赤字の箇所**は、③と④間での不一致を表す。6 箇所全てが出席確認表上は出席になっているが、アンケート上では欠席のものに相当する。同姓の他人が間違えて出席確認表に○をつけてしまったのかもしれないし、単に出席した項目の日付を間違えただけなのかもしれない。④には (①-③) で表されるアンケート未回答者の出席数を考慮していない数値であることから、**最少でも④の人数は出席していた**といえるだろう。

全体として、②は③のほぼ全てを含んでいる。また、③の出席表上は欠席となっているが、該当項目に対する「受講していない or 未選択」以外の選択肢が選ばれた出席を示唆する受講者の設問 19 などの自由記載欄を調べてみると、アンケート結果（つまり実際には出席）を支持する明確なコメントが散見された。例えば、「7 月 27 日（水）RNA-seq については、アノテーションに関する情報を聞きたかったこともあり役に立った」、「第 2 部の rna-seq は特にわかりやすかった。どのソフトをどの順番で何に使うかが図式化されていたため自分の中で整理がしやすかった」、「全日程参加した」などである。⑤は「②アンケート結果の数に (①-③) で表されるアンケート未回答者の出席者数」を足したものである。もちろん実際には出席しておらず、虚偽あるいは勘違いによる可能性も完全には否定できないためあくまでも参考値ではあるが、おそらく⑤が実際の出席者数の推定値として尤もらしく、実際の出席者の出席人数の感覚にも合致するのではないだろうか。

実施日	①出席確認表の出席者数 (126名)	②アンケート結果の数 (119名)	③出席確認表の出席者数 (アンケートに回答した119名に限定)	④「②と③」共通の出席者数 (積集合)	⑤推定出席者数 defined as ②+ (①-③)
7月19日	83	95	81	80	97
第1部	7月20日	98	97	94	93
	7月21日	84	87	82	80
	7月22日	92	93	86	85
第2部	7月25日	93	95	90	90
	7月26日	91	93	88	88
	7月27日	89	92	85	85
	7月28日	79	83	76	75
第3部	8月1日	67	68	62	62
	8月2日	60	67	57	57
	8月3日	53	58	49	48
	8月4日	51	57	48	48

表 3-2. 出席確認表とアンケート結果、受講希望者数の内訳および実際の受講者数。

今年度の締切時点での総受講希望者数（162名）は、部分受講を可能にしたことで前年比約1.5倍となった平成27年度（136名）よりさらに増加した。このうち、36名（22.2%）が結果的に全日程キャンセルした（一度も講習会に出席しなかった）。36名中22名からは事前連絡があり、残りの14名が事前連絡なしの不参加（平成26年度はゼロ、平成27年度は26名中13名）であった。

受講者の要望に応じて部分受講を認めているわけではあるが、開催側は様々な可能性を想定した準備を要するため、非常に負担が大きい。昨年度も事後の検討事項としてあがったが、今後の講習会における「連絡なし不参加者」対策として、何らかのペナルティを明記するなどの措置があったほうがいいかもしれない。

総受講希望者数162名の各種内訳は、表4-1の通りである。関東在住の受講者が約8割を占めた。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。昨年度に比べて、特に医学系の受講者が多かったように思われる。職種は、学生が84名、ポスドク・研究員が36名、助教・講師が16名であった。所属は、大学関係が119名と最も多く、東京大学が64名を占めていた。東京大学の内訳としては、農学系（獣医を含む）が27名、医学系（保健を含む）が19名、理学系（情報理工を含む）が7名、その他が11名であった。

住所	東京：93名、神奈川：18名、埼玉：9名、千葉：8名、茨城：4名、京都：4名、静岡：3名、栃木：3名、愛知：2名、宮城：2名、大阪：2名、兵庫：2名、岐阜、熊本、群馬、広島、三重、新潟、石川、長崎、長野、福岡、北海道、和歌山が各1名
研究分野（複数回答可）	生物学（総合生物学10名を含む）：98名、医歯薬学：66名、農学：52名、情報学：13名、理工学：3名、その他：10名
職種	学生（大学院生を含む）：83名、研究員：23名、講師・助教：16名、ポスドク：13名、主任研究員：4名、教授・准教授：4名、研究支援員：5名、その他：14名
所属	大学：119名（うち、東大64名）、企業：25名、研究所：16名、その他：2名
Linux利用経験	全く触ったことが無い：52名、使ったことはあるがよく覚えていない：29名、簡単なコマンドなら知っている：44名、時々or頻繁に使っている：35名、未選択：2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：38名、簡単なコマンドなら知っている：31名、時々or頻繁に使っている：52名、未選択：2名
NGS利用経験	今のところ使用予定はない：3名、使ったことはないが使用予定がある：45名、いずれ使用すると思う：28名、時々or頻繁に使っている：65名、未選択：3名
実習PC環境	Mac持込：61名、Win持込：50名、アグリバイオPC貸与希望：48名、未選択：3名
H26年度講習会受講の有無	受講していない：153名、受講した：5名、未選択：4名
H27年度講習会受講の有無	受講していない：141名、受講した：18名、未選択：3名

本講習会を知った経緯（複数回答可）	友人・知人：60名、アグリバイオインフォマティクス（Rで）塩基配列解析を含む）：53名、NBDCウェブサイト：22名、メーリングリスト：21名、キーワード検索の結果：16名、DBCLSウェブサイト：8名、SNS（Facebook, Twitter等）：2名、分子生物学会：1名、その他：18名
-------------------	--

表 4-1. 受講希望者 162 名の各種内訳。

総受講希望者数 162 名のうち、出席確認表上で実際に一日でも出席した受講者数 126 名の内訳を表 4-2 に示す。1 都 4 県（神奈川・埼玉・千葉・茨城）の減少幅（132→101 名；76.5%）のほうが、それ以外の減少幅（30→25 名；83.3%）と比べて若干大きい。その他の特記事項としては、実習 PC 環境のところで、アグリバイオ PC 貸与希望者数の減少幅が小さいことが挙げられる。一つの可能性としては、自身の PC 環境構築の自習段階で挫折した受講者が一定の割合で存在することを示唆しているのかもしれない。

住所	東京：93→69名、神奈川：18→14名、埼玉：9→7名、千葉：8→7名、茨城：4→4名、京都：4→2名、静岡：3→2名、栃木：3→3名、愛知：2→2名、宮城：2→2名、大阪：2→2名、兵庫：2→2名、岐阜1→1名、熊本1→1名、群馬1→0名、広島1→1名、三重1→1名、新潟1→0名、石川1→1名、長崎1→1名、長野1→1名、福岡1→1名、北海道1→1名、和歌山1→1名
研究分野（複数回答可）	生物学（総合生物学8名を含む）：98→71名、医歯薬学：66→53名、農学：52→41名、情報学：13→10名、理工学：3→2名、その他：10→7名
職種	学生（大学院生を含む）：83→63名、研究員：23→20名、講師・助教：16→12名、ポスドク：13→12名、主任研究員：4→3名、教授・准教授：4→3名、研究支援員：5→4名、その他：14→9名
所属	大学：119→89名（うち、東大64→48名）、企業：25→20名、研究所：16→15名、その他：2→2名
Linux利用経験	全く触ったことが無い：52→39名、使ったことはあるがよく覚えていない：29→25名、簡単なコマンドなら知っている：44→34名、時々or頻繁に使っている：35→27名、未選択：2→1名
R利用経験	全く触ったことが無い：39→31名、使ったことはあるがよく覚えていない：38→30名、簡単なコマンドなら知っている：31→24名、時々or頻繁に使っている：52→40名、未選択：2→1名
NGS利用経験	今のところ使用予定はない：3→2名、使ったことはないが使用予定がある：45→37名、いずれ使用すると思う：28→22名、時々or頻繁に使っている：65→40名、未選択：3→1名
実習PC環境	Mac持込：61→42名、Win持込：50→36名、アグリバイオPC貸与希望：48→46名、未選択：3→2名
H26年度講習会受講の有無	受講していない：153→119名、受講した：5→3名、未選択：4→4名
H27年度講習会	受講していない：141→108名、受講した：18→15名、未選択：3→3名

受講の有無	
本講習会を知った経緯（複数回答可）	友人・知人：60→48名、アグリバイオインフォマティクス（Rで）塩基配列解析を含む）：53→44名、NBDCウェブサイト：22→17名、メーリングリスト：21→16名、キーワード検索の結果：16→13名、DBCLSウェブサイト：8→6名、SNS（Facebook, Twitter等）：2→2名、分子生物学会：1→1名、その他：18→10名

表 4-2. 出席確認表上で実際に一日でも出席した受講者数 126 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, 27 年度に実施したボランティアの TA 募集をしなかった。これは TA 募集結果に一喜一憂する精神的負担軽減の意味合いが大きいこと、主催機関および共催機関より各日計 2-3 名の TA が見込めたことによる (表 5)。また、第 2 部では、アメリエフ株式会社より講師と共に TA も派遣して頂き、円滑な運営を行うことができた。

実施日	項目	NBDC/DBCLS	東京大学	明治薬科大学	Amellief	
7月19日 : PC環境の構築 (Bio-Linux 8とRのインストール状況確認)		八塚、建石	寺田	杉原	/	
第1部	7月20日	ゲノム解析、塩基配列解析	宮崎、小野	寺田		
	7月21日	トランスクリプトーム解析1	小野、守谷	寺田		
	7月22日	トランスクリプトーム解析2	小野、守谷	寺田		
第2部	7月25日	NGS解析基礎		寺田		三澤
	7月26日	ゲノムReseq, 変異解析	三橋	寺田		尾上
	7月27日	RNA-seq	川嶋	寺田		久保
	7月28日	ChIP-seq	川嶋	寺田		窪川
第3部	8月1日	Linux環境でのデータ解析 : JavaやRの利用法	河野	寺田		杉原
	8月2日	Linux環境でのデータ解析 : マッピング、トリミング、アセンブリ	河野	寺田		杉原
	8月3日	クラウド環境との連携、ロングリードデータの解析	大田	寺田		杉原
	8月4日	トランスクリプトームアセンブリ、発現量推定	宮崎	寺田		杉原

表 5 TA 体制表。

## 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つは受講者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。講習会後の自立した研究活動をより効果的に行えるのは後者であるため、本講習会は後者寄りのスタンスである。受講者のレベルを一定以上に揃えるべく、平成27年度と同程度以上の厳しい事前準備を課した。今回は3部構成としたため、受講希望項目ごとに予習内容は異なる。ここでは、受講者共通の心構えや予習事項、部ごとに課した予習事項などを述べる。

- 受講者共通の心構えや予習事項

[http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html#bioinfo\\_ngo\\_sokushu\\_2016\\_introduction](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngo_sokushu_2016_introduction)

ここではまず、謝辞、講習会の理念、全体的な方針、注意事項を述べている。受講申込み期間（4月4日～5月31日）内から随時情報を更新し、各項目の受講希望者数の変遷や、作成済みの講義資料などを見られるようにした。その他、下記事項についても事前予習を課した。

1. 平成26-27年度講習会の報告書を読み、過去の受講者がどのような感想や要望をもち、世話人がそれに対して次年度にどのような対応を行ってきたのかについて理解を深める。

[http://biosciencedbc.jp/gadget/human/h26\\_ngo\\_report.pdf](http://biosciencedbc.jp/gadget/human/h26_ngo_report.pdf)

[http://biosciencedbc.jp/gadget/human/h27\\_ngo\\_report.pdf](http://biosciencedbc.jp/gadget/human/h27_ngo_report.pdf)

2. 平成28年度講習会の概要として、昨年度との違いや、項目間の対応関係、想定受講者、および予習事項について学ぶ。

[http://biosciencedbc.jp/gadget/human/h28-h27contents\\_1.pdf](http://biosciencedbc.jp/gadget/human/h28-h27contents_1.pdf)

3. 特に7月19日に来られない受講者向けの、事前準備に関する注意喚起。PC環境の構築として、Bio-Linux8（第2部および3部で利用するovaファイル）の導入確認、共有フォルダ設定完了確認、無線LAN設定などの情報を知っておく。

[http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo\\_ngo\\_sokushu\\_2016/20160719\\_pcinfo\\_20160709.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2016/20160719_pcinfo_20160709.pdf)

[http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo\\_ngo\\_sokushu\\_2016/20160719\\_tips\\_20160704.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2016/20160719_tips_20160704.pdf)

- 第1部の予習事項

[http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html#bioinfo\\_ngo\\_sokushu\\_2016\\_about1](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngo_sokushu_2016_about1)

Rを利用するため、R関連の予習を必須とした。

1. Rおよびパッケージのインストール確認。利用予定のRパッケージ導入確認も含む。

Windows用：[http://www.iu.a.u-tokyo.ac.jp/~kadota/R\\_install\\_win.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/R_install_win.pdf)

Macintosh用：[http://www.iu.a.u-tokyo.ac.jp/~kadota/R\\_install\\_mac.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/R_install_mac.pdf)

2. Rの基本的な利用法

Windows用：[http://www.iu.a.u-tokyo.ac.jp/~kadota/R\\_install\\_win.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/R_install_win.pdf)

Macintosh用：[http://www.iu.a.u-tokyo.ac.jp/~kadota/R\\_install\\_mac.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/R_install_mac.pdf)

3. 平成 27 年度講習会の「データ解析環境 R (7 月 29-30 日分)」の自習。

4. 第 1 部の心構え(掟)を読む

[http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo\\_ngs\\_sokushu\\_2016/20160720\\_2\\_2\\_okite\\_20160618.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2016/20160720_2_2_okite_20160618.pdf)

● 第 2 部の予習事項

[http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html#bioinfo\\_ngs\\_sokushu\\_2016\\_about2](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngs_sokushu_2016_about2)

Linux 環境 (Bio-Linux8) を利用するため、Linux 関連の予習を必須とした。日本乳酸菌学会誌の NGS 連載第 4 回終了時点の ova ファイルをベースに、第 2 部で用いるプログラムやファイルを追加したものを利用。

1. Bio-Linux8 (第 2 部で利用する ova ファイル) の導入。第 2 部を担当するアメリエフ株式会社が作成した ova ファイルの URL 情報を通知し、下記手順に従って事前にインストールしておく。

Windows 用 :

[http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB6\\_BioLinux8\\_ova\\_win.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB6_BioLinux8_ova_win.pdf)

Macintosh 用 :

[http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB6\\_BioLinux8\\_ova\\_mac.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB6_BioLinux8_ova_mac.pdf)

2. 共有フォルダ設定確認。

3. Bio-Linux 環境での作業に一通り慣れておく。Bio-Linux 特有の不具合への対処法を身につけておく。

● 第 3 部の予習事項

[http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html#bioinfo\\_ngs\\_sokushu\\_2016\\_about3](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngs_sokushu_2016_about3)

日本乳酸菌学会誌の NGS 連載第 4 回終了時点の ova ファイルをベースに、第 3 部で用いるプログラムやファイルを追加したものを利用。第 2 部用 ova ファイルとは独立に提供した。

1. 第 1 部と第 2 部の予習事項内容

2. 共有フォルダ設定確認

3. 8/3 参加者のみ、日本乳酸菌学会誌の NGS 連載第 6 回ウェブ資料の DDBJ Pipeline 関連部分を眺めておく。

[http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB6\\_suppl\\_win\\_20160617.pdf](http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB6_suppl_win_20160617.pdf)

## 2-6. 講習内容

講習会初日（7月19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のウェブサイト](#)より公開予定である。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】2016年7月19日（火）10:30-18:15

【題目】PC環境の構築（Bio-Linux 8とRのインストール状況確認）

【担当】主催・共催機関

【内容】自習形式。講習会会場を解放し、予習事項の最終確認を各自で行ってもらった。また、持ち込みPCの無線LAN設定や名札配布などの各種手続きが行われた。受講者滞在時間を概ね2時間単位で区切り、時間帯ごとの来場予定者を把握していたため、混乱はなかった。

第1部のRについては、インストール作業を行ってこなかった場合や事前インストールに失敗している場合、講義室で改めてトライしてもらった。また、実習で必要となるライブラリファイルをUSBメモリに準備しておき、必要に応じて貸し出した。管理者権限でパッケージをインストールすることが必要な場合に、その権限が無く、インストールに失敗するケースが散見された。

第2, 3部の予習内容は、VirtualBoxおよびその拡張機能を事前にインストールし、メールで指定されたURLから実習環境設定済みの仮想環境ovaファイルをセットアップしておくこと、ホストOS・ゲストOS間の共有フォルダ設定を行うことであった。7/19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者に対しては、メールで指定していたのと同じ実習環境設定済みのovaファイルをUSBメモリで渡し、7/25以降の講習を受講可能となるPC環境を構築してもらった。詳細な手順書（[Windows版](#)、[Mac版](#)）を配布し、資料を見ながら自力で行うことを基本とした。一部の機種（VAIOなど）ではBIOS設定の変更が必要であった。7/25 第2部開講後に、指定したovaファイルではなく、乳酸菌NGS連載第5回終了時点のovaファイルや、Bio-Linuxの本家サイトから公開されているovaファイルのみを導入してきた受講者が数名おり、講義中にUSBメモリからの環境構築が必要なケースがあった。

全体として、自習形式の受講者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も大きい。この方式は受講人数規模が大きく、長期間にわたる他の講習会でも積極的に取り入れたほうがいだろう。



【日時】 2016年7月20日（水） 10:30-18:15

【題目】〔第1部〕 統計解析：ゲノム解析、塩基配列解析

【講師】 門田幸二（東京大学）

【内容】 NGS解析手段をリストアップ、ウェブツール（DDBJ Pipeline）との連携について紹介。乳酸菌ゲノム解読論文（PMID: 25879859）を元に、MiSeqデータをDDBJ Pipeline上で利用可能なPlatanusを用いて*de novo*アセンブリした結果をRで解析し、アセンブリ過程の理解を深める。公共NGSデータベースDDBJ SRA（DRA）について簡単に説明し、DRAからダウンロードしたデータを解析。用語（リード、single-end、paired-end、コンティグ、scaffold、gap close）のおさらい。*de novo*アセンブリのステップをプログラムVelvetとPlatanusの違いと共に確認。ステップ間での配列数の増減も含め、データ解析の全体像の理解を深める目的でRを利用。

用意しておいたPlatanusでアセンブルを実行後のデータplatanusResult.zipを用いて、1塩基ごとの出現頻度を解析するk-mer解析（k=1）、2連続塩基の出現頻度を解析するk-mer解析（k=2）を行った。さらに、塩基配列解析を行うための基本スキルの復習や作図について、具体的には出現確率とboxplotについて各種オプションを含めて説明。

*de novo*アセンブリ時のエラー補正やゲノムサイズ推定の基本的な考え方を学ぶため、塩基の存在比率A:C:G:T = 22:28:28:22のランダム塩基配列を生成。仮想ゲノム・仮想NGSデータを作製して、k-mer解析の応用としてk=6,8,10,12の場合について説明。またカバレッジを変更したゲノムサイズ推定や、シークエンスエラーを含む場合を仮定し、任意の位置の塩基置換（G→C, C→G, A→T, T→A）を行ってからk-mer出現頻度分布の解析。置換前全500リード、置換なしの450リード、置換後の50リード、置換なしの450リードおよび置換後の50リードをマージした場合のそれぞれについて、ゲノムサイズを推定して比較。シークエンスエラーを多く含むk-merの出現頻度は低い傾向にあるため、それらを除けばよい。



【日時】 2016年7月21日（木） 10:30-18:15

【題目】 [第1部] 統計解析：トランスクリプトーム解析1

【講師】 門田幸二（東京大学）

【内容】 カウントデータの定義は「マップされたリード数をカウントしたデータ」で数値行列であること、RPK（**Reads per kilobase**）補正、総リード数を100万に揃えるRPM（**Reads per million**）補正について確認。補正せずに発現変動遺伝子（DEG: **Differentially Expressed Gene**）検出を行うことは間違いであることを確認。

ReCount DBにあるヒトのデータ、bodymap、gilad、これらをマージしたデータのサンプル間クラスタリングを実行。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）ことを示した。

様々な2群間比較。最も類似度が低いヒト（HS） vs. アカゲザル（RM）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）の考え方を説明。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、HS vs. RM の結果は2,489 DEGs、類似度が近いHS vs. チンパンジー（PT）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）と同じ傾向にあることを示した。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を実行。実際にDEG がわずかししか検出されないこと、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】2016年7月22日（金）10:30-18:15

【題目】〔第1部〕統計解析：トランスクリプトーム解析2

【講師】門田幸二（東京大学）

【内容】反復あり3群間比較としてTCCによるANOVA的な解析、すなわちどの群間で違いがあるかは問わずにどこかの群間で発現に差があるものを検出する解析を行い、ヒト（HS）、チンパンジー（PT）、アカゲザル（RM）のどこかの群間で発現変動している遺伝子を検出する方法を示した。

ANOVA後にどこの群間で差があるかを調べるために行う一般的な方法はpost-hoc test。クラスラベル情報を基に作成したデザイン行列について解説。デザイン行列から特定の列を除いた2群間比較と、コントラスト情報を与えるやり方で同じ結果が得られることを確認。3群以上の比較ではANOVA解析を先に実行するのが一般的。またpost-hoc testの2群間比較と通常の2群間比較の違いについても説明。

遺伝子間クラスタリングについて、ANOVA的な結果が得られる「MBCluster.Seq単体での利用」、頑健な正規化を組み合わせた遺伝子間クラスタリング結果が得られる「TCC正規化とMBCluster.Seq」を組み合わせた方法、さらに「TCC正規化とMBCluster.Seq」に「TCC発現変動解析」を組み合わせた最もおススメの解析方法について紹介。FDR閾値を満たすものに限定して各クラスターに属する遺伝子数を概観可能であることなども解説。

反復あり3群間比較についてbaySeqによる発現パターン分類、ANOVA的な解析のため発現パターン分類できなかったTCCの発現変動解析結果とbaySeqの結果を合わせて解析可能であること、実際にそのような解析を行った原著論文中の表を参照しながら解説。最後に反復なし3群間比較について、TCC、DESeq2のそれぞれについて実行し、q-valueの算出方法の違いについて説明。若干結果が異なる理由は、反復なしのデータでは、TCCは内部的にDESeq2の解析手順を繰り返して頑健な正規化を行っているためであることを説明。



【日時】 2016年7月25日（月） 10:30-18:15

【題目】 [第2部] NGS解析（初～中級）：NGS解析基礎

【講師】 山口昌雄、窪川美雪（アメリエフ株式会社）

【内容】 午前は山口講師による講義、午後は窪川講師による実習。NGSデータ解析で主に使用するファイル形式（FASTA, FASTQ, SAM/BAM, VCF, BED, GFF/GTF）や圧縮形式を解析フローと共に紹介し、それぞれの見方や特徴などについて詳しく解説。またSAM/BAMファイルを扱うsamtoolsやBEDファイルを扱うBEDtoolsについても紹介した。

実習で使用するソフトウェアや解析データ等を全て含んだ仮想環境Bio-Linux上で、データの可視化手段としてよく用いられるIGV（Integrative Genomics Viewer）について説明。インデックスの作成、samtoolsコマンドを用いたBAMファイルのソート、VCFファイルの読み込みなどを紹介。NGSデータ解析において最も重要なことは、解析データのクオリティコントロール（QC）。クオリティチェックやクオリティクリーニングについて説明。

シーケンスリードのクオリティを確認するFastQCについて利用法などを説明。QC Reportの見方を詳しく解説。実際にBio-Linux上でクオリティチェックとクリーニングを行うべく、samtoolsを用いたマッピング率やマッピング状況の確認し、結果（idxstats）の見方を詳しく解説。NGSデータのマッピングには専用の高速なソフトが使用されることを紹介。主なマッピングソフトは、Reseq用はBWAやBowtie、RNA-seq用はSTARやHISAT、Methyl-seq用はBSMAPやBisulfighter。

最後に新しいソフトウェアの導入する場合の手順について、アダプター除去などの機能を持つTrimmomaticの導入を例として、様々なファイル圧縮・解凍形式についての解説や、javaを用いた実行方法について説明した。ソフトウェア開発共有サービスGitHubについて紹介。gz圧縮ファイルを扱うコマンドとして、圧縮したままファイルの中を見るzlessや、複数の圧縮ファイルをまとめて一つのgzファイルにするテクニックも紹介。



【日時】 2016年7月26日（火） 10:30-18:15

【題目】〔第2部〕 NGS解析（初～中級）：ゲノムReseq、変異解析

【講師】 山口昌雄、三澤拓真（アメリエフ株式会社）

【内容】 午前は山口講師による講義、午後は三澤講師による実習。講義では、ReseqとはDNAの変異検出を目的としたワークフローの総称。検出可能な遺伝子変異は、SNV (single nucleotide variant)、InDel (insertion and deletion)、SV (structural variation)など。公開データ取得方法としてSRA, ENA, DRAを紹介し、DRAからのダウンロードを実際に示しながら、wgetによるダウンロード方法について詳しく解説。ダウンロードした公開データをそのまま使用するのは危険であり、クオリティコントロール (QC)が必要である。ゲノム解析でよく用いられるQC用ツールは、FastQC、FASTX-toolkit、PRINSEQ、Trimmomaticなど。マッピングとは、各リードがリファレンスゲノムのどこに位置するかを調べること。主なツールや特徴、たとえばBWAは高速だがエラーが多いことなど。変異検出とは、マップされたリードを元にリファレンスゲノムとの比較を行うこと、実行前にRealignment、Base Recalibrationが必要である。様々な検出ツールの中ではGATK HaplotypeCallerを使用すればRealignmentが不要であることなど。

実習では、解析パイプラインとは、ある出力結果が次の解析ソフトの入力ファイルとなる連続した解析処理の流れであることを説明。Reseqの解析パイプラインの一例として、公開データのFASTQファイルから1. FastQCによるクオリティチェック、2. TrimmomaticによるQC、3. BWA memによるマッピング、インデックス作成、4. GATK HaplotypeCallerによる変異検出と出力されたVCFファイルの確認、5. GATK Variant Filtrationによる変異のフィルタリング、6. snpEffによるアノテーション、という1つのフローを紹介。各手順について実習をしながら詳細を解説し、IGVによる解析結果の可視化をBio-Linux環境で行った。



【日時】 2016年7月27日（水） 10:30-18:15

【題目】 [第2部] NGS解析（初～中級）：RNA-seq

【講師】 山口昌雄、尾上広祐（アメリエフ株式会社）

【内容】 午前は山口講師による講義、午後は尾上講師による実習。RNA-seqとは、mRNAをキャプチャしてNGS機器でシーケンシングする方法。リファレンスがある場合は、既知遺伝子にマッピングするか、リファレンスにマッピングして既知遺伝子の発現量を定量。リファレンスがない場合は、アセンブリして転写物構造を予測した配列に対してマッピングするか、近縁種のゲノムのリファレンスにマッピングする。1つのRNA-seq解析フローを示し、各ステップでどのようなソフトウェアがあるかを示した。まず当日の実習で解析に使用するリファレンスゲノム、解析対象のシーケンスデータについて、公開データの取得方法と共に紹介。解析対象とする公開済みシーケンスデータの取得方法についても説明し、SRA形式ファイルのみしか提供されていない場合は自分でFASTQファイルへの変換が必要である。QC用プログラムFastQCを実行し、結果のhtmlファイルの見方を解説。別のQC用プログラムPRINSEQについても機能や使用方法を説明。HISAT2のインストール、基本的な使い方の確認法の説明、およびTopHat2よりも精度・速度ともに向上していることなどの特徴説明。発現定量として、各種バイアスの補正の必要性とよく使われる指標（RPKMやFPKM）の説明。発現定量によく用いられてきたCufflinksの解説。

実習では、実際にPRINSEQでQC、HISAT2でマッピング、IGVを用いたマッピング結果の可視化、Cufflinksで発現定量、cuffmergeとcuffdiffによる発現比較、そしてcummeRbundの紹介まで行った。マッピングではHISAT2のアルゴリズムや特徴（スプライシングを考慮してゲノム配列にマッピングする、hierarchical indexingを用いることで高速で高感度のアラインメントが可能）、オプション説明も含めて紹介した。講義で紹介した流れはあくまでも一例であり、ツールの選択に「正解」はないことも強調した。



【日時】2016年7月28日（木）10:30-18:15

【題目】〔第2部〕NGS解析（初～中級）：ChIP-seq

【講師】山口昌雄、久保竜一（アメリエフ株式会社）

【内容】午前は山口講師による講義、午後は久保講師による実習。ChIP-seqとは、**Chromatin Immuno Precipitation + NGS sequencing**の略で、クロマチン免疫沈降により濃縮したゲノム領域をシーケンスする方法。主な解析対象は、タンパク質とDNAの相互作用、ヒストン修飾など。転写調節領域が主な解析対象。免疫沈降(IP)のバックグラウンドノイズを低減するため、inputと呼ばれるIPを行っていないサンプルをコントロールとして使用する。検出したピークを抗体に非特異的なものとして取り除き、抗体が特異的に結合した領域を信頼度の高いピークとして得る。

ChIP-seq解析フローについて、各ステップで使用される代表的なソフトウェアを示した。QCでは、75bp以下など短い場合が多いのでリード長に注意すべし。マッピングでは、Reseqでも使用されるソフトがよく使われる。最近ギャップアライメントに対応し、マッピング精度が高いbowtie2が使用されることが多い。ピーク検出ソフトは、濃縮した領域のリードの頂点を検出する。MACS2が被引用数の点でデファクトスタンダードだが、SICERというプログラムも用いられる。ピーク検出後に、ピークがゲノム上のどのような位置に存在するかをアノテーションするソフトやピークに共通のモチーフ検索をするソフトについても解説された。また、NCBI SRAを例として、公開データの取得やデータの探し方についても解説がなされた。

実習では事前に準備された公開データを用いて、FastQCによるQC、Trimmomaticによるトリミング・低クオリティリードの除去、Bowtie2によるマッピング、MACS2によるピーク検出では詳細なオプション説明もなされた。IGVで可視化し、検出したピークを確認する方法を学んだ。アノテーションはSnEffで行い、モチーフ検索にはRのrGADEMパッケージが用いられた。



【日時】 2016年8月1日（月） 10:30-18:15

【題目】 [第3部] NGS解析（中～上級）：Linux環境でのデータ解析：JavaやRの利用法

【講師】 門田幸二（東京大学）

【内容】 第3部で使用するソフトウェアや解析データ等を全て含んだ仮想環境Bio-Linux上で行った。日本乳酸菌学会誌に連載中のNGS連載第5回の内容に沿って行い、乳酸菌*Lactobasillus casei* 12AのRNA-seqデータ（の一部）を使用した。具体的には、paired-endファイル（forward側107bp、reverse側93bp）の最初の100万リードを用いたが、解析データの作成など予習事項として課した連載第4回までの内容の復習も兼ねて改めて説明した。FastQCによるQC、そしてFaQCsによるアダプター配列除去を行った。その結果を入力として再度FastQCを実行して、アダプターやプライマー配列情報が消滅していることを確認した。

トランスクリプトーム用*de novo*アセンブラのTrinityとRockhopper2を紹介し、Javaプログラムである後者の設定と実行方法を解説。GUI版でps、kill、nohupといったLinuxのプロセス管理についてのTipsも紹介した後、コマンドライン版でクラスパスを設定。paired-end、single-endのforward側データのみ、reverse側のみのアセンブリを実行し結果を確認した。

Linux環境でのRの利用法をRの起動や終了から解説。基本的な利用法をおさらいして、第1部の予習事項がわかっているならば、ゲストOS環境でのR起動後の状態はホストOS環境と基本的に同じことを示した。RのQuasRパッケージのインストール状況確認、source関数を使用とバッチモードについて説明した。TipsとしてRの終了方法やバージョン確認、文字化け対策についても触れた。最後にバッチモードの発展形としてslaveオプションやRスクリプト内での入力ファイルの絶対パス指定、gz形式圧縮ファイルのまま解析する方法について説明、実行した。



【日時】2016年8月2日（火）10:30-18:15

【題目】〔第3部〕NGS解析（中～上級）：Linux環境でのデータ解析：マッピング、トリミング、アセンブリ

【講師】門田幸二（東京大学）

【内容】Bio-Linux上のR環境で、QuasRパッケージを用いてRNA-seqデータの乳酸菌ゲノム配列へのマッピングを行い、forward側の100-107bp付近に問題があることを特定。問題のある領域を2つの手段（RおよびFASTX-Toolkitのfastx\_trimmer）でトリミングする作業を実行。トリム後のデータでRockhopper2による*de novo*トランスクリプトームアセンブリを再度実行。トリミング前後のアセンブリ結果とマッピング結果の評価を行い、トリム後は大幅に改善していることを確認。環境設定ファイル（.zshrc）へのechoコマンドを用いたクラスパス設定情報の追加書き込みなど、各種Tipsも交えて紹介。

Illumina MiSeqデータ（乳酸菌ゲノム；DRR024501）の特徴を説明。このデータはIllumina Universal Adapterが混入しているが、アダプター除去済みのデータとして提供される場合もあるため、一般的にはアダプター関連は気にしなくてもいい場合もあることなどを紹介。FaQCs 実行前後のFastQC結果を比較してアダプターが除去されていることを確認。

Bio-Linuxにプレインストールされている*de novo*ゲノムアセンブラVelvet (ver1.2.09)を、k値の上限であるk=31で実行。k値の意味についても簡単に説明。上限値を超えたk=141では同じ結果になることを、復習を兼ねてRで確認。k=31より大きい値を用いたアセンブリを実行するため、Velvet (ver1.2.10)をダウンロードしてmakeでインストール。コンパイル時のオプションの意味やインストールマニュアルの見方も説明。k=31を超えてVelvetを実行できるようになり、様々なk値で実行して比較した。仮想環境のデフォルトはメモリ2GBだったが、使用可能な最大メモリを4GB以上に設定すると計算時間が大幅に短縮可能である。使用メモリ設定変更方法についても説明し、メモリ2GBで約1時間かかるところを、4GBに変更して約10分で終了するところまで。



【日時】 2016年8月3日（水） 10:30-18:15

【題目】 [第3部] NGS解析（中～上級）：クラウド環境との連携、ロングリードデータの解析

【講師】 門田幸二（東京大学）

【内容】 前日までにVelvet実行時にk=171がいいと主観的に判断したが、客観的な判定結果を返すプログラムKmerGenieのインストールからスタート。KmerGenieは、makeコマンドを利用してインストールする2例目として、また最適k値と同時に出力するゲノムサイズ推定周辺を主目的として紹介。forward側のみに対し、探索範囲としてk=31～191を対象として実行し、推奨値k=87を得た。次にpaired-endで実行する方法を調べて実行し、推奨値k=141を得た。forward側のみの場合と大きく異なったが、ゲノムサイズの推定結果はほぼ同じであることを確認。

配列長によるフィルタリングを行うPythonプログラムの実行に際し、「パスを通す」別の手段を伝授。この結果、配列長フィルタリング（300 bp未満のコンティグを除去）前はゲノムサイズに相当する総塩基数が最大約5.7MB（k=111）であったが、フィルタリング後は最大約2.6MB（k=151）となり、真の値に近づいたことを実感してもらった。DDBJ Pipelineを紹介し、Bio-Linux環境下でのVelvet実行結果と同じになることを確認。また、DDBJ Pipeline内のプログラムPlatanusを実行し、Velvetに比べてコンティグ数が大幅に改善されたことを確認。配列長フィルタリングを行って得られた52配列という結果は、原著論文の結果（53配列）とほぼ同じ。

PacBioデータと公共DBのデータ形式の説明。FASTQへの変換時にクオリティ値が高いIlluminaデータ（エラー率 約1%）はリード数が若干減る程度だが、PacBioデータ（エラー率 約13%）は相当減る。また公共DBから提供されているPacBioデータのFASTQファイルを入力としてFastQCを実行。エラー率が比較的高いPacBioデータにも関わらず、クオリティスコアが高かったのは、フィルタリングを行って生き残った少ないリードデータのみを対象としたため。NCBIが提供するSRA Toolkitをapt-getコマンドでインストールする方法を説明し、最後にDDBJ PipelineでPacBioデータを解析するHGAPを実行する場合について資料のみで解説した。



【日時】 2016年8月4日（木） 10:30-18:15

【題目】 [第3部] NGS解析（中～上級）：トランスクリプトームアセンブリ、発現量推定

【講師】 門田幸二（東京大学）

【内容】 第3部のこれまでの内容や解析データのおさらいと問題設定。近縁種などゲノム配列を利用せずとも、paired-end RNA-seqデータのみから合理的にそれなりの結果を得られるという内容。FastQC実行結果のKmer\_Content項目を眺め、forward側の99-101塩基目あたりが怪しいと目星をつける。Rockhopper2によるアセンブリ結果ファイル群をgrepでざっくり眺める基本戦略を紹介したのち、fastx-trimmerで様々なトリミング条件を実行した結果をgrepと組み合わせで全体像を把握。このとき、出力ファイル名の名前も一目でトリム条件がわかるようにしたほうが便利であるという実用上のTipsも紹介。結論として、forward側の5'末端はそのまま、3'末端のみ100塩基目以降をトリミングすると配列数が多くなることを確認。どの条件が最も良いかを選ぶには悩ましい結果であったが、paired-endで実行することの利点は示すことができた。

Rockhopper2以外の*de novo* Transcriptome Assembly用プログラムとしてTrinityとBridgerを紹介。Trinityのインストールから実行まで。apt-get経由のインストールだとバージョンが相当古いことなど。事前確認ではBridgerのインストールに失敗していたが、講義中に受講者より依存関係にあるBoostパッケージのバージョンを落としてapt-getでインストールすればうまくいくという報告があり、数人の受講者が実際にうまくいくことを確認した。最後に発現量推定プログラムTIGAR2の紹介。トランスクリプトーム配列とRNA-seqデータを入力として実行し、得られた結果の解釈について説明。FPKM値の手計算や、発現変動解析TCCとの組合せも述べた。



## 2-7. アンケート結果

講習会終了時の受講者の課題、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 164 名うち、1 項目以上の受講者 126 名を対象とした。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 2016 年 7 月 19 日～9 月 16 日であった。9 月 16 日時点で未回答であった 20 名について、個人宛メールによる督促を行い、最終的に 119 名（回収率 94.4%；未回答者 7 名）から回答を得た。最終的な調査期間は 2016 年 7 月 19 日～9 月 29 日である。

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、事前予習に関する設問と合わせて表 6 に示す。「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（⑦は未選択）。設問 2 に関してのみ、①少ない、②やや少ない、③適量である、④やや多い、⑤多い、⑥事前予習していない、⑦未選択である。①がよい結果というわけではないため、参考までに灰色で示した。

設問Noと設問内容	①	②	③	④	⑤	⑥	⑦
設問1. 「事前予習」について、内容を十分理解できた	59	47	5	5	3	0	0
設問2. 「事前予習」について、分量は適量であった	2	5	66	38	8	0	0
設問3：PC環境の構築（Bio-Linux 8とRのインストール状況確認）（7月19日）	76	17	1	1	0	21	3
設問4：第1部ゲノム解析、塩基配列解析（7月20日）	63	30	2	2	0	18	4
設問5：第1部トランスクリプトーム解析1（7月21日）	49	36	0	1	1	28	4
設問6：第1部トランスクリプトーム解析2（7月22日）	51	39	1	1	1	23	3
設問7：第2部NGS解析基礎（7月25日）	60	26	5	4	0	21	3
設問8：第2部ゲノムReseq, 変異解析（7月26日）	61	28	2	2	0	23	3
設問9：第2部RNA-seq（7月27日）	54	27	7	3	1	24	3
設問10：第2部ChIP-seq（7月28日）	46	27	6	4	0	33	3
設問11：第3部Linux環境でのデータ解析：JavaやRの利用法（8月1日）	35	23	5	5	0	44	7
設問12：第3部Linux環境でのデータ解析：マッピング、トリミング、アセンブリ（8月2日）	36	21	6	3	1	45	7
設問13：第3部クラウド環境との連携、ロングリードデータの解析（8月3日）	27	19	7	5	0	53	8
設問14：第3部トランスクリプトームアセンブリ、発現量推定（8月4日）	26	23	4	4	0	54	8

表 6. 講義内容の理解度に対するアンケート結果。

講義全体についての結果を表 7 に示す。計 4 つの設問（設問 15～18）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（⑥は未選択）。また、上記 4 設問について自由記述欄（アンケートの設問 19 に相当）も設けた。寄せられた計 67 件の生の記載内容は、付録 1 に示した。

設問Noと設問内容	①	②	③	④	⑤	⑥
設問15. 講義内容の分量は適切であった	69	37	8	5	0	0
設問16. 講義のレベルは適切であった	65	44	7	3	0	0
設問17. 講義で取り上げられた事柄は興味ある内容であった	80	32	6	1	0	0
設問18. この講義で学んだことは今後役に立つと思った	100	14	3	1	0	1

表 7. 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 8 に示す。計 7 つの設問に対し、上記と同じく 5 つの選択肢を用意した（⑥は未選択）。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 24 に相当）も設けた。寄せられた計 75 件の生の記載内容は、付録 2 に示した。設問 26 に関しては、①がよい結果というわけではないため、参考までに灰色で示した。

設問 No と設問内容	①	②	③	④	⑤	⑥
設問 20. 説明の仕方はわかり易かった	91	22	6	0	0	0
設問 21. 講義の進捗や時間配分は適切であった	73	35	6	4	1	0
設問 22. 配布資料などの教材は適切であった	98	17	4	0	0	0
設問 23. 講義に TA は必要である	89	18	7	4	1	0
設問25. 講義の内容は全体的に理解できた	67	46	3	1	1	1
設問26. 講習会の講義資料や講義動画があれば、講習会に参加しなくても良いと思う	23	10	23	45	17	1
設問27. 総合的に評価して、あなたはこの講義に満足しましたか	96	21	0	1	0	1

表 8. 講義方法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 つの設問を用意した。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。

提案	回答件数	付録番号
提案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	71	<u>付録3</u>
提案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	69	<u>付録4</u>
提案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	43	<u>付録5</u>
提案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	69	<u>付録6</u>

提案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	30	<u>付録7</u>
--	----	------------

表 9. 自由記述方式の 5 つの設問（提案 1～5）、回答件数、および付録番号との対応関係。

### 3. 総評

#### 3-1. 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、平成 27 年度に計 14 日間かけて施行実施した NGS ハンズオン講習会の結果を踏まえ、平成 26, 27 年度受講者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）。昨年度ほぼクリアされた課題「項目間の連携強化」は、今年度も重要課題として継続実施した。具体策はできる限り少人数の講師陣で実施することであるが、この体制構築に多大な貢献をいただいたのが、初年度からご協力いただいているアメリエフ株式会社の諸氏である。アメリエフ株式会社（代表：山口昌雄 氏）には、7/19～8/4 の計 12 日間のうち第 2 部の全て（4 日分）を担当いただいた。第 2 部各日午前の講義を担当頂いた山口昌雄 講師、午後の実習を担当頂いた窪川美雪 講師（7/25）、三澤拓真 講師（7/26）、尾上広祐 講師（7/27）、久保竜一 講師（7/28）の諸氏に深く感謝する。なお、実習を担当頂いた諸氏は、前日の TA としてサポートして頂き（7/28 は窪川氏が TA）、円滑な進行の助けとなった。アメリエフ株式会社には、7 月 25 日夜の「情報交換会」も主催していただいた。予想を超える約 40 名の参加があり、会費のみでは到底足りないクオリティのメニューも提供していただいた。窪川氏には、講義内容や講義資料作成の全体統括を担当していただいたことも記しておきたい。通常業務のみでも大変な中、講義資料も全面的にアップデートしていただき、最新情報提供に尽力いただいた。昨年度に引き続き、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。

本講習会では、昨年度まで実施したボランティア TA（講義補助）の公募を行わなかった。これは、①上述のように第 2 部はアメリエフ株式会社からの TA 派遣が見込まれたこと、②残りの第 1 部と 3 部は一定のスキルを持つ受講者に限定していたこと、③例年不具合が生じる Macintosh ユーザ対策として共催機関 2（明治薬科大学）の協力があつたことが大きい（表 2）。また、受講申し込み時に「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。」というお願い事項も示したが、実際に TA が足りない局面では受講者同士でサポートしあう場面が多々見られた。これは講習会の早い段階で「情報交換会（7/25 の夜）」を開催したおかげで受講者同士の横のつながりが強化されたためであろう。TA 的な役割を果たしてくれた多くの受講者諸氏に感謝したい。また、アンケートにも記載されていたが、今年度は macintosh ユーザの不具合はほぼ皆無であった。これは、主に講義資料事前チェックや講習会当日のサポートをいただいた明治薬科大学の杉原 稔氏の尽力によるものである。

強調しておきたいのは、本講習会は JST-NBDC（主催機関 2）による全面的なサポートで成り立っている点である。JST-NBDC は、昨年同様、受講申し込み業務、講義補助（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の統括を担当した。受講者の多くにとっては、本講習会は顔の見える世話人のイメージしかないかもしれないが、主な裏方業務を全面的に JST-NBDC が担当し、講義内容の充実により多くのエフォートを費やせたおかげである点を重ねて強調しておきたい。昨年同様、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ていないが、実務担当の諸氏（白鳥、館澤、眞後、三橋、大波、川嶋、宮崎、櫛田、信定、佐久間、森、八塚、松平、建石）に感謝したい。特に白鳥亜希子 氏には、昨年度に引き続き、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。昨年度に引き続き、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。JST-NBDC 統括のもと、DBCLS (共催機関 1) の諸氏 (小野、守谷、河野、大田) には、主に第 1 部と 3 部の TA を担当いただいた (表 5)。安定の仕事ぶりで、質問の無茶振りにも対応していただいた。

本講習会の主な予習事項は、日本乳酸菌学会誌の NGS 連載原稿に基づいている。本連載は、鈴木チセ編集長 (当時)、坂本光央編集長 (2016 年現在) をはじめとした編集委員諸氏の原稿に対する建設的なコメントおよびご協力により、毎回高いクオリティを維持した状態でタイムリーな原稿 PDF の公開が実現されている。連載 2-5 回の筆頭著者は孫建強氏、第 6-8 回の筆頭は谷澤靖洋氏である。他の共著者 (清水謙多郎、西岡輔、三浦文、湯敏、神沼英里、中村保一、遠野雅徳、大崎研、寺田朋子) の貢献、そして原稿内容に対する有意義なコメントやアドバイスをいただいた諸氏 (有田正規、伊藤武彦) にも触れておきたい。また、講習会で利用させていただいた各種プログラムの開発者にも感謝申し上げる。特に上記以外の contributor として、第 3 部の最終日に利用した TIGAR2 開発者の成相直樹氏からは有意義なアドバイスをいただいた。JSPS 科研費では、22128008, 24500359, 15K06919 の成果物が講習会で紹介されている。これらは、主に世話人が関係する TCC、サンプル間クラスタリングと発現変動遺伝子の関係、ロングリードデータ解析関連のグラントのみであるが、遺伝研で提供されている DDBJ Pipeline など数多くのグラント成果物が日本の研究の進展やバイオインフォマティクス教育に役立っていることを改めて記しておきたい。

世話人と同じ所属機関 (主催機関 1) ではあるものの、本講習会に尽力いただいた諸氏にも触れておきたい。非常に厳しい状況の中で計 3 週間に渡って講習会会場を確保できたのは、清水謙多郎先生の尽力によるものである。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。アグリバイオインフォマティクスが関与する講習会の安定した開催実績は関係諸氏の尽力によるものである。これらの 3 氏に加えて、今年度は寺田朋子氏が本講習会を主務として参画した。寺田氏は、主に Windows PC 側の動作確認、USB メモリ管理、TA、報告書のドラフト作成などを担当した。昨年度の講習会会場での実働部隊は、三浦氏と世話人のみであり、相当疲弊した。しかし今年度は、安定した仕事ぶりが保証されている三浦氏に加え、昨年度世話人が行っていた仕事の多くを寺田氏が担当した。そのおかげで、非常に穏やかな精神状態で講習会を迎え、滞りなく講習会を終えることができた。

今年度は JST-NBDC に対し、本講習会を万全の態勢で実施するために必要な要望を数多く示した。結果として、それらは全て受け入れていただいた。ノート PC を含む講習会で使用した機材の多く、そして三浦・寺田両氏の人件費の一部は、JST-NBDC との共同研究経費で賄われている。バイオインフォマティクス人材育成は、講師や TA などの受講者らの目に触れる受講者たちのみで行われているわけではない。基礎教育の重要性を正しく理解し、必要な資金と人員を派遣していただいた JST-NBDC 抜きに本講習会は成立しえない。

## 3-2. アンケート結果について

### 3-2-1. アンケート全体 (表 6-8)

昨年度の結果と比べると、各講義（項目）についての講義内容の理解度に関する結果（表 6；設問 1～14）は、明らかにネガティブな方向（①の割合が減り、④や⑤が増加）にシフトしていた。講義全体についての結果（表 7；設問 15～18）についても、ややネガティブなほうにシフトしていた。昨年度との大きな違いは、講義のレベルを全体的に引き上げた点である。特に、第 1 部および第 3 部の内容は、主に昨年度受講者が引き続いて受講することを想定して作成した。ネガティブな結果は、この難易度引き上げが主因であろう。尚、講義方法についての結果（表 8；設問 20～27）は、ポジティブな方向にシフトしていた。全体としては、予習という義務を果たした受講者の本講習会に対する満足度は高いと結論づけることができる。

「ここが分からなかった、分かりづらかった」といったコメントが散見されたのは残念である。一見正論かもしれないが、本講習会に限って言えば受講者側にも責任の一旦はあると思われる。なぜなら、わからないことはその場で聞くように幾度となく呼びかけ、また TA を配置して「講師側に手を挙げる全体質問」以外にも「後ろ向きで手を挙げる個別質問」ができるようにしたためである。受講必須要件に「分からないことがあればその場で質問します。」を掲げてよかったかもしれない。

### 3-2-2. 事前予習について（表 6 の設問 1-2）

表 6 の設問 2 は分量が適量であったかどうかに関する問いであったが、「③適量である」が 66 名、「④やや多い」が 38 名という結果であった。世話人は、昨年度以上の事前予習を課していたため、「③適量である」よりも「④やや多い」のほうが多い程度を想定していた。「あれだけ多い予習分量を適量と感じる受講者が多数派なのか！」というのが率直な感想であるが、裏を返せばそれだけ本気で学びたい意識レベルの高い受講者が多数派を占める講習会だったのであろう。

### 3-2-3. ネガティブな回答結果に対する詳細な分析

講義内容も異なるため昨年度の結果との直接的な比較は難しいが、批判的な回答に相当する「④あまりそう思わない」や「⑤そう思わない」が比較的多く選択されていた点は気になる点である。特に、表 7 の設問 17 および 18 で「④あまりそう思わない」を選択した受講者や、表 8 の設問 25 で「⑤そう思わない」を選択した受講者、そして表 8 の設問 27 で「④あまりそう思わない」を選択した受講者が存在する点については論評すべきであろう。調査の結果、これらを選択したのは同一人物（受講者 A とする）であり、他の項目についても全般的に批判的な回答を寄せていた。具体的には、表 6 中の設問 1, 5, 6, 9 で⑤、設問 4, 7, 8, 11, 12 の計 5 項目で④、表 7 の設問 16 で④、表 8 の設問 25 で⑤であった。それ以外の項目（設問 10, 13, 14；7 月 28 日、8 月 3, 4 日に相当）は「⑥受講していない」であり、表 6 全体にわたり（設問 2 の③適量であるを除いて）ポジティブな評価は皆無であった。つまり、最低ランクの評価は、ほとんどが受講者 A のみに起因するものであった。これは、おそらく単純に受講者 A は予習という義務を果たしていないことに起因すると思われる。その理由は、事前アンケートで Linux および R 利用経験がないと自己申告した人物が、表 6 の設問 1（「事前予習」について、内容を十分理解できた）で「⑤そう思わない」と回答しているためである。しかし、本来、表 6 の設問 1 で「⑤そう思わない」にならない努力が受講者側の責任において受講前までになされておくべきである。設問 19 では、「今回の講習はパッケ

ージ化されたソフトを使うのではなく、自分でデータを操作し、解析する玄人向けの講座と感じました。ややマニアックな印象がありましたが、運営の都合上、講習会のレベルを下げられないのかもしれないと回答している。「パッケージ化されたソフトを使わない玄人向けの講座」かどうかは、開催概要や予習事項をよく読み、事前公開されている講義資料を眺めてきちんと予習すればわかることである。

表6の設問1で「⑤そう思わない」を選択した受講者は、他に2名存在した。そのうちの1名(受講者Bとする)は、表6の設問1と2以外全てで「④あまりそう思わない」としていた(設問2の事前予習の分量は「⑤多い」と回答)。受講者Bは、表8の設問25でも「④あまりそう思わない」としていた(設問27は②であった)。ただ、自由記載の設問19では、「NGSを実際に使ったことがなかったので、よくわからないこともあったが、実際に使用する前に受講する機会があってよかったと思う。2年ぐらい受講すると良いのではないかと思った。」と回答し、また設問24でも「早口な先生の日は、ほぼ理解できなかったが、非常に丁寧な資料と講義で分かりやすかった。」とポジティブな回答がなされていた。

残りの1名(受講者Cとする)は、計3日間(7/25, 7/27, 8/4)のみの断片的な出席状況であったが、いずれも「②ややそう思う」であった。設問2の事前予習の分量は「④やや多い」と回答し、設問3に相当する7月19日の自習日に「⑦受講していない」。また、自由記載の設問24で「解析初心者にとっては、このコマンドを入力することによって何が起きているのがを理解するのが難しかった。解析の原理?概念?について、予習してから望むべきだった。」と書かれていた。

表6の設問12(第3部8月2日の講義)で「⑤そう思わない」とした受講者は、上記A~Cとは異なる人物であった(受講者Dとする)。受講者Dは、設問19の自由記載欄に「第一部の講義はしっかりと予習をする時間がとれたため、講義の内容を理解することができた。しかし**第三部の方はあまり予習をする時間がとれなかったため、初回から内容をよく理解することができず、結局行かなくなってしまった。**」と書かれていた。実際、設問11(8月1日)が「④あまりそう思わない」、設問13, 14(8月3, 4日)が「⑥受講していない」となっていた。表6の設問1は「②ややそう思う」ではあったが、予習をしていない状況で第2部に出ることなく、いきなり第3部の内容を聞いたという流れであるため、全体として正直かつ妥当な(受講者Dによる)感想であろう。

その他、ネガティブな回答の原因を考察する上で有意義だと思われる受講者Eについても触れておく。受講者Eは、表6上の3か所で「④あまりそう思わない」を選択した。設問19では、「去年と比べ、どのように予習をすれば良いのかが分かりずらかったように感じました。いくつかの連絡が来たこと、ホームページでいろいろなところに飛ぶこと、資料をいくつかいただいたことのためかもしれません。もし一つにまとまったなにかの指針があると安心しながら予習ができるかと考えました。」とコメントされていた。また、設問24では、「...今年は同時期に用件(大学内での薬品廃棄業務、後期だけの私が行う講義のための事務手続き)が重なってしまったので、ほとんど参加できなかったことを、残念に、そして申し訳なく思っております...」とコメントされていた。実際、**受講者Eの出席状況は、出席簿上は3日間のみであり、しかも断片的なものであった。**具体的には、第1部の初日(7月20日; ②ややそう思う)、第2部の最終日(7月28日; ④あまりそう思わない)、そして第3部の初日(8月1日; ④あまりそう思わない)であった。アンケート上は、8月3日にも出席したことになっており、「④あまりそう思わない」となっていた。さらに、受講者Eは、世話人が把握しているだけでも一日を通して出

席していたわけでもない。本講習会は、制度上は部分受講を認めているが、少なくとも同一部内は連続して出席することを前提とした講義を行っている。つまり、昨年度に引き続いて目指した「項目間の連携強化」は、受講者 E のような断片的にしか出席できない状況は想定していない。この意味において、受講者 E の講義内容の理解度に関する「④あまりそう思わない」という感想は至極当然であり極めて妥当である。

ただ、アンケート自体は、第一義的には講義内容の理解度に関する結果が講師側の責任によるものかどうかを見極めるためのものである。本講習会では、アンケート回答者を 1 項目でも出席したものとしたが、「開催側が予め指定した予習を誠実に実施し、関連のある前日までの講義内容を適切に理解したうえで、少なくとも受講当日は一日を通して聴講した講義のみについて答える」ように変更したほうがいいかもしれない。今後の検討課題を浮き彫りにした受講者 E に感謝したい。

#### 3-2-4. 第 3 部の受講人数減少や理解度低下について (表 6 の設問 11-14)

第 3 部は、受講希望者数自体が他に比べて相対的に少なかったこともあるが、全体的に受講希望者数と実受講人数の乖離が大きかったといえる。これは大きく以下に述べる 2 つの理由が考えられる：

1. 年度初めはやる気に満ちて講義に参加するが、徐々にやる気レベルが下がっていくという、大学でよく見られる一般的な経時変化
2. 講習会後半の予習まで辿り着けなかったため、予習事項を理解済みという前提で行われた講義についていけずに脱落というパターン

特に、理由 2 に該当する受講者が多かったのではないかと推測している。実際、受講者 D に代表されるように表 6 の第 3 部内のアンケート結果で徐々に理解度が低下 (①→⑤の方向) し、第 3 部後半で脱落というパターンが散見された。しかし、「4-3. 講習会受講対象、受講定員、受講条件など」の受講必須条件として「講師指定の事前予習を実施していただける方」を掲げており、予習のための資料は充分事前に公開した。にも関わらず実際には予習しきれなかったという理由であった場合は、**第一義的には受講者側の意識レベルの改善以外にはない**だろう。表 6 の設問 1 で「⑤そう思わない」と回答した受講者 A~C の主張として、「事前予習を行ったが内容を十分に理解しきれなかっただけ」だということは想定される。しかし本講習会の規模やアンケート結果を踏まえると、**予習を適切に行った圧倒的多数派の満足度のさらなる向上に多くのエフォートを費やすべき**であろう。

前述のように、第 3 部はさらにスキルアップを目指したい中上級者向けとして実施したため、“初心者を含めた”全体としての満足度向上はそもそも目的としていない。本講習会の第 1~3 部の部間では、第 3 部の相対的な受講者数は少ないものの、中~上級者向け講習会でこれだけの受講人数規模を誇るのはおそらく本講習会のみである。もちろん理念や正論、そして保身的なディフェンス“のみ”を述べるのは建設的ではない。あくまでも日本全体の研究力向上という大局的な視点に立ち、現実的かつ効率的な講習会運営を心がけるべきであろう。今回改めて再考するきっかけをつくった受講者 D には感謝したい。

#### 3-2-5. 講義方法 (表 8) や講習会の存在意義について (表 9)

講義方法（表 8）に関する結果は、全体的として昨年度よりもポジティブな方向にシフトしていた。説明の仕方や配布資料など全体的な理解度や満足度は向上したといえるだろう。この理由は、項目間の連携強化など昨年好評であった部分を維持しつつ、受講者からのリクエストにできるだけ応えたためだと思われる。具体的には、第 1 部として統計解析の強化、第 2 部として講義内容のアップデート、第 3 部として Linux 環境での R や Java の利用、そしてプログラムのインストール手順やエラーへの対処法などが該当する。

設問 26（講習会の講義資料や講義動画があれば、講習会に参加しなくても良いと思う？）は、今年度新たに設けたものである。アンケート回答者 118 名（⑥の未選択 1 名を除く）のうち、講習会への参加の必要性についてポジティブな回答を示したのは、「④あまりそう思わない」の 45 名と「⑤そう思わない」の 17 名であった（割合は 52.5%）。一方、講義資料や講義動画があれば十分であると思ったのは、「①そう思う」の 23 名と「②ややそう思う」の 10 名であった（割合は 28.0%）。28% という数値の多寡については意見の分かれるところではあるが、**これだけの割合の潜在的な受講者が受講申込みを行うことなく、本講習会の講義資料や講義動画を自習用教材として有効利用している（将来的にも利用する）ことを示唆している。**そして一般に、多くの利用者は（本教材を含む）この種の自習用教材の有用性に関して意見表明を行わない。

一般論として、この種の講習会の主な評価軸は、（受講希望者数を含む）受講人数やアンケートによる満足度の高さであろう。満足度調査自体は、分母がアンケートに答えた受講者数となり割合での議論となる。通常、受講者は知識やスキルを身につけるといふ明確な目的をもって参加しているため、講習会内容の理解度や満足度が低いという評価にはならない。このため、過去や他所の講習会との明確な差別化は困難である。それゆえ、「**受講人数**」は費用対効果の観点から有力な評価指標といえる。しかし、**本講習会はフルオープンな方針（講習会開催前に早々と講義資料を公開し、講習会後に講義動画も公開）である。そして本方針は、一般に「受講人数」が減る方向に作用する。本講習会の運営方針は、講習会自体への積極的な参加を促してもいないため、ある意味自縄自縛である。表 8 の設問 26 で掲げた、本講習会への参加の必要性有無についての質問に対しては多様な反応が寄せられているが、これは本講習会が提供する講義資料や講義動画の存在を否定するものでは全くない。**付録 1~7 の自由記述方式のコメントの多くがポジティブであったことがその証左であろう。恣意的な解釈を排除するため、コメントに関してはこれ以上論評しない。

### 3-3. 運営全般について

本講習会では、全般的にこれまで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・運営については全体的には問題なかったと思われる。想定以上の受講希望者数だったのは、おそらくこれだけの内容を集中的に無料で受講できる講習会は、他には無いためであろう。

#### 3-3-1. 受講人数の正確な把握は不可能

今年度は、出席確認表上の出席者数と実際の受講人数の感覚的な乖離が非常に大きかったため、詳細な分析を行った（表 3-2）。実際に、受講者自身による自己申告のみでも、出席確認表とアンケート結果から得られる出席状況が大きく異なっていた。これはおそらく大規模な本講習会ならではの

の問題である。100名規模のハンズオン講習会は、会場のオープンから講義開始までに30分程度は見込まなければならない。受付を済ませた受講者が講義開始までに講義室を出入りし、休憩時間などの区切りや講義途中での入退室もある。講習会期間の二日目以降の名札着用を促しても、持ってくるのを忘れて、紛失したりなど様々である。出席確認表エリアに受講者がたむろしているため、○をつけるのを断念してそのまま忘れて帰ったりすることも現実にはいるだろう。運営側としても、出席確認要員を配置する余力があればTA要員に回したい。そのため、講義日ごとの出席人数の正確な把握ですら事実上不可能である。

### 3-3-2. 質疑応答時のマイクの不具合は反省すべき

アンケートにも書かれていたように、**質疑応答時のマイクに関しては今年度も実質的にうまく機能しなかった点は反省点として挙げられる（付録6のコメント47, 64, 65）**。昨年度購入した拡声器と同機種のを新たに2台購入して対応したが、ほとんどの場合においてうまく機能しなかった。具体的には、ボタンを押して話始めても拡声機能がONにならずに、結局講義室に付属しているマイクに持ち替えるパターンがほとんどであった。スタッフによる事前・事後の動作確認時はうまく機能するのだが、本番（受講者の質問時）ではうまく機能しないの繰り返しであった。もし来年度も開催されるのであれば、別の機種を購入して試す以外にはないだろう。

### 3-3-3. 講習会の枠組みは重要だが受講人数は減らす方向がよい

一定数の受講者は、JST-NBDCのサイト上で公開している講義資料や講義動画の継続的な提供に期待している。また、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。本講習会においても、世話人担当部分（第1部および3部）は、講義日の約1ヶ月前から公開した（タイプミスなどの差し替えは随時行った）。本講習会では受講希望者のニーズを最大限満たすために受講者の調整等は実施しなかったが、次回以降も同様の方針とする場合、開催側が無理なく対応できる**適正なハンズオン講義の受講人数規模（30名程度）に近づけるためには、今年度採った、講習会自体への積極的な参加は促さない方針を堅持したほうがよい**と思われる。講習会の枠組みのみ維持し、講習会のライブ感を求める受講者や、誰にも邪魔されることがなく一受講者としてじっくり学びたい方に限定して講習会を行う方向性に移行していくのがいいのかもしれない。

### 3-3-4. 約束を守らぬものにはペナルティを与えるべき

都合が付かなくなった場合や公開済みの講義資料のみで充分と判断した場合に受講をキャンセルするに当たって事前連絡するよう再三依頼していたが、36名の不参加者のうち14名からは事前連絡がなかった。運営側としては講習会当日までのある程度のキャンセルを見込みつつも、最大電力需要と同様に全員が参加してしまった場合を想定した準備をしなければならない。**単に講義室の収容人数だけの問題ではなく、電源容量の問題、配布物の印刷、無線LANの安定的な動作への対応など、運営側の負担が大きい**ためである。

事前連絡以外の事柄として、受講必須条件として掲げた「受講者アンケートにご協力いただける方」にチェックを入れて参加した126名のうち、当初のアンケート回答期限まではおろか（13名が

当初の回答期限後に提出)、個人宛メールによる督促を行っても回答しなかった7名も論外である。事前キャンセル連絡のなかったものやアンケート未回答者には、今後の講習会などへの申込を認めないなど何らかのペナルティを検討すべきであろう。

### 3-3-5. アンケート対象者や回答させる項目は改良の余地あり

本講習会のアンケート対象者は、1項目でも出席したものであった。講義内容の理解度に関する結果は、狭義には講師に対する評価であり、広義には開催側に対する評価である。これは「3-2-3. ネガティブな回答結果に対する詳細な分析」の最後のほうの議論と本質的に同じであるが、講習会の有効性を厳密に検証したい場合は、「予習を誠実に実施し全日程フル参加した受講者」と「断片的にしか参加していない受講者」で切り分けるなどしたほうが良いと思われる。アンケート対象者はそのままとしても、「開催側が予め指定した予習を誠実に実施し、関連のある前日までの講義内容を適切に理解したうえで、少なくとも受講当日は一日を通して聴講した講義のみについて答える」ように変更したほうが良いかもしれない。一日を通して聴講した講義の理解度として「④あまりそう思わない」や「⑤そう思わない」という結果であれば、第一義的には開催側の責任とみなしてよいであろう。しかしながら、(項目間の連携を強化した講習会の受講前日にはまるまる参加せず) 当日の午後ふらっとやってきた結果として「④あまりそう思わない」や「⑤そう思わない」のは、感想としては正しいだろうが、第一義的には受講者側の責任である。両者は明確に区別すべきであろう。

### 3-3-6. アンケートの文字数や保存ボタンを改良

上記と関連して、受講者がアンケートボックス内の文字数制限に引っ掛かってしまった旨の記載があった(付録6のコメント48)。保存ボタンの追加など、アンケート回答者の視点に立った改良の余地はまだあると思われる。

### 3-3-7. 講習会の運営方針や理念は正しく理解し納得した上で受講すべし

昨年同様、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。これは、全ての受講者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・納得した受講者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者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者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、それでも東大大学院アグリバイオの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、今年度のアンケート結果で希望として挙げたものについては、世話人のNGS連載記事や、今後JST-NBDCが実施する講習会の枠組みなどでできるかぎり取り上げられるよう善処していきたい。

### 3-3-8. 冷房が寒い

受講者向けのお知らせで予め羽織るものをもってこさせるなど周知はしていたが、少なからぬ人

数が冷房の寒さについてのコメントを寄せていた（付録6のコメント9, 41, 60, 61, 62, 63）。講義室の場所による違いが大きいため、場所の移動で解決するはずである。休憩時間に移動を試みさせるなど周知やリマインドを徹底させたほうがよいだろう。

### 3-4. 今後の NGS に特化した人材育成実施方針について

JST-NBDC が実施主体の本講習会（NGS ハンズオン講習会）は、前身の「NGS 速習コース」の試行実施を含めて3年間にわたって行われてきた。アンケート結果からもわかるように、本講習会の存在は広く知られており、JST-NBDC の講習会サイトもよく利用されているようである。昨年度のアンケート結果では、「統計解析、Linux 環境での R の利用法、そしてプログラムのインストール方法やノウハウ」に関する要望が一定数見出されたため、明確に今年度実施方針として設定することができた。費用対効果の点からも、目指すべきは多数派の要望や需要を満たす効率的な講習会である。しかし世話人としては、一通りの**多数派の要望や需要を満たす**という役割は終えたのではないかという感想を持っている。今年度の内容でとりあえず満足、あるいは事足りているというコメントが少なからず存在したためである（付録3のコメント4, 9, 26, 51, 70; 付録4のコメント5, 7, 9, 27, 30, 31, 50, 53, 64）。また、今年度のアンケート結果（特に付録3と付録4）に書かれていた要望が多岐にわたっていたことも理由として挙げられる。したがって今後は、**多様な少数派の要望**に対して**きめ細かな対応**を行っていくべきかもしれない。例えば、以下で示すような枠組みや取組みを取捨選択して行うことが考えられる：

1. **中規模で短期間の NGS ハンズオン講習会**。3部構成で開催した平成28年度講習会は、計3週間にわたって講義室を確保し、そのうち第1部（7/20-22）と第3部（8/1-4）をアグリバイオインフォマティクスの大学院講義（「先端トピックス」の2科目分）と連動させた。これは、講習会会場の講義室を確実に確保するためであった。逆にいえば、あの講義室を講義とは無関係の講習会という位置づけで長期にわたって確保するのは現実問題として相当厳しい。それゆえ、もし来年度もあの講義室を利用して行うのであれば、今年度と同様に大学院講義と連動する形で世話人（門田）のみを講師として実施する形式にせざるを得ないだろう。例えば日本乳酸菌学会誌の NGS 連載第7回後半～第8回の内容であれば、平成28年度第3部受講者を主な対象とした、継続的かつ発展的な内容になるであろう。これらは、付録3のコメント2, 5, 7, 11, 19, 34, 44, 48, 56, 60、そして、付録4のコメント3, 12, 15, 19, 46, 56, 58, 61の要望を一部満たすものと考えられる。
2. **JST-NBDC のデータベース講習会とアグリバイオ大学院講義の「基礎」科目との連携**。今年度は、アグリバイオインフォマティクスの大学院講義「先端トピックス」の科目として講習会と連携させたが、年度初めにアグリバイオ単独で実施されている「基礎」科目では NGS データベースについても解説を行っている。これらの内容についても、JST-NBDC と連携して効率的により幅広く教えていく方向性は検討に値する。付録3のコメント29, 52の要望を満たす。
3. **遺伝研スパコンの講習会**。Bio-Linux ではサブセットの解析に限定されるため、スキルの向上した受講者向けに開催してもいいかもしれない（JST-NBDC と遺伝研主催）。付録3のコメント20、付録4のコメント16, 18, 23, 39, 49の要望を満たす。アグリバイオは、ノートPCの提供、事前の環境構築、当日のTAでの貢献などが可能と思われる。
4. **初級者向け NGS 講習会**。一定の需要は存在する。今年度は、第2部として世話人が示した枠組みに沿っていただく形でアメリエフ株式会社の講師陣が実施したが、例えばアメリエフ

株式会社は彼ら独自のやり方がある。様々な教え方があり、相性は受講者によって異なる。世話人の方針は基本的に厳しめであるため、例えば、3~4 日程度の連続開催のような形で、予習を課さずに Linux コマンドの基礎から教える比較的やさしめの講習会があってもよいだろう。一般に中上級向けの需要は初心者向けに比べて減少傾向にあるため、NGS 解析入門の間口を広げる意味でも重要であろう。今年度開催の第2部講義資料の有効活用や講師を務めた若手のさらなる講義力向上の意味も込めて、枠組みも含めて自由にやってもらってもいいかもしれない。あるいは今年度講習会の予習事項として掲げた内容そのものを、世話人が一から丁寧に教えるのもいいだろう。

5. **講師公募講習会。**受講者から要望のあった項目をテーマとして掲げ、広く（ハンズオン）講義可能な講師を公募。応募のあったものについてのみ、日程調整をして随時開催。メタゲノム解析、ゲノムワイドアソシエーション解析、多変量解析、モチーフ解析、DNA メチローム解析、C 言語、copy number や融合遺伝子の解析、Hi-C 解析、コマンドラインを使用しない Galaxy などを用いた NGS 解析などが想定される。各所で独自の講習会が開催されているため、費用対教育効果という点ではそれほど大きくはないかもしれないが、広く知られている NGS ハンズオン講習会の枠組みで行うところに意義を見出す受講者もいるかもしれない。
6. **個別指導・共同研究の仲介。**受講者の要望は多岐にわたる。しかしながら、必ずしも需要の絶対数は多くはないため、個別指導や共同研究レベルで事足りる局面も少なからず存在するのではないだろうか。JST-NBDC のサイト上に掲載が許可されたバイオインフォマティクス系研究室や個人をリストアップしておき、JST-NBDC の担当者や世話人らと相談しながら適切な専門家とのマッチングを行うような枠組みがあってもいいかもしれない。
7. **困ったときに気軽に質問できる場の提供。**実はこれが一番重要かもしれない。解析プログラムのインストールで躓いたり、見たこともないエラーに戸惑うことは中~上級者でもよくある。バイオインフォマティクス専門家のウェブページ上で「お気軽にご質問ください」とウェブ上に書かれていても、実際には嫌な顔をされるかもしれないと躊躇する場面もあるだろう。それゆえ、(本当は誰でもいつでもメールなどで問い合わせてもらってよいのだが) NGS ハンズオン講習会受講者向けの公式なフォローアップ相談会を計2週間程度設けてもよいかもしれない。期間をあえて区切ることで「この期間なら世話人などの専門家が“業務として”相談に乗ってくれる」という精神的な安心感も与えられるであろう。あえて「指定した期限内にアンケートに答えた受講者のみ対応する」と明記し、権利と義務について再認識させるのもいいかもしれない。

また、アグリバイオインフォマティクスの業務として行っている、「アグリバイオ事務局内で実習用ノート PC を貸与して自習」という枠組みをそのまま利用して、NGS ハンズオン講習会内容の自習用として開放してもいいかもしれない。講習会の講師に聞くほどではないが講義補助 (TA) に気軽に聞きたいという要望もあるだろう。アグリバイオ側の TA は事務局に常駐しているため、まずは TA に対応してもらい、だめな場合は世話人が対応、それでもだめな場合は6.の個別指導の仲介という3段階の体制は受講者にとっても敷居が低いのではないだろうか。

### 3-5. よりよいハンズオン講義のために

過去の報告書中の記載内容と大卒と同じであるが、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。受講者は、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者の多数派を満足させられるよう、受講者側の気持ちになって最大限の努力をすべきであろう。受講者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）を果たすべきである。講習会にかかる最終的な責任の所在は開催側である以上、開催側が考える常識の範囲、理念や運営方針、現実的な対応方針が優先される。もちろん建設的な提案はぜひ寄せていただきたい。

- **開催側：大容量 USB メモリを複数種類用意。** 事前準備不足の受講者に対応するため。自前 PC を持ち込んで受講する場合、事前準備を行ったつもりで持参した PC に準備不足や不具合が発生する場合がある。例えば、R はインストールしたが必要なパッケージをインストールしていない、間違った ova ファイルをインストールしていた、ネットワークが繋がらないといったケースで、いずれも実際に発生した。USB メモリに実習で必要な R パッケージや ova を準備しておけば、速やかに対応することができる。ova ファイルは計約 15GB 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。また、貸出時には受講番号と名前を確認し、USB メモリの紛失を防ぐことも重要である。
- **開催側：事前予習内容を分かり易く示し徹底させる。** 平成 28 年度講習会では、非常に多くの予習事項を課した。昨年度に引き続いて受講するもの、今年度初めて受講するもの、初級コースのみ受講するものなど受講者の状況・レベル・モチベーションは様々であるため、場合分けを多く行った。しかしながらこの場合分けは、結果として 1 名のみ（受講者 E；[付録 1](#)の改善・検討の余地あり 7）ではあったが不満を招いた。実際、特に、ova ファイル導入手順書の例として示した「**乳酸菌連載第 5 回終了時点の ova ファイル**」をそのまま講習会で利用した受講者がおそらく 10 名程度はいた。実際には第 2 部用・第 3 部用の ova ファイルを導入しない限り、確実に途中で行き詰るため、講義室が一時的にざわつき、TA が一時的に足りない状況に陥った。第一義的には、予習事項をよく読まなかった受講者側の責任ではあるが、開催側としても改善の余地はあるだろう。
- **開催、受講者側：質問関連。** 実は配布資料が間違っていただけというオチだったとしても、受講者は自分だけが間違っているのではという不安感をもつ。受講者の質問も、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。本講習会で行っているように、「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」に切り分けてもらうのがよいだろう。
- **開催、講師側：配布資料やスライド。** 18 pt 以上の大きさをスライドを作成すべし。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。平成 27 年度までは「A4 横、4 アップ、両面印刷」の印刷物を配布したが、平成 28 年度は「A4 縦、8 アップ、両面印刷」の印刷物を配布した。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。受講者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。

- **開催、講師、受講者側：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、受講者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、前方の受講者の頭に隠れてしまい、後方の受講者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者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（つまり後方の受講者が見づらい位置に重要な情報を載せない）。開催側で徹底しきれない部分もあるため、受講者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。平成 28 年度講習会では開催側の徹底もあり、本件に関する不満は全くなかった。
- **開催側：予備の実習用 PC を可能な限り確保。**アンケート（付録 1 のポジティブなコメント 3）にも書かれていたが、OS（Macintosh と Windows）ごとの不具合遭遇率の違いはほぼゼロであり、トラブル自体もほとんどなかったようである。しかし、今年度も持込 PC 以外に開催側の貸与 PC も利用することで、様々な動作環境での挙動を学んでもらうことができた。可能な限りゆとりを持った代替 PC は確保しておくといいだろう。
- **開催、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 と受講者持込 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 の両方で開催側が行っておき、講師にエラー事例を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とうよい。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 ユーザの受講者は Mac の GUI 画面を、そして Mac ユーザの受講者は 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者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つた場合でも、開催側のチェックや修正依頼に対応する余裕が生まれる。直前にいろいろ言われても対処できません。
- **講師側：受講者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。開催側は、受講者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、早めに開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、受講者

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とうい。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。これでハマる受講者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 年度以降は、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。** 受講者の多く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者は一定の割合で存在する。エラーに遭遇した場合、受講者は自身の 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、正しい入力がわからない、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者は、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。ハンズオン講義に慣れていない講師は、通常の講義のように自分のペースのみで実習が進むわけではないということを肝に銘じておくべきである。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者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、受講者側：受講者の PC 環境は多様だということを認識する。** これは OS の違いに気を付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者が現実存在する。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方？」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が起きていると思ったほうがよい。受講者も積極的に 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 19 (上記設問 15~18 について、講義名も含め、内容を具体的にご記入下さい) について、計 67 件の生の記載内容を以下に示す。主催側の主観で、19 件の改善・検討の余地ありのコメントと、48 件のポジティブなコメントに分けた。尚、アンケートでは、「設問 12~15 について...」となっていたがこれは間違いであり、正しくは「設問 15~18 について...」である。この点についてはお詫びして訂正させていただきたい。コメント内容は、基本的に記述内容を(漢字変換ミスも修整せずに)そのまま掲載しているが、改行はできるだけ取り払い、「。」や「:」の挿入のみ適宜行った。

全体にとって有意義だと思われる部分はハイライトさせた。

設問 1 5. 講義内容の分量は適切であった。

設問 1 6. 講義のレベルは適切であった。

設問 1 7. 講義で取り上げられた事柄は興味ある内容であった。

設問 1 8. この講義で学んだことは今後役に立つと思った。

**改善・検討の余地あり 1:** ゲノムアセンブリに関しては、やや分量が多く、RNA-seq, ChIP-seq などは、少なめに感じました。門田先生に、RNA-seq, ChIP-seq の講義をお願いできましたら、大変ありがたいと思いました。

**改善・検討の余地あり 2:** 設問 15→19 と読み替えて回答します。設問 15 に関して、第 2 部の NGS 解析基礎と Reseq の内容は重複する部分が多く、両講義を併せて 1 講義にまとめても良いのではないかと感じました。また、第 1 部と第 3 部の分量に比較して第 2 部の分量が少なかったため、やや物足りなく感じました。設問 16 に関して、それぞれの講義が想定する受講者レベルに合っていたと思います。設問 18 に関して、講義中に生じたトラブルのシューティングはとてもよい勉強になりました。

**改善・検討の余地あり 3:** Linux から R から全て本当に素人でしたが、申し込み時からコツコツ勉強をし始めて、講義内容を理解することができました。自分はヒト検体を扱う人間なので、実際にヒト検体を解析するとしたらどれくらい時間がかかるのか、どれくらいの規模であれば受託解析にしないといけないものなのか、といった実情も知りたかったです。(ヒトの解析であれば現実にはほぼ全て受託解析にするべきなのかもしれませんが) また、例えば研究費が取れたとして、解析環境を整えたいと思ったら何を参考にすれば良いかといったことも興味がありました。

**改善・検討の余地あり 4:** 第二部については解析パイプラインの話が中心であり、もう少しトラブルシューティングを含めた作業が欲しいと感じた。第三部の最後が第一部に繋がるのは非常によい構成だと思ったが、第一部のみ受講した人には教育効果が薄いと感じた。ハンズオンの趣旨と離れるのは重々承知だが、1 日だけでも各社代表的な NGS の動作の仕組みやライブラリーの作成法まで紹介できれば wet の話まで繋がって面白くなるのではと思う。

**改善・検討の余地あり 5:** 講義の分量とレベルについては、限られた期間で初心者でもついていけ

るプログラムとしてはこれぐらいが最大限ではないかと感じました。門田先生のご専門が RNA-seq ということで、トランスクリプトーム解析の内容が主でしたが、私自身は DNA を用いた変異解析をメインに行っているため、個人的にはゲノム解析についてももう少し詳しく学べればありがたいと思いました。ただ、これまで R も Linux も触れたことがなかった者としては、これらを実際に動かしてある程度慣れることができたというのは非常に重要な経験だと思います。講義中で教えていただいた Tips など今後役立つと思います。

**改善・検討の余地あり 6:** 今回の講習はパッケージ化されたソフトを使うのではなく、自分でデータを操作し、解析する玄人向けの講座と感じました。ややマニアックな印象がありましたが、運営の都合上、講習会のレベルを下げられないのかもしれないかもしれません。

**改善・検討の余地あり 7:** 去年と比べ、どのように予習をすれば良いのかが分かりずらかったように感じました。いくつかの連絡が来たこと、ホームページでいろいろなところに飛ぶこと、資料をいくつかいただいたことのためかもしれません。もし一つにまとまったなにかの指針があると安心して予習ができるかと考えました。

**改善・検討の余地あり 8:** 第1週の「統計解析」の講義は全般的に短めでしたが、私が de novo アッセムブリー未経験で k-mer や FDR の検証の意味を理解するのが難しく、消化にかなり時間がかかったため、時間に余裕があって助かりました。第2週の「NGS 解析初?中級」は、新しいソフトウェアを知れた点では大変良かったのですが、特に最終日は無駄に講義が伸びた印象です。第3週の「NGS 解析中?上級」は、より実践的な RNA-Seq や de novo アッセムブリーを取り扱っていたので、データ解析を始めたばかりの人にはかなり参考になると思います。内容的には第1週と連続して行った方がより良いかと思います。

**改善・検討の余地あり 9:** 全体を通して講義分量としては適切であり、大変興味深いものであった。第3部講義レベルに関して難しい箇所はあったものの、丁寧な説明をしていただき理解することができた。第2部の「7月25日: 第2部 (NGS 解析初~中級) NGS 解析基礎」に関しては、説明が早かったことと、PCのコピペのトラブルがあったことにおいて行かれた部分もあった。

**改善・検討の余地あり 10:** 設問 15: アメリエフさんの講義は全体的にボリュームをもっと増やして、より実践的な解析方法を追加して欲しいです。設問 16: 個人的には、講義内容をレベルアップして欲しいです。より実践的に。設問 17: de novo assembly 関連の講義が非常に興味深かったです。設問 18: データ解析の基礎部分を学ぶことができましたので、今後非常に役立つと思います。

**改善・検討の余地あり 11:** 0 から解析スキルを身に付けたい思いで事前準備からしっかり取り組んだ。Linux 使用経験がなかったが、丁寧な事前準備資料や昨年度までの講義資料に目を通して PC のセットアップや、本年度講習会前までにある程度の知識を得ることができた。本年度講習会資料も練り上げられた内容になっていたかと思うので、全体を通して日々学ぶことがあった。第2週のア

メリエフさんによる講習は、もう少し講習内容の練り上げがされていたらさらに受講側も得るものが多かったのではないかと思う。具体的には、近年の解析手法の動向や各出力ファイル内部の解説等含まれていればさらに踏み込んだ理解に繋がったと思う。

**改善・検討の余地あり 12**：私は日モデル生物の RNA-seq 解析を行っており、初回の統計解析編は大変有意義な講義でありました。2部では扱ったことのない解析も触れることができ、大変参考になりました。ただ、非モデル生物での rna-seq についても触れて欲しいと感じました。

**改善・検討の余地あり 13**：第3部は初心者には分量が多く感じました。

**改善・検討の余地あり 14**：一通り解析がどのようなものか見られたのは良かったが、自分のサンプルを解析する場合のセットアップはできないかもしれない。例えば chip seq のみ3日間、すべてをダウンロードするところからやったほうが身につくような気がします。

**改善・検討の余地あり 15**：アメリエフさんの講義は分かりやすかったが、内容をもう少し盛り込んでほしい。

**改善・検討の余地あり 16**：NGS を今後使用する予定でいたので、先にデータ解析方法を学ぶために参加しましたが、所々で実際に実験を行っていないと理解が難しい箇所がありました。その他については全体的に良かったと思います。

**改善・検討の余地あり 17**：全体で見たら、大変勉強になったと思います。ただ、ときどき、無駄に時間を使っているときがあったのが残念でした。1日の集中力が続く時間はそれほど長くないので、もっと詰め込んでコンパクトにやっても良かったように思う。開催側も準備を要求しているわけなので、準備する側ももっと準備できるんじゃないかと思える瞬間が何度かあった。特に第2部の初日と3日目の講義は時間を多く無駄に使っていて、もっと早くやれたし、もっと中身を充実することもできたと思う。

**改善・検討の余地あり 18**：第1部は、特に TCC を使ってみたいと思いました。第2部は、実際、現場の方の考え方や、awk を用いた command の実例、各ソフトパラメータ設定など、具体的に示されてよかったです。応用したいと思うコマンドが多々ありました。第3部は、使用ソフトが、すでにやや時代遅れになっている (Velvet, Trinity など) というお話でしたので、せっかくインストール法や使用法をマスターしても、実際には使う機会がなさそうなのが残念でした。

**改善・検討の余地あり 19**：講義内容は多かったですが、せっかく 3 週間の時間をとり出席しているため、内容が少ないよりは多いほうが良いと思います。事前に公開されているスライドが非常に丁寧なので、門田先生の仰る通り講義内容については自習でもよいのかもしれませんが、スライドに記載されている以外の、温度感・雰囲気 (コマンドのオプションは把握しておくものではなくそ

の場その場で調べて使う、prerequisite はインストールしてみて error が出たら考える場合もある、etc) がわかり、参考になりました (私の周囲にはバイオインフラマティクスを専門とされている方があまりいないためです)。

**ポジティブなコメント 1:** 講義分量：第二部のみ参加。初心者なので第二部の中で繰り返しが多いのが良かったです。新しいことをどんどん詰め込まれると追いついていけないと思います。講義レベル：第二部は私のような初心者と結構上級の人が混ざっているのでレベルの設定が難しかったです。講義内容：これから行いたいと思っている実験の範囲を網羅してあって助かりました。今後役に立つか：さらに勉強をして使いこなせるようになりたいと思っています。

**ポジティブなコメント 2:** まず、設問は 15 から 18 の間違いでは？ 15 から 18 と考えてお答えします。今回、自分が行っている RNAseq について、頭ではわかっていたものの、実際どのようにコマンドラインで操作するのか、敷居が高く感じていたが、25、27日の講習でわかりやすく説明していただき、疑問が解けてよかった。早速解析がスタートできた。

**ポジティブなコメント 3:** 医学部の臨床系の大学院生、Mac ユーザーです。今回初参加でした。統合 TV で偶然存在を知り、それ以降ぽつぽつと動画を視聴して臨みました。過去と比べて今回は Mac がトラブルに陥ることがほとんどなく全体にスムーズでした。全体の満足度も高いですが、一番は先生方の講義資料の準備が素晴らしいという点に尽きると思います。講義に出ずウェブ資料だけで自学でもかなりの知識が得られると思いました。その上で初日に仰っていた「なぜわざわざ講習会に来るのか？」について、実際に受講してみて感じたメリットをいくつか回答したいと思います。・実際に来ないと雑用に邪魔されまとまった時間を確保するのが難しい。・トラブルシューティングを学べる。直接質問できる。・周りの他の受講生からのコメント、質疑応答も興味深い。・講師の小話がおもしろい。一方でかなり時間がとられるのも事実で、同じ内容なら複数回参加する意義は乏しいと感じました。次回開催時は内容をよく吟味し、自分に関係する興味があるコマだけの参加になると思います。それから、一定量の予習を必須としそれを前提に実践的な内容にしているのは共感もてました。

**ポジティブなコメント 4:** 第3部全体としてだが、使用するソフトの①ダウンロード、②インストール手順の確認作業、③使い方の探し方、④実践、⑤実践でわざとコケる、⑥コケたときの対処法、⑦まとめ (ソフトの使用感や評価など) という流れで進んでいたように思うが、②③⑤⑥のステップが自分にとっては特に有用だった。自分のような素人が一人で解析を行おうと思ったときにどこかしらで躓くのは間違いなく、その場合どう解決するかを示していただいたことで、講習内容の理解が深まっただけでなく、自分のデータを取り扱うときにどうするべきかというイメージを常に持ち続けることができた。

**ポジティブなコメント 5:** 内容のわかりやすさも素晴らしかったが、コマンドや少しの工夫でできるような tips がちりばめられていてかなりタメになった。

**ポジティブなコメント 6:** 三週目の中上級者向けの講義はアドバンスな内容であったが非常によく理解できかつすぐに使える技術ばかりでよかった。今後はこのレベルの講義の需要が増えてくるとおもいます。

**ポジティブなコメント 7:** 今まで、NGS 解析はした事はありませんでしたが、春からのアグリバイオからの講義の流れで受講することにより、内容の理解や、ハンズオンの対応も可能だったと思います。また、第2部に関しては、データ解析の一連の流れが実践できたことは非常に良かったと思います。

**ポジティブなコメント 8:** 上記15?19について、、、ということでしょうか。私は RNA-seq を行う予定であるため、受講しました。Linux は多少予習をしてきたこともあり、当日はとてもよく理解できました。ちょっと講師の方の声が小さかったかな・・・でも許容範囲です。本当はもっと受講したかったのですが、既に予定が入っており、3日間のみでした。3日間のみでも受講をご許可いただけたのは大変ありがたかったです。

**ポジティブなコメント 9:** 参加したどの日程も内容、時間的にも満足しました。講義のレベルですが、第一部では解析をこれから始めたいなと思っている私でもわかるように講義してくださったので、理解することができました。しかし、すでに解析をされている方からの質問やそれに対する返答にはなかなかついていけませんでした。また、第二部のアメリエフさんの講義は、実習でいまから何をやるのか、どのパッケージを使うのかなどははっきりしていてわかりやすかったです。

**ポジティブなコメント 10:** 第2部、第3部で Linux コマンドに関する tips 的な話はだいぶ役に立ちました。特に、パスの通し方は自分もよく理解していなかったところだったので、第2部2日目の窪川さんの丁寧な解説はもはや感動すら覚えました。

**ポジティブなコメント 11:** 第1部: RNA-seq を行ったことがありませんでしたが、mapping 後のデータからの解析方法でしたので、理解しやすかったです。方法論の比較もあり、勉強になりました。第2部: NGS のファイル形式など基礎的なことがわかってなかったので、その辺りから説明してもらえて助かりました。Linux は初めてでしたが、Reseq, RNA-seq, ChIP-seq の解析の具体例を教えて頂けたので、全体の流れを良くつかむことができました。私は ChIP-seq の解析を主にする予定なので、ChIP-seq の解析についてもう少し深く教えて頂けたらと思いました。ノート PC 環境なのでデータを小さくして解析する必要があるのはわかりますが、実際のデータで行うとどのくらいのスペックのコンピュータでどれくらいの処理時間がかかるのかについての情報も教えて頂けると良かったと思います。第3部: De novo アセンブリは行ったことがありませんでしたが、いくつかのプログラムを実際に使う充実した内容でした。

**ポジティブなコメント 12:** fastq からのトリミング、QC についての基礎知識が自分の研究上今後必

要となりそうであり、その意味でとても参考になった

**ポジティブなコメント 13:** 以上の設問 15?18 について共通して、第一部の統計解析は、門田先生  
の思想が良く理解できました。そのことは、エンドユーザーとして解析を行う上で現在も、今後も  
役に立つと思います。

**ポジティブなコメント 14:** 様々なパッケージが存在する中でどれを利用すべきかわからないという  
のが、初心者にとってひとつの大きな壁になっていると思います。そのため、複数のパッケージ  
を取り上げて比較したり、新しく入手したパッケージの取り扱いを確認していく過程を授業の中で  
体験できるのは、非常に貴重な機会だったと思いました。

**ポジティブなコメント 15:** 設問 12?14 : 第3部は結局受講しておりません。設問 15 : 1 日中  
パソコンに向かう作業はそう楽なものではなかったですが、今後必要な知識であることを考えると  
全体としてあれくらいの分量になるのはもっともなところなのかと思いました。

**ポジティブなコメント 16:** 仕事があるので参加できない回も多かったです。それでも、受講したも  
のについては理解できましたし、ずいぶんためになったと感謝しております。ありがとうございました。  
また、機会があればぜひ参加したいです。

**ポジティブなコメント 17:** 全く不満はありません。構成も非常に考えられていると思いました。

**ポジティブなコメント 18:** 非常に考えられた構成であったと思いました。この長期間に渡る講習な  
らではだと思いました。アメリエフさんの部分は割合と一般的なところで、門田先生のところは一  
つを深く探るといところで、1部~3部まで全て出ると理解が深まるなと思いました

**ポジティブなコメント 19:** 第2部を受講しましたが、資料・説明が丁寧でとても分かりやすかった  
です。一通りの解析方法について理解することが出来ました。

**ポジティブなコメント 20:** 7月28日 : 第2部 (NGS 解析初~中級) ChIP-seq データ処理に使用さ  
れるソフトウェアなどに関する最近のトレンドや課題点などをよく理解することができました。

**ポジティブなコメント 21:** Linux 環境でのデータ解析 : これまで Linux を触ったことはあったが、  
直ぐにエラーで行き詰まってしまっていた。具体的な事例をハンズオンで経験できたことで Linux  
での不安が軽減され、より使っていけるスキルがついたと思う。

**ポジティブなコメント 22:** 第2部のみ参加。午前中に講義があったため、実習にとっかかりやすく  
て良かった。

**ポジティブなコメント 23** : 難易度をどのレベルにあわせるかが非常に難しい問題かと思いますが、この業界に入って3か月の私にとっては1-3週目ともに程よく頭脳を刺激され、置いていかれるわけでもなく、退屈なわけでもなくちょうどよかったです。

**ポジティブなコメント 24** : 全体のうち、特に、第3部の内容が今後の解析に役立つものであったと思います。第1部での解析の原理・考え方の説明があった後でしたので、第2部、第3部の理解が深まりましたし、これまでなんとなくの理解で使っていたツールの特徴やそれぞれの位置付けが整理できたと思います。

**ポジティブなコメント 25** : きょうみ

**ポジティブなコメント 26** : 現在、所属研究室にてChIPseqとエクソーム解析を行っており、自分たちで解析を行うことを検討していたため、その2つの講義が特に参考になりました。

**ポジティブなコメント 27** : 7月28日の最後に「パスを通す」ことについての説明をしていただいととてもよかったです。普段、GUIばかりで、NGS解析のみにCUIを利用するので、基本的なことが分かっていませんでした。

**ポジティブなコメント 28** : アメリエフさんの授業はとてもわかりやすくよかった

**ポジティブなコメント 29** : 受講する前は、なぜ、毎年受ける人がいるのかと思っていたが、RNAseq解析に関してはTophat(Bowtie)から、HISAT2に移行していくことや、Rを使った解析でも新しい技法の紹介など、ハンズオンを通して、最新の知識や業界の流れのようなものを知ることもこの講習の意義であると感じた。Biolinuxの存在も知らなかったので、ぜひ今後活用してみたいと考えている。

**ポジティブなコメント 30** : NGSを実際に使ったことがなかったので、よくわからないこともあったが、実際に使用する前に受講する機会があつてよかったと思う。2年ぐらい受講すると良いのではないかと思った。

**ポジティブなコメント 31** : 第2部のma-seqは特にわかりやすかった。どのソフトをどの順番で何に使うかが図式化されていたため自分の中で整理がしやすかった。

**ポジティブなコメント 32** : NGSのデータ解析を経験したことがあるが、実験デザインや結果の解釈等についての学びたいと思っていた。今回参加した講習会ではそれらのことが聞けたためとても勉強になった。なによりNGSのデータを100%正しく解析する方法・プログラムは確立されていないこと、そしてそのことを念頭に入れて自分の目的に沿った解析手法を選択しなければいけないということが学べたことは、今後NGSのデータを扱う上で役立つと思う。一方、実際のプログラム

を動かす操作等は、講師の言葉通り講義を受けなくともウェブ資料を利用するだけで十分だと感じた。

ポジティブなコメント 33: 全然イメージがわからなかった解析のイメージがわくようになった。実際に自分で解析する際は環境設定、データサイズなどで手こずりながらやることになるとは思いますが、それは自力でトライアンドエラーでどうにか頑張ろうと思う。

ポジティブなコメント 34: RNA-seq 解析など、以前に行った経験があるものでも知らない事柄がところどころにあり、参考になった。

ポジティブなコメント 35: 予習を行っていたため、どの講義も適切な内容量であると感じた。現在の私の研究では今回の講義で得た知識の一部しか役に立たない。今後役に立つ可能性はある。

ポジティブなコメント 36: 第1部の統計解析(トランスクリプトーム解析)は実際に自分のデータを扱う際に必要となってくると思うので、勉強になった。反復あり/なしの違いや、正規化方法による結果の違いを実際に体験できたのは良かった。また、遺伝子クラスタリングをついやってしまいがちだが、直感的理解とかけ離れることも実感できた。

ポジティブなコメント 37: 第1部: この日に取り扱ったデータが第三週のアセンブリ解析で実際に取得できるデータであったり、K-mer に関して理解が深まった。発現解析の際の統計的な考え方について、丁寧に教えていただき理解が深まった。第2部: 講義内容の分量は少し少なく易しいと感じた。取り上げられた事柄への興味; 午前中の講義部分が非常に興味深かった。講義資料には載せきれてない、実際の NGS 解析動向やデータ取得先などについての解説はありがたい。第3部: 受講人数も少なくなり、より実践的であったのでためになった。解析の最中にアセンブリできたコンティグが0という状況から、原因を特定し、コンティグを伸ばしていくかの過程は、NGS 解析できればすべて解決ではないという良い事例であると自戒。Bridger が動作させるのは結構難しいという解説中にフロアから解決法が提案されていたのが印象的でした。これは、ハンズオン解析の現場で得られた重要な知見。

ポジティブなコメント 38: 7月27日: 第2部(NGS 解析初~中級) RNA-seq について、特に今後使う可能性が高い内容であったので、手法を詳細にお教え頂く事が出来て良かった。

ポジティブなコメント 39: 7月27日(水) RNA-seq については、アノテーションに関する情報を聞きたかったこともあり役に立った。8月3日(水) クラウド環境との連携、ロングリードデータの解析・8月4日(木) トランスクリプトームアセンブリ、発現量推定については、DDBJ pipeline が思った以上に簡単に使え、platanas を用いた de novo assembly を今後使っていく予定である。

ポジティブなコメント 40: ツールのインストールは実際に行ってみることでうまくいかないことが

あることを経験し、どのような方法で解決するか学ぶことができ、非常に今後役に立つと感じた。

**ポジティブなコメント 41**：実際に行っている ChIP-seq 解析や RNA-seq 解析についての講義が特に役立ちました。

**ポジティブなコメント 42**：全てにおいて、講義の資料、説明ともに丁寧に作っていただき理解しやすかった。

**ポジティブなコメント 43**：細かいコマンドの説明まで懇切丁寧にしてくださり、自分のような超初心者には大変ありがたかったです。おかげさまで大変有意義な講習となりました。

**ポジティブなコメント 44**：特定の講義だけではなく、全体の講義に関するのですが、QCの手順をしっかりと指導されていることが大変良かったと思います。実験生物学関連の方は（10 数年前の私もそうですが）、定量的なデータの取り扱い経験が少ない場合も多く、「データの品質」の概念自体がない場合も多いかと思います。当然、前処理や QC 無しでの解析、ごみデータの解析は時間と労力の無駄になりますので、QC 手順を念入りにご指導くださったことに大変感謝いたしております。また、私個人は治験のクリニカルデータマネジャー（DM）として、統計解析部門に渡す前のデータの精査をしておりましたので、逆に「公のデータベースに格納されたデータにも前処理が必要なものがある」という事実をあまり意識できておらず、しっかりとご教示頂き、感謝いたしております。

**ポジティブなコメント 45**：第一部の講義はしっかりと予習をする時間がとれたため、講義の内容を理解することができた。しかし第三部の方はあまり予習をする時間がとれなかったため、初回から内容をよく理解することができず、結局行かなくなってしまった。しっかりと予習をすれば講義のレベルは適切であると感じた。

**ポジティブなコメント 46**：特に第 1 部は、NGS での解析が初心者であったため、言葉の定義を含めて講義があり、大変有意義であった。NGS に関する書籍では、なかなか触れられないような感覚的なことであつたり、思考過程であつたりが講義されて、思考過程を学ぶことができた。同じ手法を用いて解析することが可能となるような講義内容であった。

**ポジティブなコメント 47**：全ての講義に参加しました。

**ポジティブなコメント 48**：特に統計の部分や BioLinux は教わって良かったです。今後役に立つと感じました。

## 付録 2

設問 24：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。計 75 件の生の記載内容を以下に示す。主催側の主観で、38 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。尚、「特になし」のみだったコメントはどちらにも含めていない。全体にとって有意義だと思われる部分はハイライトさせた。

**改善・検討の余地あり 1：**25日と27日の先生は声が小さめで聞こえづらく、特に25日は喋りも速かったので、しかも25日はこちらがまだ慣れていなかったのも、ついていくのが大変でした。講義資料はそれぞれのコマンドやオプションの説明がたくさん載っていて良かったです。

**改善・検討の余地あり 2：**お忙しかったようですが、アメリエフさんの講義資料にタイプミスが多かったこと、若干マイクの音量が小さく感じられたのが改良点ではないかと思います。

**改善・検討の余地あり 3：**門田先生の講義はスタイルが一貫してついていきやすかった。特につまづきポイントが事前に押さえられていて、優しさを感じた。また、データをどう解釈するか、門田先生なりの解釈思考過程も参考になった。色々試してみても眺めるという姿勢も重要だと再認識した。アメリエフの方は、山口氏や三沢氏の講義はテンポよくわかりやすかった。一方初日や3日目の講師の先生はあまり慣れていらっしやらなかったり、受講生がどこでつまづいているかを汲み取って頂けなかったりで、ややテンポが悪かったのが残念であった。改善に期待したい。

**改善・検討の余地あり 4：**門田先生の講義は、大変わかりやすかったです。アメリエフさんの講義一日目（7月25日）は、早すぎて、ついていけませんでした。TAの方々には、大変お世話になりました。

**改善・検討の余地あり 5：**分かり易かったところ：講義スライドとコピペコマンド掲載ページがほぼ完ぺきにリンクしているところ、講義中の質問と答えを受講者全員で共有できるようにしていたところ。分かりにくかったところ：仕方ないことはわかっているので不満というわけではありませんが、1，3部と2部で若干スタイルが違うので少し戸惑いました（すぐ慣れましたが）。

**改善・検討の余地あり 6：**音声小さい時は聞きづらくわかりにくく感じた。

**改善・検討の余地あり 7：** $n=1$  で  $q$ -value を計算させるところが、必要性の有無を含めてわかりにくかった。

**改善・検討の余地あり 8：**全体的に分かりやすい講義であったと思います。ただ、第二部3日目の講師の方があまり質問にちゃんと答えられておらず、回答がうやむやになっていたのが気になりました。

**改善・検討の余地あり 9:** 第1部: Rを使ったマイクロレイ解析の講義を以前に聞いていたので、RNA-seqのデータ解析についてもすんなり理解することができました。第2部: 講義資料のコマンドに間違いがあると、ついていくのにいっぱいいっぱいになってしまいました。お忙しい中資料を準備して下さったのは分かりますが、初心者には結構きつかったと思います。第3部: 単に解析の方法論だけではなく、解析をするにあたっての考え方のプロセスも教えてもらえたので、分かりやすかったです。

**改善・検討の余地あり 10:** 統計解析と系統樹作成についてわかりにくかった。実際に何の違いが判るのだろうかと思った。k-mer 出現頻度解析は実際にゲノムの倍数性やサイズ予測では何を読み取ればいいのかわかりずらかった。

**改善・検討の余地あり 11:** ・全ての日の午前講義およびゲノム Reseqの回の担当の先生の講義は非常に分かりやすかったです。・NGS解析基礎の回の先生の声が最後まで小さくて聞き取りづらかったです。内容は分かりやすかったです。・RNA-seqの回の先生の講義は正の流れがあまり良くなくて分かりにくかったです。

**改善・検討の余地あり 12:** 第一部統計解析では、1日の最後まで講義内容を受講すると、納得できて勉強になったと感じていました。その反面として、1日の途中過程では先のゴールが見えない状況でした。第2部の講義内容は、解析をしたことのない人向けには、一通りの流れがわかりやすかったと思います。一方で、物足りなく感じる側面もありました。物足りないと思った具体的な理由は、実習で行った解析のゴールが、「一通りの流れを体験する」ということで、実際に知りたいことをNGSデータから抽出するというようになっていたように感じたからだと思います。

**改善・検討の余地あり 13:** 講義は全体的に分かりやすかったと思いますが、第二部が先にあり、第三部、第一部だとより理解しやすいと思いました。

**改善・検討の余地あり 14:** 門田先生の授業は、レジュメも充実していて、難しい内容ものすごく丁寧に教えてくださっていたと思います。第2部のRNA-seqは研究に直接関わるところで一番興味を持っておりましたが、レジュメと実際に打つべき内容が違うところがあり、勉強し始めて日の浅い私にとってはついていくのが大変でした。

**改善・検討の余地あり 15:** 分かりやすかったところ・・・予習が全てだと思います。予習で分かりにくかったところで、プログラムのインストールがよく理解できた。分かりにくかったところ・・・ナンセンスかと思いますが、あえて挙げるとすると、実際の解析でのオプションのデフォルトが知りたいなと思いました。おそらく、データの種類や解析の目的に依存するのだろうと思います。ただ、その理論を知りたかったです。

**改善・検討の余地あり 16:** 分かり易かったところ: 今何をやっているかがスライドの番号と、コ

PIPE用のソースコードに表示された番号とでリンクしていたため置いていかれることがなかった。分かりにくかったところ。門田先生の思考回路を事細かにトレースしている内容となっていたため、主目的を達成するまでに様々な寄り道があつて本来の目的を忘れかけてしまうことがあつた。この点については今の目的は～でしたよねと確認があつたが混乱してしまっている人もいたのではないかと思う。

**改善・検討の余地あり 17 :** 門田先生の講義は、途中何度か確認が入り分かりやすかつたです。一方第2週は実技担当者が慣れていないせいか、声が聞き取りにくかつたり説明が走つたり、質疑応答がかみ合わないこともありました。時間配分は、全般的に前半が間伸びし、後半走りがちだったので、皆元気な前半は早く進めて、疲労してくる夕方以降はゆっくりと進めると良いと思います。TAに関しては、持ち込みPCだったのでそれほどお世話にならずに済みましたが、講義内容に直接関係ない質問（私の場合はパスを通す方法の選択基準）も答えてもらえ、とても助かりました。

**改善・検討の余地あり 18 :** 門田先生が実施される講義は、非常に分かり易いです。講義資料の作り込みから、時間配分、休憩のタイミング、解析目的の再確認のタイミング等、大変満足しております。アメリエフさんの講義は、昨年に比べ非常に分かりやすくなっていました。特に講義資料の作り込みが改善されていて助かりました。NGS 解析基礎(7月25日)、RNA-seq(7月27日)は、プレゼン方法を改善するとともに分かりやすい講義になるのではないかと感じました。

**改善・検討の余地あり 19 :** 全体を通して、ゆっくりと進めすぎではないかと感じた。あれだけ予習をしてくるようアナウンスしていたので、それでもついてこれない人は自力で努力するべきだと思う。周りの人の様子を見ている、しっかりとできる人は時間を持て余しているように見えた。

**改善・検討の余地あり 20 :** お世辞でなく門田先生の講義は非常に分かりやすく、素人がどう躓いて、何が分かっていないかを把握してもらっているなあと感じました。try and error を中心に小ネタも混ぜていただき、汎用性が高くなるよう意識されていて、大変勉強になりました。アメリエフの方々も丁寧な講義で申し分なかつたのですが、今後の質向上のためにあえて提言させていただくとすれば、午前中の講義において もう少し図示や具体例表示を増やして頂ければ、分かりやすいかなあと感じました。具体例を書いたら字数制限にひっかかつてしまったので省略します。

**改善・検討の余地あり 21 :** それぞれのツールの実際の使い方、インストール方法などに時間がよく割かれていて、実際の解析で一番つまづくところでしたので、わかりやすいご説明でよかつたです。分かりにくかつたところは、第2部のRNAseqの回で議論があつた、GFF ファイルをマッピング時に与えてマッピング結果を遺伝子ごとのカウントデータとして得るか、発現定量時に与えるかという、2つのそれぞれの戦略の利点と欠点が不明瞭でした。GFF ファイルをマッピング後に与える場合の、カウントデータの取得の仕方を説明していただきたかつたです。

**改善・検討の余地あり 22 :** 配布資料は見やすかつたのですが、もう少し補足説明があつたほうが

より理解が進むとおもいました。しかし、情報量が多くなると説明も大変でしょうから、なんとも言えないですが。

**改善・検討の余地あり 23**：門田先生の講義は非常にわかりやすかったですが、アメリエフの講師は声が小さかったりわかりにくかった講師がいたように思いました。

**改善・検討の余地あり 24**：講義もテキストも全体的にわかりやすく工夫されており、大変わかりやすく感じました。しかし、7月25日 NGS 解析基礎と7月27日 RNA-seq の午後の実習講義の中で、講師の方自身が混乱していた部分が見受けられ、その箇所で言いたいことがうまく伝わってこず、少しわかりにくいように感じました。

**改善・検討の余地あり 25**：早口な先生の日は、ほぼ理解できなかったが、非常に丁寧な資料と講義で分かりやすかった。

**改善・検討の余地あり 26**：講義の進度はハンズオンで問題が起きた時のためにかなり余裕をもっておられたかと思えます。ただ、講義がだれてしまいますので、もう少し、詰め込んで4日間→3日間程度にできるような気がしました。

**改善・検討の余地あり 27**：普段 R を使わないため、R のコマンドやオプションの話になるとなんとなくしか理解できなかった。

**改善・検討の余地あり 28**：コマンドの説明がプログラム言語にかなり慣れている人向けのように思われるのでもう少し詳しい説明がほしい。

**改善・検討の余地あり 29**：第二部の講師の方、特に NGS 解析基礎と RNA-seq の解析の講義が理解しにくかった。講師の方の進め方によって講義の分かりやすさに差が生まれていた（仕方のないことだと考える）。

**改善・検討の余地あり 30**：生データ取得の際に SRA に取りに行ったり、DRA に取りに行ったりしていた。講義中でも口頭で説明していただいたのだが、それぞれの長所短所があるのであれば詳しく教えていただきたいかった。RNA-seq のデータセットについて。かなり大部分のリードが酵母の rDNA に張り付いていました。IGV で眺めて、ほぼすべてのゲノム領域ではマッピングリードがなく寂しかった。最初から1万行の設定だったため？もしくは全リードをマッピングした結果を、三分クッキング的にあとから眺めてみるでもよかったかも。

**改善・検討の余地あり 31**：R と統計解析を行ってきたので、ついていけるだろうと思いましたが、慣れ親しんだ数値データと講義で扱われた配列データでは異なるところも多く、データの構造がうまくイメージできなかったところもごさいます。今後、「統計や R を行ってきた人間であっても、

配列データを扱った経験がなければ経験者以上の予習が必要である」等の注意喚起を最初にされると、受講生の予習や準備がより入念なものになるかもしれません。

**改善・検討の余地あり 32**：第一部に関してですが、ゲノム解析、塩基配列解析では k-mer 解析とゲノムサイズ推定の結びつきが分かりづらいように感じました。トランスクリプトーム解析では TCC と edgeR で DEGs 検出数の違いが何に起因するのかを説明してほしかったのと、RPM 補正の部分は不要だと感じました。3 群間比較のマトリックス行列に関しての説明は非常にわかりやすく、応用できそうだと思います。

**改善・検討の余地あり 33**：プログラムに関しては既に用意されているものを用いるので、どのプログラムを使えば欲しい結果が返ってくるのかはわかりやすかった。一方、プログラムに関して関数をまとめたものがあれば、自分で新たに組む際に参考にしやすくなると思う。

**改善・検討の余地あり 34**：第1部と第3部に関しては、講義内容と資料が非常にリンクしており、講義中に理解を深めることができた。第2部は当初は少し流れが理解しづらかったり、専門用語があったりと初心者ではすぐに理解しづらい点があったが、講義の進行に従って改善があり、またフォローアップの追加講義があり非常にわかりやすかった。

**改善・検討の余地あり 35**：アメリエフさんの若手講師は、経験不足から来るプレゼン能力の低さが際立ったことから、内容の把握に少し困ったこともあった。

**改善・検討の余地あり 36**：・核心部分になるとすっ飛ばす傾向は気のせいでしょうか？教える分量が多く大変かと思いますが、ペース配分は改善の余地ありかと。・基本的なコマンドや設定を次の日には忘れていて、うまくいかず TA に聞いたりしていました。基本的なコマンドや設定は常時黒板に張り出しておくともよいかもしれません。・講師の先生、TA の方とも、質問には大変丁寧に答えてくださいました。感謝です。

**改善・検討の余地あり 37**：全体的に、なぜそのコマンドを使うのか、その結果としてどんなファイルができるのか、をもう少し解説してほしかった。

**改善・検討の余地あり 38**：解析初心者にとっては、このコマンドを入力することによって何が起きているのを理解するのが難しかった。解析の原理？概念？について、予習してから望むべきだった。

**ポジティブなコメント 1**：1 部のみを受講です。反復なしの比較のところは、講義資料のみではよくわかりませんでしたが、講義の説明を聞いてよく理解できました。（群間でばらつきを見積もっているということ）

**ポジティブなコメント 2:** 第1部は、いずれも門田先生の専門分野ということもあって、よく伝わる内容でした。特に TCC の実際の使い方と TCC がどのようなことをやっているのかという点について、書籍を読んだときよりも理解できました。第2部は、いずれも入門編としてよく練られていて、それぞれ講師の方は、説得力があり、みなさま、魅力的でした。第3部は、どのように考えながらデータを見ていくかという点で、重要な研修と理解しました。

**ポジティブなコメント 3:** 講義資料はいずれも丁寧に作りこまれていたのでわかりやすかった。自分では対応しきれないエラーなどが生じることもあるため、TA は必要であると思う。例えば私がお借りしたアグリバイオの貸与 PC は一度ビジー状態に陥ってエラーを吐き出したが、これを参加者全員が自らの力で原因を特定して対処できるかは疑問。

**ポジティブなコメント 4:** 全くの初心者のところから、予習、受講の流れでした。初めは PC への抵抗感がありましたが、過去の講義は初心者でもわかる部分から始まっており予習する上でやりやすかったです。予習を含め 1-2 か月くらいついていくと、解析がもうできてしまうのでは！という気持ちになる作りだと思います。実際には色々なところつまづくのだと思いますが、エラーへの対処方法もかなり授業に盛り込まれておりすごく良かったです。インストール方法やブラウザの使い方も説明も良かったです。

**ポジティブなコメント 5:** 普段からインフォマをしている人間からみて全体的によく練られていてわかりやすい講義だったとおもいます。

**ポジティブなコメント 6:** 第3部に関しては、申し込みをしておりましたが内容が難しかったため、最終的には欠席させていただきました。すいません。

**ポジティブなコメント 7:** Pacbio の存在の意義は多分よく理解できていないと自分で思います。

**ポジティブなコメント 8:** TA については、私は 1 度も質問しませんでした。マシントラブル発生時を考えると必要な、と思います。

**ポジティブなコメント 9:** 第一部の k-mer 解析がとてもわかりやすかったです。実際に k=1 から増やしていくとどうなるのか、自分のパソコン上で操作して、見て確認できたのでよかったです。第二部では Reseq の講義がわかりやすかったです。今何をしているのか、随時確認してくださり、自分の作業している場所、コマンドなど理解することができました。また、わからなくなったときの TA も親切に教えていただきました。でも、解析は初心者であるため、パソコンがうまく動かないなど一度躓くと講義に追いつくのが大変でした。

**ポジティブなコメント 10:** 毎回、門田先生の資料や説明の仕方が非常に丁寧で、初学者にも理解し

やすかったです。実際の画面のスクリーンキャプチャが資料に示されているのはものすごく助かりました。

**ポジティブなコメント 11:** 講義は2つのスクリーンを用いて、スライドと実際の画面と両方をうつし、かつスライドのプリントアウトを配布していただけるので、非常に受けやすく、わかりやすかったです。

**ポジティブなコメント 12:** 資料が非常にわかりやすくできていたため、ちょっと気持ちが途切れても、すぐに catch up できた。

**ポジティブなコメント 13:** 門田先生の資料が分かりやすかったです。(R で)塩基配列解析のページと対応づけられていて、スクリプトのコピペが可能のため、自分で打ち込む労力が減り、その分内容の理解に注力できたため予習復習をスムーズに進行できました。アメリエフ様の資料も非常に分かりやすかったと思います。

**ポジティブなコメント 14:** 大まかな NGS の解析の流れの中に Linux R の TIPS が散りばめられており、非常にわかり易かった。

**ポジティブなコメント 15:** 門田先生が、去年と異なる授業をなさったのには、度肝を抜かれました。準備がここまでしっかりした講義は、生まれて初めてではないかと思っております。今年は同時期に用件（大学内での薬品廃棄業務、後期だけの私が行う講義のための事務手続き）が重なってしまったので、ほとんど参加できなかったことを、残念に、そして申し訳なく思っております。できましたら来年度もよろしくお願いいたします。

**ポジティブなコメント 16:** 講義内容はもちろんのこと、受講者の方々のレベルが高く、皆さんの質問内容や質問への回答がかなり勉強になりました。

**ポジティブなコメント 17:** どの回でも前半に講義があり、解析の流れやメインストリームとして使用されているソフトウェアについて紹介して頂き、解析の流れを事前に理解できた事が良かった。また実習でもコマンド集が用意されており、講義をスムーズに進めることが出来た点が良かった。

**ポジティブなコメント 18:** 7月28日：第2部（NGS 解析初～中級）ChIP-seq。殆どの説明がわかりやすかった。

**ポジティブなコメント 19:** 7/25 と 7/27 は実習の途中しばしば置いていかれてしまった。休み時間があったので救済してもらえて助かった。7/26 と 7/28 は講師の方がちょくちょくフィードバックをしてくれたおかげで詰まることなく実習を進められたし、講義内容も頭に入る余裕があった。TA の方には何度も助けていただきとても感謝している。

**ポジティブなコメント 20:** 準備頂いていた例題コマンドの解説を十分にしていただけだったので分かり易かった。データの可視化の部分で IGV を使用しましたが、

**ポジティブなコメント 21:** 講師の先生がスライド番号をこまめにおっしゃってくださったり、2つの前方のスクリーンのうち、どこについて説明しているかも非常に気を使って伝わるようにして下さっていたので、先生は疲れると思いますが講義を受ける側にとってはありがたかったです。

**ポジティブなコメント 22:** 講義資料だけで理解できるほどに丁寧で良かった

**ポジティブなコメント 23:** コマンドなど一個一個説明してくれるので頭によくはいる

**ポジティブなコメント 24:** 1部の R のハンズオンに比べて、第2部の NGS 解析のハンズオンの進行が遅かったようにも思う。しかし、それが悪いわけではなく、私のような素人には、そのくらいゆとりがあったほうが、資料を前に戻ったり先に行ったりする時間があり理解度を高められたように思う。おそらく、熟達している人にとっても、そのような時間に色々と進んでやることがあり(そのように質問から感じた)、一意に悪いというわけではないと思った。

**ポジティブなコメント 25:** コピペできるようにコマンドが用意されていて良かったです。エラーが出なければついていけるのですが、ちょっとしたことでエラーが出るとどんどん講義とずれてしまうので、TA さんがいてくれて助かりました

**ポジティブなコメント 26:** 全般的に講義は分かりやすかった。習得には、講義の繰り返しや自分のデータを解析してみるなどの自学習が必要と思う。

**ポジティブなコメント 27:** 最初は資料のスライドの見方がよくわからなかったが、2回ぐらい講義を受けた後はすんなりスライドを理解できるようになった。

**ポジティブなコメント 28:** ChIP-seq 解析については、今年の方が理解しやすかったが、内容は昨年の方が現在の自分の研究の役にたった気がする。

**ポジティブなコメント 29:** amellief さん担当回で、パスの通し方やソフトウェアのインストール方法を毎回丁寧にやって頂いてとても勉強になった。

**ポジティブなコメント 30:** 全体的に特にわかりにくいことはなかった。講師の方も言っていたが、これぐらい説明してくれるなら、集まらなくても web 上で十分理解できた気がする。あとは、気軽に質問できる場がどこかにあれば、わざわざ東京まで行って参加しなくてもいいと思えました。ありがとうございました。

ポジティブなコメント 31：全体を通じて、適切なペースでお教え頂けたと考える。

ポジティブなコメント 32：講義資料が非常に丁寧に作り込まれていて、全体を通して非常にわかりやすかった。特に統計解析の部分で R を使ってデータを様々な方法で眺める手法はわかりやすかった。

ポジティブなコメント 33：門田先生の講義資料は必要十分な情報が載っているので、予習してわからないところを重点的に聴講させていただきました。

ポジティブなコメント 34：第1部に参加できなかったため、第2部の作業に必要な PC 環境を整えることができず、追いつくのが大変でした。しかし、全てにおいて、講義の資料、説明ともに丁寧に作っていただき理解しやすかった。実際の解析を今後始める予定なので、自分のデータが出た際に改めて取り組みたいと思っています。

ポジティブなコメント 35：わかりやすかったところ：門田先生の講習はつまづくポイントを見透かして安心してきいていました。また、久保さん、三澤さんの講習はとても sharp でまわりの受講生の間でも大変好評でした。

ポジティブなコメント 36：配布資料が、大変わかりやすく、自習、復習に役立ちました。

ポジティブなコメント 37：全体を通して非常にわかりやすかった。

### 付録 3

提案 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。計 71 件の生の記載内容を以下に示す。全体にとって有意義だと思われる部分はハイライトさせた。

コメント 1: R のプログラミングに対する理解が足りないため、コピペで全く同じ解析はできても、応用が効かないので、前年度講習会にあった、R によるプログラミングについても少しあれば有用と思います。

コメント 2：関連の解析とその流れや本編と異なる解析ツールが必要な場合はその紹介。ゲノムの de novo のアッセムブリー、バイサルファイトシーケンスの解析など。

コメント 3：NGS 解析の他の様々な講習会を見ても、真核生物を対象に行っていることが多く、原核生物やマイクロバイームなどのパイプラインについて取り扱っていることがほとんどないので、そのような講義を取り入れてほしい。

コメント 4：今後解析を進めていくとさらに疑問が生じるのかもしれませんが、現時点ではほぼ満足です。一つ挙げるとするなら、今回、RNAseq の結果を可視化するやり方の方を多く取り上げておられましたが、我々ウェットの研究者としては、そのような俯瞰的な解析より、得られたカウントデータから実際に変動している遺伝子は何かを見つけ機能解析につなげるの方が目的です。cuffdiff によって出たファイル群の内容の説明をもっと行って、具体的に遺伝子の単離に至る方法の解説が欲しかったです。

コメント 5：・エピゲノム解析（メチローム解析）。・R の作図についてもう少し。実際の論文にあるような、一見簡単には作れなさそうな figure を、R でこういう風に作れますよ、という話をぜひ聞いてみたい。

コメント 6：トランスクリプトーム解析の中に含まれる応用事項と考えられますが、ncRNA の解析に関し、配列長が短い miRNA などのトリミングの注意点など、機会がありましたら、教えていただきたいです。

コメント 7：awk, sed, grep や、条件文などを使った有用なシェルスクリプトや、R のスクリプトの具体例について。ゲノムにトランスクリプトをマップした場合のリード数の数え方(Ht-seq その他 R のパッケージ)に、ソフトごとの差があるかについて。また発現量比較で、biological replicates が多い場合（例. 8 vs 8）、目的により、2 vs 2 の解析を 4 回繰り返して段階的に絞り込んで有意な DEG を抽出していく方法の是非、さらに発現に影響を与える因子が 2 つ以上ある場合のモデル化について。

コメント 8：今回の講習会の内容から更に発展させたものとして、受講者全員が同じ生データを用

いて各々解析し、その手法や結果の交換、トラブル遭遇事例の解決討論などを行う機会が欲しいです。参加者数によっては開催が困難になるため、数名の解析担当者によるパネルディスカッション形式でも十分実りがあるものになるのではないかと思いますがいかがでしょうか。また、パイプラインを追っていくのではなくパイプラインのパーツに注目して、複数の解析ツールを紹介し、それぞれを手元で動かして挙動や結果の比較を行う講義があるとうれしいです。

**コメント 9 :** 今回はトランスクリプトーム解析を学ぶことが主な目的だったので満足できる内容だった。

**コメント 10 :** ChiP seq のさらなる解析(少数波ではあると思うのですが)encode の使い方。無料でなくても全然いいと思います。位置づけるに難しとは思いますが、お金払ってでもまた受講したいです。門田先生は今年で終わりとおっしゃってはいましたが、来年も是非やっていただきたいです。

**コメント 11 :** 第三週のような講義を増やして欲しいとおもいます。逆にいえば第2週の講義内容はネット等で学習できる項目です。

**コメント 12 :** 世の中のアカデミアに属するバイオインフォマティクスの実際の解析環境が、どういう規模のものなのか、どのように構築したら良いのか、知りたいと思いました。また、バイオインフォマティクス系の論文を投稿する場合にレビュワーとのやりとりでよく問題となる点など、解析して論文投稿する流れで出てくる問題などについても知れたらと思いました。

**コメント 13 :** 講義資料どおり行えば出来るか、といわれれば、たぶん習得できる範囲はあると思いますが、実際に実習形式で手を動かしながら講師の話聞く、というのはとても大切で得られるものは多いように思います。自習だと、動画は早送りするところが増えそうな気がします。

**コメント 14 :** 今回の講習会でも、すでに現場で解析をしている方からの質問がありました。講義内容に限定せず、質問に答える(全員で議論する)時間をもう少しとっていただけると、参考になります。

**コメント 15 :** 3次解析まで終わったデータからどのように有益なものを抜き出してくるのか、その実際の方策を聞いてみたい。

**コメント 16 :** がんの体細胞変異検出

**コメント 17 :** 菌層解析 (qiime など)

**コメント 18 :** NGS で RNA-seq を行っている人が多数派だったようなので(自分もそう)、RNA-seq 解析の一連のパイプラインとしてメジャーな tophat-cufflinks と、quasR-TCC などの R 上でカウント

データを得て統計解析へと持って行く流れの比較やそれぞれの利点などを紹介してもらえればと思います。特に後者のカウントデータに関しては、肝心のカウントデータの取得に関する方法についての話があまりなかったように感じたのと、こういったことに関する質問が多くなされていた印象を受けたので、需要はあると思います。

**コメント 19 :** PacBio についてはインターネット上にあまり情報がないように感じますので、今回の講習会では扱わなかった、第 7 回乳酸菌学会原稿の内容は、非常に気に入りしました。(第 7 回の資料を拝読します) また、copy number や融合遺伝子の解析についても、個人的には興味があります。

**コメント 20 :** DDBJ pipeline だけではなく、国立遺伝学研究所のスパコンに直接アクセスしてマッピングなどの NGS 解析をする方法について

**コメント 21 :** パスウェイ解析や、gene set enrichment analysis などの、transcriptome データに対する解析手法について、独学で学んでいますが芳しくありません。NGS で作られた omics データに対する解析なので、NGS 解析という趣旨にはそぐわないかもしれませんが。

**コメント 22 :** 座学のところで、実際の応用例について教えてもらってもよかったかもしれません(シーケンス情報が実際に役立った例は少ないのですが)。現在、こんなプロジェクトが進められているとの紹介でも構いません。

**コメント 23 :** 発表された論文のデータを実際に解析し、再現する。というのは、実践的かついろんな議論ができるように思います。どの論文を取り上げるのかが現実的に難しいとは思いますが。

**コメント 24 :** 時間的に難しいかもしれませんが、シーケンサーで読まれたデータの基本的な処理(Quality check、アセンブリ、マッピング、発現量推定)に加えて、その後の解析、例えばクラスタリングに加えて GO enrichment、主成分分析などがあると、有り難いです。

**コメント 25 :** RNA seq のパスウェイ解析の仕方を教えてほしい

**コメント 26 :** 今の内容に満足しています。

**コメント 27 :** NGS の仕組みそのものについての説明がもう少し聞きたかった。毎回の講義の中で少しでも触れていただけると助かったと思います。

**コメント 28 :** 各種ツールの比較、流行りなどをまとめて教えていただけると嬉しいです。また、インストールに躓きやすいツールについては、その TIPS をおしえていただければ嬉しいです(自分で環境構築を行おうと思っているので)

コメント 29 : 実際に練習をするのに public なデータの選び方は重要かと思います。

コメント 30 : パスウェイ解析、GO 解析、de novo 解析についてさらに詳しく知りたいです。R を使った統計解析をさらに詳しく知りたいと考えております。

コメント 31 : 発現差解析など次のステップを重点的に行ってほしいと思います。FASTQC や trimmomatic など、解析に必要なソフトのインストール法や使い方は分かったので、それらの実用時のトラブルシューティング、オプションコマンドの調整の仕方のコツなどを盛り込んでいただきたいです。

コメント 32 : pajek によるネットワーク解析

コメント 33 : 過去のハンズオンで出た QA 集があれば講義時の TIPS として配布して欲しいと思いました。

コメント 34 : 実際は他の講義を受講すれば習得できるかもしれませんが、シェルスクリプトや R のトレーニングについてもう少し勉強したいと思いました。また、R で描いたグラフや IGV を使って出したピークから、実際に論文にする Figure を作成するまでの基本工程も習えると良いかと思いました。さらに、解析サーバーを立ち上げる際の基本知識なども教わる機会があればよいと思いました。

コメント 35 : 系統樹とドットプロットを講義で取り上げてもらいましたが、R のパッケージを使用して他にどんな図が書けるのか簡単に紹介してもらえると良いと思います。2 群、3 群間比較で、違いの大きい遺伝子や領域を多面的に評価する方法を教えてもらえればと思います。

コメント 36 : ・RNA-seq のカウントデータの取得方法。・メタゲノム解析の一連の解析方法。

コメント 37 : それぞれの解析手法を細かく説明していく前に、ウェットの実験をどうやって行うか概説した方が良いかと思う。ドライでの解析で、何故その作業が必要か、不必要かの理解が進むと思う。

コメント 38 : バイオインフォ系論文以外の生物・医歯薬系既報論文にある生データからパスウェイ解析、ヒートマップ、PCA、ヴァイオリンプロット等までの図の作成までの解説。

コメント 39 : メタゲノム解析

コメント 40 : アセンブラーが内部的に用いている原理の説明をもう少し詳細に取り入れていただき

たいです。アセンブルした配列データを、RNAseq などのマッピング時に参照配列として用いる場合、参照配列の構築時に生じたバブルがどのように処理されているのかがマッピング条件に関わるかと思うからです。また、ヘテロ性の高い生物種を対象にした解析についてももう少し取り入れていただきたいです。具体的には、kmergenie が diploid の配列でもゲノムサイズ推定が可能な理由や、その正当性について踏み込んだお話があるといいと思います。

**コメント 41** : 非モデル生物の RNA-seq

**コメント 42** : 第 1 部の Reseq ではシングルリードでマッピングされていましたが、ペアエンドでの方法がわかりづらかったです。講義ではソフトのオプションを確認し、ペアエンドでもできます、ということでしたが、自分でやってみると、forward と reverse の 2 つのファイルをどのように並べたらよいのか、というところで躓いてしまいました。また、このつまづきによって、まだまだ Linux の常識(スラッシュや改行について)が身につけていないと感じました。そこについても予習しておけばよかったと思いました。

**コメント 43** : 時間がなく、コピペになるのは仕方がないと考えておりますが、できれば実際にコマンドを打って確認する作業も一部取り入れていただけたらと感じました。

**コメント 44** : コマンドやオプションの説明。マニュアルを読んでも意味が良くつかめないものが多いため。非モデル生物様に、アセンブリー後の遺伝子モデルの構築や blast 等によるアノテーション方法。

**コメント 45** : 特になし。

**コメント 46** : メチル化の解析

**コメント 47** : NGS の時系列解析を行う予定があるので、2 群間、3 群間の平均値の比較だけでなく、時系列解析があれば、ぜひ受講したい。

**コメント 48** : プラスミドの解析方法。8 月 3 日の講義資料の、最後のスライド (176 枚目) の内容。

**コメント 49** : Linux の説明は軽くて良いので、perl にもう少し時間割いても良いと思います。

**コメント 50** : 情報処理分野出身のため、ウェットな分野の内容があまりピンと来ていませんでした。受講の前提条件にはなっているのですが、可能であればシーケンサの原理、ペアリード、RNA の発現とその計測方法、CHIP シーケンスの原理等を冒頭で少し説明して頂けると情報処理分野の人間にも優しくなるのではと思います。

コメント 51 : 特にない。おおむね満足。

コメント 52 : 第 1 部の導入で少しお話された公開されている NGS データの入手法について詳しく聞きたかった。(データベース上で目的のデータセットどういうキーワードで検索するか等)

コメント 53 : まだ自身の研究で NGS の使用の実践段階に入っていないので、さらにどのような講義内容が必要なかわからない。

コメント 54 : 一細胞 RNA-Sequence の統計解析

コメント 55 : 第一部の統計解析で、例えば edgeR はスクリプト内部でどのような処理を行っているか、正規化にはどのような種類のものがあり具体的にどのような方法であるかなど、詳細な説明をより多く盛り込んでほしいと思いました。

コメント 56 : (最後の二日間の講義に出席していないため扱っていたかもしれないが) PacBio と illumina シーケンサーを組み合わせたマッピング、アセンブリについての講義。

コメント 57 : 今回の講義ではクオリティコントロールやアダプター除去などの前処理段階にそれなりの比重が置かれていたと思うので、ゲノム解析や発現解析において、実際にどのようなデータセットでどのような問題に直面し、それを解決したかというケーススタディのようなものが学べるといいです。

コメント 58 : ChIP-seq や DNase-seq のようなエピジェネティクス解析によりフォーカスしてほしい。(特に、モチーフ解析やピークの位置の分布などの出し方)

コメント 59 : もっと非モデル生物を扱う農学系の生物に特化して、ゲノムサイズがある程度大きい非モデル生物等の参照ゲノム配列が不完全なものの場合の、そのゲノム配列を利用した発現量定量解析 (RAN-seq) など。非モデル生物での近縁参照ゲノム配列を利用したアセンブリ方法など。

コメント 60 : denovo アセンブリの内容をさらに強化してほしいです。例えば、微生物ゲノムのコンティグを整形して一本につなげていく作業などです(ウェブ資料としては公開されているのは把握しております)。DNA メチル化解析についても取り入れていただけたらいいなと思いました。今年 ChIP-seq の授業もあったので、エピゲノム解析の次候補としていかがでしょうか? 解析に要求されるメモリ数などが大きくなるかもしれませんが。

コメント 61 : ma-seq について、よく使われるパイプライン或いは計算ソフトの原理とアルゴリズムの説明を入れて欲しいです。

コメント 62 : perl を用いて、fasta ファイルの整形やエラーが起きた配列の除去や確認など、細かい実例を見てみたい。ゲノム未解読生物種を扱うことを前提とした解析方法を聞いてみたい。

コメント 63 : 可能であれば、解析の原理や概念についての講義を取り入れてほしいです。

コメント 64 : 実際の RNA-seq や CHIP-seq を行う際の解析の流れや、解析ツールをどのように使い分けたらよいか講義していただけると嬉しい。

コメント 65 : ・実際の NGS 解析では、biolinux を使用しているバイオインフォマティシャンは少ないと思いますが、実際の解析環境や方法についての紹介。・コマンドラインを使用しない Galaxy を用いた ChIP-seq 解析など

コメント 66 : 講習最終日にでも、個人の NGS データを解析することができれば、ぜひ参加したい。

コメント 67 : 昨年度の講習ビデオを予習内容に組み込んでもよいかも？と思いました。なかなか予習が増えて大変かもしれませんが・・・

コメント 68 : 1) 配列データの構造のイメージをつけるためのトレーニング 1)については作業を行いながら身に付けていくしかないとも思いますが、なにかトレーニングカリキュラム等を開発されると、より社会的に価値ある講習会の展開ができるのではないかと思います。

コメント 69 : TCGA などのデータベースなどを用いて、複数のデータの遺伝子発現をグループ間で比較し、有意差のあるものを抽出するなど、マーカーとなる遺伝子をビックデータから抽出する過程を具体的に学んでみたい。

コメント 70 : 特にありません。今年度の講義内容はかなり充実していたと思います。

コメント 71 : NGS のメーカーによる違いとか、NGS だけでなくマイクロアレイや様々な分子生物学的手法をどう使い分けるか、といった話。R のプログラム解説。エラーを正しく理解し対処する方法。パラメーターをどう修正すればきちんとプログラムが動くか? とか。

## 付録 4

提案 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。計 69 件の記載内容を以下に示す。全体にとって有意義だと思われる部分はハイライトさせた。

コメント 1：実験者（がん、免疫）なので、研究を進めるに当たり、自分の実験データを公開 DB などを利用した human サンプルでの解析結果と比較することによって、次の実験計画を立てることが当面の目的なので、基本的な解析スキルが身に付けられれば十分です。但し、自分の解析結果の検証は難しいと思いました。

コメント 2：上記 R のプログラミングを少し勉強できたらと思います。

コメント 3：Bio-Linux を使いこなせるようになりたいです。できれば R も挑戦したいです。

コメント 4：・シェルスクリプトや perl・python 言語などを身につけて、NGS 解析後に出てくる膨大なデータを自由に解析できるようになりたい。・NGS 解析から出てきた膨大なデータの可視化方法を身につけたい

コメント 5：現時点では特にはないです。

コメント 6：今回の講習では、当然ながらすでにそれでうまく行くことが分かっているデータを用意してもらって解析したので、つまづくことなくついていけた。しかし実際の自分たちの研究では、検証されていない生のデータを扱う必要があるので、どこまでうまくいくかはやってみないとわからない。つまり、今回の講習で、聞いて理解できた気ではいるが、実際に自分でも出来そうと言えるレベルではまだない。幸い、様々なトラブルシューティングを伝授して頂いたのと講義資料やウェブサイトが充実しているので、苦戦しながら解決していきたいと思う。

コメント 7：現時点では、本講習会の内容をまずは身に着けたいと思います。将来的には、Hi-C 解析ができるようになりたいと思います。

コメント 8：python, perl など、昨年度の講習であったと思いますが、塩基配列の扱いのための、インタープリター型言語についての入門。R 言語のスクリプトに関するスキルを向上させることを望みます。それらのスクリプトを見たり、書いたり、動かしたりすることによって、既存の NGS 用ソフトで実行されて得られる結果が、本当に妥当なのかどうか、確かめながら進められるスキルが得られるとよいと思います。

コメント 9：R の概要に関して理解を深めることが出来た。

コメント 10 : 本講習会の後は Linux 等の情報技術に係る基礎知識が必要になると思います。参考書籍はたくさんあるので特に学習で困ることはないと思います。前項と関連しますが、自分の目的に合ったパイプラインの適切な構築や結果を検証する能力 (ツールなどの"引き出し") を身につけたいです。

コメント 11 : 現在、植物の転写因子を研究対象として扱っているので、関連した ChIP-seq を習得したいと思う。これまではウェットな方面のハードルが高いと思っていたが、講習会ではそのハードルがかなり下がってきているとのことなので、ぜひ挑戦したいと思う。

コメント 12 : シェルスクリプトを使いこなす方法。Linux 環境の構築

コメント 13 : 配列解析に特化したスクリプトの書き方のスキル。

コメント 14 : とりあえず R を練習中なので、目的のデータの発現変動解析を自力でできるようにするつもりです。

コメント 15 : BioLinux で出来ることを増やしていきたいと、今回参加して初めて思いました。

コメント 16 : de novo assembly を実際のデータで、スパコンを用いて実行したいと考えています。

コメント 17 : ChIP-seq

コメント 18 : R を用いた作図。スパコンなど大型計算機を用いた解析

コメント 19 : どの解析ツールを使えばよいか身につけたい。

コメント 20 : 菌層解析 (qiime など)

コメント 21 : 生データから発現量を得るところまでのパイプラインはだいぶ理解が深まったので、次はそのデータを用いた統計処理(からの DEG 抽出)の勉強をしなければ、と思っています。統計解析に関しての話は第 1 部でもありましたが、やはりどうしてもコードのコピペに終始してしまい、統計そのものの知識や原理を理解できていないため、結局何をやっているのかよくわからない状態から脱することができていないというのが自分の現状です。

コメント 22 : copy number や融合遺伝子の解析について

コメント 23 : DDBJ pipeline だけではなく、国立遺伝学研究所のスパコンに直接アクセスしてマッピングなどの NGS 解析

**コメント 24:** ヒトの癌組織から採取した DNA, RNA に関する解析データを扱いはじめました。癌組織は実際には正常細胞のコンタミが必ず含まれるため、copy number をもとに腫瘍細胞混合率を解析するソフトが開発されています (ABSOLUTE など) それらのソフトウェアを自分で使えるようになりたいと考えています。

**コメント 25:** 自分はあまりデータを扱ったことがないので、次のステップに行くレベルではありません。R や Linux の学習に時間を費やせないで、パッケージされたソフトで実際に作業を行ってから、R や Linux での解析に臨み、それから考えたいと思います。

**コメント 26:** 有意差のあるデータを抽出する方法の概論を学ばせて頂いたので、可能であればそれらのデータの中で本当に究明する意義のありそうなものを更に見分ける手法について学びたいです。

**コメント 27:** 講習で学んだスキルを研究現場で実際に使うということが最優先事項です。次のステップについては、研究現場で実際に使っている中で見えてくると思いますので、現在はありません。

**コメント 28:** 手元にある RNA-seq データから Biological insight を見出す解析、さらに解析結果を論文などでプレゼンテーションするための figure 作成技術を身につけたいと考えています。

**コメント 29:** がん細胞の体細胞変異の call の仕方を教えてほしい。

**コメント 30:** まずは習ったことを自分のデータで出来るように、自分のパソコンで再現しているところです。その後の解析スキルについては、今私と同じ分野の研究をしている方たちの先行研究などを参考にしていますが、具体的にこれというのはまだわかりません。

**コメント 31:** まずは、自分の Linux で環境を構築し、自分の分野で今回の講習会の内容を再現しようと思います。

**コメント 32:** データの評価を出来るようになりたいです。つまり、発現値を得た後に、クラスタリングなり、PCA なり、DEG を出したりすると思います。一つのデータで発現値を出した後、そこまでやっていただけると嬉しいです。

**コメント 33:** パスウェイ解析、GO 解析についてもっと詳しく知りたいです。

**コメント 34:** オープンソースでの pathway 解析について

**コメント 35:** C 言語。別講義で良いので、R のコピペだけではなく、R の基本的な使い方について

(宣言、初期化、など) 学べる授業があると嬉しいです。

**コメント 36:** 一通りの解析作業を実施し、必要な計算リソースの見積を自分でおこなえるようになりたいです。

**コメント 37:** スパコンを使う際のシェルスクリプト、バッチ処理や R、モチーフ解析、DNA メチローム解析、Single Cell RNA-seq のデータ処理等。

**コメント 38:** Linux を用いた解析に技術的に不安を感じているため、複数の公知データを用いてさらに Linux 活用技術を向上したい、また副次的に必要なになってくる Parl も学びたい。

**コメント 39:** ラボでスパコンを用いて解析を行っており、スパコン特有のコマンドを打ち込む必要があると言われた。そのためスパコンを動かすためのコマンドを学んでいる。解析したいデータや解析手法によってフリーソフトを使い分ける、との話だったが、その使い分け方法を知る必要があると思う。

**コメント 40:** メタゲノム解析の解析スキルを、今後学びたいと考えております。

**コメント 41:** 本講習会で紹介されていないパイプラインでの解析は、自主的に身に付けていこうと思う。また、データベースの活用についても色々と試しながら学んでいきたい。

**コメント 42:** シェルスクリプト作成からデータ解析までの半自動化するためのスキル

**コメント 43:** 公開データをテストデータとし、論文とは異なるアプローチで解析するスキルを身につけたいです。

**コメント 44:** ゲノム・エピゲノム・トランスクリプトームデータを統合して、生物学的な情報を得るための統計学。

**コメント 45:** 真核生物ゲノムの de novo アセンブリとアノテーション

**コメント 46:** Linux 環境で、ある列について sort して行を抽出するといった、Linux コマンドを自在に使うスキルを身につけたいです。perl や shell スクリプトの基本的な知識から、Blast 検索の結果をアセンブルで構築した参照配列の description 行に入れ込むというような、応用スキルを身につけたいと思います。

**コメント 47:** 中長期的には、いずれ細菌叢のメタゲノム解析やエピジェネティクスについても解析できるようになりたいと思います。フリーソフトで可能かは見当もつかないのですが…。

コメント 48 : perl や python による解析スクリプトの作成

コメント 49 : スーパーコンピューターを利用した解析。

コメント 50 : 特になし。

コメント 51 : メチル化の解析

コメント 52 : 今回の講習会では、すべて用意されたデータで、コマンドもコピペで行うことができたが、実際に自分のデータで、今回習得したコマンドをカスタマイズして使えるようになりたい。

コメント 53 : 今学んでいる内容でしばらくは事足りそうです。

コメント 54 : ゲノムワイドアソシエーション解析

コメント 55 : まだ自身の研究で NGS の使用の実践段階に入っていないので、さらにどのような講義内容が必要なのかわからない。

コメント 56 : R による Figure 作製

コメント 57 : 多変量解析のスキルアップをしたいと思っています。

コメント 58 : 新規生物の de novo アッセムブリを行ってリファレンスを構築していきたい。

コメント 59 : 複数の生物が混ざって存在するサンプルを用いた発現解析において、適切なノーマライズや DEG 解析方法を身につけたい。

コメント 60 : R と python と Linux

コメント 61 : ・NGS 解析の結果（変異や発現変動遺伝子など）に関して、図表に落とし込むスキル。  
・SAM ファイルから、自身のほしいリード（例えば unmaped read）のみを抽出する

コメント 62 : 大まかな解析スキルは得られたと思ったので、イレギュラーな条件での対応力を身に付けたい。例えば、パッケージを使わずに GO 解析を行うことや、リシーケンシングで SNP や indel を検出した後にその情報をゲノム情報に反映させ、新しいリファレンスを作製するなど。

コメント 63 : RNA-seq 解析において、新規転写産物を探索するスキルなどを身につけたい。

コメント 64 : まずは実際の解析での実戦経験を積んで行きたいと考えています。

コメント 65 : マウス : 遺伝子改変マウス (KO, TG など)の組織を用いた RNA-seq をもとに、解析  
ヒト : 腫瘍組織を Reseq で解析し、新規変異遺伝子を探索する。できれば、シグナル解析を NGS  
で行い、候補遺伝子を絞り込む、という技が使えると wet な実験にもものすごく有用なのですが、ま  
だ自分にはよくわかりません。

コメント 66 : 講習会の内容を消化するのでまだまだ精一杯ですが、可視化やレポーティングといっ  
た、解析結果を如何にほかの方に分かりやすく伝えられるか、そのための技術の習得をしたいと考  
えています。

コメント 67 : 3 群間比較のマトリックス行列を応用した多群間比較解析

コメント 68 : 提案 1 と同じ (付録 3 : コメント 69 : TCGA などのデータベースなどを用いて、複数  
のデータの遺伝子発現をグループ間で比較し、有意差のあるものを抽出するなど、マーカーとなる  
遺伝子をビックデータから抽出する過程を具体的に学んでみたい。)

コメント 69 : 大量に得られた NGS データのサーバー管理術。

## 付録 5

提案 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。計 43 件の記載内容を以下に示す。全体にとって有意義だと思われる部分はハイライトさせた。

コメント 1 : マイクロアレイの解析やオープンソースの解析ツールの使い方など

コメント 2 : NGS 解析などに特化した、シェルスクリプトの講習会、perl, python などのプログラミング言語の講習会

コメント 3 : 難しそうですが、興味があるのは、ゲノム編集や、共焦点顕微鏡を使ったカルシウムイメージングの定量的解析などでしょうか。

コメント 4 : データベースを構築する講習会など。

コメント 5 : 機械学習の講習会

コメント 6 : NGS データと各種データの紐付け解析

コメント 7 : 公共データベースの利用方法。プロテオミクス・メタボロミクス・マイクロバイオーームデータとのトランスオミクス解析。

コメント 8 : Perl, Python, Shellscrip の基礎 (去年はあったようですが未受講です)

コメント 9 : 提案 1 に書いたように、NGS 解析で得られた網羅的発現データを用いて、機能やパスウェイ、ネットワークについて解析する手法に触れてみたいです。

コメント 10 : NGS 解析に臨むので、精一杯なので、考えが及びません。

コメント 11 : 統計解析の講習会、MATLAB の講習会など

コメント 12 : 統計解析、機械学習などです。機械学習は麻生川先生の講義が素晴らしいと思います。もっと時間を増やして講義いただくわけにはまいりませんか。また、できれば**基本的理論**とハンズオンの中間を埋めるような**講義**はをしていただくことはできませんでしょうか。

コメント 13 : オープンソースでの pathway 解析について

コメント 14 : NGS 以外だと、思いつきません。

コメント 15 : プログラミング

コメント 16 : メタ解析、シミュレーション

コメント 17 : 統計解析の基本 (考え方など)

コメント 18 : Linux のコマンド作成の初心者講座があったら参加してみたいです。

コメント 19 : 中級者向けの Linux の使い方の講習会

コメント 20 : アレイ解析、メチル化、GWAS など

コメント 21 : Python や Perl の講義をしっかりと身につけたい。

コメント 22 : 生物実験に必要な基礎的な統計学

コメント 23 : R, Python による動的レポートおよび論文 Figure 作成法

コメント 24 : 今は思いつきません。NGS 講習会をどんどん進化させていってほしいと思います。

コメント 25 : R で論文によくあるような画像を描画するような講習会があれば参加したい

コメント 26 : コマンドラインを使用しない Galaxy を用いた ChIP-seq 解析などの講習会

コメント 27 : david など web 上で使える基本解析ツールの特性に触れる。ほかに有用なツールの紹介。(wet なひと、特に単科に近いような医学部では、主に内輪の話をもとに使っているので..)

コメント 28 : セキュリティーの問題もあり、難しいかもしれませんが、データベースが置かれているセンターの見学会があるとよいと考えています。「実体験」の記憶は強力に頭に残り、また興味・関心を惹起されればされるほど技術内容の習得が進むと思います。例えば、「京」を実際に見たことがあるのとないのとでは、クラウド経由で使ったとしても、技術習得に差が出てくると思います。

コメント 29 : GO 解析などのアノテーション情報付与の方法と解析結果のまとめ方、NGS トランスクリプトーム解析のアウトラインについて

コメント 30 : R の活用法を集中的に行う講習会

コメント 31 : ゲノム編集。実験部分(ウェット)の講習会。

コメント 32～43：ありません／特になし／とくに思い当たりません。／これは希望ありません。／  
今のところ特にありません。／今のところ思いつきません。 等、12件

## 付録 6

提案 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。計 69 件の記載内容を以下に示す。全体にとって有意義だと思われる部分はハイライトさせた。

コメント 1：商業ベースでないにも関わらず、100 名以上収容の大規模な、且つ中身の濃いセミナーでした。先生方の労力も多大なものがあると思いますので、特に要望はありません。

コメント 2：通常の業務のため、第一部のみ、それも完全には受講できない状況で、門田先生がお勧めされていたように講義資料で自習にしようかと迷いました。しかし、一点は、研究室にいて講座の人達から、家にいると子供から、常に中断されるため、まとまった時間をとって集中して自習することは不可能であるため、受講することを決めました。結果的にはそれに加え、講義資料は非常に詳細に作っていただいているとはいえ、講義資料のみではよくわからなくても、説明を聞くとよく理解できた部分もあり、受講すればそれだけのものは得られると感じました。

コメント 3：設備、人数等は良いと思います。28 日の実習時間のように「遊び」の時間があると心の余裕ができて助かります。すでに置いてあるファイルのダウンロードや使い方の説明も講義の時間にしていただけただけなので実際自分で行う時に困らないと思うのでありがたかったです。

コメント 4：設備や講師 TA その他スタッフの方々の仕事ぶりには非常に満足です。また来年、興味がある内容が追加されれば参加させて頂きたいと思います。

コメント 5：人によって、知識・経験が全く異なるので、いくつかのチェック項目を事前に主催者側でご用意いただき、初級・中級・上級者レベル別にさせていただくのもよいかと思いました。初級の講義を受ければ、事前評価が中級レベルに達していなくても受講できるなど。

コメント 6：TA の方は、遅れた人のサポートをするだけでなく、受講生の質問に対して、別の角度から意見を述べるなど、講義自体に参加していただけると、さらに得るところが大きいと思いました。インターネット環境、電源、コンピュータ、事前の ova ファイル配布など、とても準備が素晴らしく、感服いたしました。ありがとうございました。

コメント 7：適切であった

コメント 8：どちらかという受講生に対してだが、休憩中に質問している方がいたが、どうせなら講義中にしてほしいと思った (実際に質問内容を聞いたわけではないのでなんとも言えないが)。明らかに個人的な質問は TA の方に尋ねることもできるので、そうでない場合はできるだけ講義中に質問して受講生全員で共有することで、満足度はより上がると思った。

コメント 9：場所によってかなり寒かった。途中からは場所の把握ができたので良かったですが門田先生をはじめ、アメリエフの講師の先生方、TA の方々、全ての方々がどんくならない質問にも親切に対応してくださり、本当に感謝の一言です。講義資料の作成はとんでもなく大変だろうと思い、そして3週間もみっちりやっていただいて、本当にありがとうございました。(余談ですが門田先生のスライドの犬やネズミが絶妙な位置に入っていたりして、楽しかったです。)

コメント 10：私は本講習会がきっかけでバイオインフォマティクスの道に入り運良く就職することができました。現場をみていると、ウェットの人間でドライ解析がしたいという、熱心かつ有能なドクターの学生やポストドクが多いですが、教育できる人がまだまだいない状況です。このクオリティでの継続は大変だともいますがわが国の生命科学の根幹を支える人材がきっと多く生まれるとおもいますのでぜひ継続していただきたいとおもいます。私ができることは協力していきたいとおもいます。

コメント 11：受講人数はちょうどいいと思います。

コメント 12：全くございません。素晴らしい講義で感動いたしました。

コメント 13：講義の準備はとても大変だったと思います。誠にありがとうございました。受講人数の制限がなかったことは、非常に助かりました。

コメント 14：この講義のさらに上のレベルの講義を聴いてみたい

コメント 15：アメリエフさん主催の情報交換会は非常に有り難かったです。結局 NGS 解析ってよくわからないですね、というお話を周りの方とただけで終わってしまいました。笑、苦悩を共有できたことだけでも収穫はあったと思っています。

コメント 16：特に問題ありませんでした。お菓子があるのがありがたかったです。

コメント 17：人数は適度だったと思います

コメント 18：受講人数、設備、講義室、TA につきましては問題ございません。お菓子は別に要らなかったと思います。

コメント 19：・講義資料のコマンドの訂正が多過ぎると感じました。いろいろ事情はあるかと思いますが、後で見直す際に面倒ですので講義開始前にスライドを修正して頂いていればより良かったと思います。・午前講義が非常に興味深かったのもっと内容を増やしてもよいかと思います。4日間非常に勉強になりました。このような機会を頂き誠にありがとうございました。

**コメント 20 :** TA の先生方がとても親切に教えてくださったので、つまづいた時にすぐついていけるようになったのは本当に有難かったです。寺田さんや、門田先生、アメリエフの山口さんもわからないところを直接質問させていただいた際に、とても丁寧に教えてくださり、初心者には本当にありがたく感じました。このように至れり尽くせりの講習会を無料で開いてくださって、大変感謝しております。また懇親会では、普段話す機会の無いような業種の人とお話できて、とても刺激を受けました。どうも有難うございました。

**コメント 21 :** (講習会では貸し出し PC を用いていたので)資料のコピペするだけでアウトプットを得られるのは非常に良いと思いますが、予習復習のために自身の PC を用いた際には、(wget コマンドやその他もろもろの formula のインストールなどの)環境設定にかなりの時間を費やしてしまいました。講義に必要な分に関しては導入(設定?)の手引きなどがあれば便利だと思いました。

**コメント 22 :** TA さんのおかげで、基礎的な環境設定で躓いていた時にはなんとかりカバーできました。とても助かりました。

**コメント 23 :** 人数も丁度よく、講義室も快適で全く不満はありませんでした。

**コメント 24 :** 去年の方が TA の方が多く、お伺いしやすかった雰囲気でした。

**コメント 25 :** 設備、人数等は適切であったと思います。

**コメント 26 :** 後方の席だと、マイクが聞きづらいと思いました。スクリーンは二つあって非常に見やすかったです。貸出 PC も非常に助かりました。

**コメント 27 :** 無料の講習会ながら非常に充実した受講環境だったと思う。

**コメント 28 :** TA の方にはとてもお世話になった。講義の途中にもかかわらず、気負いなく質問ができる実習環境だったため、実習についていくために一生懸命で講師の話が頭に入らないという事態がかなり少なかったので助かった。

**コメント 29 :** NGS データの取り扱いについて、これまでは非常に長い時間をかけてグーグル先生に聞いていたことを、今回の講義では効率よく、分かりやすく教えてもらえ、大変有意義でした。また、第3部を受講して、長年バイオインフォマティクスを専門としている方(門田さん)も、失敗したり悩んだりしながら壁を乗り越えていっていることが分かり、インフォマも実験と変わらないんだなと感じました。設備に関しては、マイクの数を増やすともっと質問しやすいと思います。

**コメント 30 :** 講義の環境、運営方法など大変満足しました。お菓子を準備して頂いたお蔭で、休憩時間に栄養補給して、集中して講義に臨めました。関係者の皆様、ありがとうございました。敢え

てお願いするならば・・・、講義のカリキュラムを難易度順（初級編→中級編→上級編）にアレンジして頂けると、遠方からの参加者はスケジュールし易くて助かります。諸事情により実施されていないのはお察しています。

**コメント 31**：ネット環境も十分整っており、席も適度な余裕があったのでとても良かったと思う。TAの方々も迅速に対応してくださっており、とても満足できる講義であった。

**コメント 32**：設備、運営体制について非常に満足しています。

**コメント 33**：結果の解釈の仕方や小話など、講義資料に必ずしも収まりきれない情報を手に入れることが出来たため、実際に講義に参加することは大いに有意義であったと考える。

**コメント 34**：設備は音声やスクリーン、ネット環境なども含めて全く問題を感じませんでした。TAの方々の存在は大変心強かったです、ありがとうございます。個人的には交流会を前半に設定していただいたことが非常に大きかったです。一旦顔見知りになるとその後の講習会の合間やお昼休みなどを通じて、更に交流を深めることができました。特にこの講習は様々な分野のかたが参加しているので、変な話ではありますがライバル関係にはないので、より積極的な情報交換が行えました。

**コメント 35**：講義と講義の間の空いている日などに、貸出パソコンで自習できるとありがたい。

**コメント 36**：大変詳細な講義資料と、前年までの講義を無料で公開して頂き、とてもありがたいです。復習などに利用させていただきます。カバンを床に直接置きたくない人のためにビニール袋があったり、マウスの貸し出しもあり、お菓子まで置いてあり、本当に至れり尽くせりで驚きました。秘書さんがきれいで癒されました。アメリエフの久保さんが質疑応答の内容をツイッターにアップしてくださったのがとてもいいアイデアだなと思いました。

**コメント 37**：今回の講習では全て適切であるように感じました。

**コメント 38**：講義資料がとても充実していて、ありがたかったです。

**コメント 39**：ちょうどよい。実践形式のためTAの方々と準備は大変だと思いますが、上手くオーガナイズされており、スムーズでとてもよいと思いました。

**コメント 40**：昨年も参加させていただいたのですが、今年も参加させていただきまして本当にありがとうございます。全国の勉強会にいくと大抵の人がこちらのセミナーに参加されてるとのことで、よく話題になります。引き続きこの業界の開拓者として率いていって頂ければとおもいます。

コメント 41：教室が非常に寒かった。受講人数は第1部は非常に多いと思ったが、第2・3部ぐらいなら良いのではないかと思う。TAはいて良かった。

コメント 42：スタッフの方々は非常に細かいところにまでいろいろ配慮していただき、快適に受講することができました。ありがとうございました。今後、自力で解析を行っていくこととなりますが、いろいろわからない点が出てくると思います。その際に、この講義のフォローアップとして、講師の先生方への質問をメール等で受け付けていただければ幸いです。

コメント 43：長時間座っているのはつらかったので、こまめに休憩を入れてくださって助かりました。

コメント 44：今回の規模で不都合はなかったと思います。

コメント 45：大変ご丁寧にご対処いただき、誠にありがとうございました。

コメント 46：特に不便なところもなく、快適に受講出来ました。休憩時間にちょっとしたお菓子をつまめて、嬉しかったです。

コメント 47：マイク関係のことを除けば、これ以上要求してはいけないと思うぐらい素晴らしかったと思います。ありがとうございました。

コメント 48：このアンケートは内容が多いので、保存ボタンなどを取り入れていただけると、途中の状態でも気づいたところから埋めていけるので、良いかと思いました。今回はワードファイルに前もって作成したファイルを分割して質問ごとにコピペしました。ワードカウント機能があるとうれしいです（書きすぎて泣く泣く削りました笑）。

コメント 49：TAを独占してしまう方がいらっしゃり、中々声をかけ辛かったため、可能であればTAの方の人数を、もう少し増やして頂ければ幸いです。

コメント 50：全体として、全員の進度に合わせながら非常にわかりやすい講義だったが、個人的には、わからないくらいの進度で追いつくように頭を使っていく方がありがたかった。

コメント 51：TAの方が非常に丁寧に教えてくださったのがとても助かりました。

コメント 52：大変満足しました。門田先生をはじめ関係者の方々には大変お世話になりありがとうございました。

コメント 53：講義内容もスクリーンがよく見え、環境は快適で、リラックスできとてもよかった。

TA の方も十分で、すぐに対応していただける人数だったと思う。本格的に解析を始める前に体験できるので、PC を貸し出してもらえるのもありがたかった。

コメント 54 : 申し分ないです。準備くださった門田先生、アメリエフの方々をはじめとする皆さまに心から感謝いたします。ありがとうございました。あと納税者にも感謝します。

コメント 55 : 「ライブ感」の中で先生の思考の流れや Tips を聴くこと（言い換えると本では学べないことを学ぶ）が重要だと考えていましたので、個人的な意見ですが TA はそれほど必要ないと考えています。講義室中の TA の数を減らし、その分の時給を先生の資料づくりのアシスタントを雇用する費用にされてもよいかもかもしれません。

コメント 56 : 第一部に関しては TA はそこまで人数がいらないと感じました。講義室への行き方がわかりづらかったです。

コメント 57 : 第 2 部は私は自習にしました。27 年度と同じように、今年の第 2 部が特論 2 の講義にあたったら嬉しかったです。

コメント 58 : 講義によっては参加人数が多く、講義室がやや狭く感じられることがあったが、全体としては適切であった。

コメント 59 : 後半仕事の都合で参加できず残念だった。遅い時間のイブニングセミナー、土休日開催などあれば社会人には嬉しい。お盆など長期休暇を取りやすい時期にして頂くのも良いと思う。PC もお借りでき、設備や受講環境は素晴らしいと思う。

コメント 60 : 部屋で寒い座席があるのは私も承知していますが、それにしても寒いので、なんとかならないものかな、、、と思いました。私も農学部職員ですので、働きかけができれば、と思います。

コメント 61 : 席によるのかもしれないですが、毎回冷房がとても寒かったです。

コメント 62 : 寒いです

コメント 63 : 寒かった。

コメント 64 : ちょっとマイクの調子が悪かったようですね。音量の大小があり聞こえづらいところがありました。

コメント 65 : 講習会での、質問時に使うメガホンですが、印象的で良かったのですが、聞き取りにくいので改善を期待します。

コメント **66～69** : 特になし／特にありません。 4件

## 付録7

提案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。計30件の記載内容を以下に示す。全体にとって有意義だと思われる部分はハイライトさせた。

コメント1：オープンソースの解析ツールについて、どういうときにこのツールを用いるのが便利という網羅的なフローチャートのようなものがあると助かります（既にあるのかもしれませんが。）

コメント2：特にはないですが、前回や今回の講義資料はもっと声を大にして（学会での宣伝など）周知してもいいかと思えます。

コメント3：ChiP seq. ATAC seq などの epigenetic の分野は今度多くなってくると思うので、是非そちらもお願いします。

コメント4：皆さんのご尽力は大変すばらしいと思います。

コメント5：これまでの講習会のデータを使わせていただき、本当に助かっています。全ての年度の資料を、まとめてキーワード検索できるようにしていただけると助かります。

コメント6：perl の stand alone で行う EnsemblVEP か R で扱う EnsemblVEP

コメント7：PacBio の解析について。copy number や融合遺伝子の解析について

コメント8：門田先生の公開している情報は大変勉強になっております。その点に対して感謝しております。以上です。

コメント9：すでに十分な情報を提供していただいていると思います。

コメント10：第二週の twitter のハッシュタグのように、Facebook のグループのような最新の話題について軽く聞ける環境があると嬉しいです。

コメント11：データベースの種類やありかについて教えて欲しいです。

コメント12：パスウェイ解析、GO 解析について、詳しく学びたいです。特に、発現解析以外の候補遺伝子を、どのようにすれば入力情報とできるのかについて、知りたいと考えております。

コメント13：オープンソースでの pathway 解析について

コメント 14 : 各解析に必要な計算リソース量などを纏めた資料があれば見てみたいです。

コメント 15 : R や Linux での理解を深めるために、簡単なコマンド集などがあるととても助かる。

コメント 16 : フリー以外の解析ソフトのお勧めなどについて教えて頂ければ幸いです。

コメント 17 : 現在も充実したものがアップされており、これを続けて頂ければ、参加者のみならず、自習している日本の方々に大変役にたつと思います。

コメント 18 : 講義を web で実況生中継とかしてもらえたら遠方より足を運ばなくていいのでうれしい。

コメント 19 : If possible it would be better if you could provide the handouts in English so that it would be easy for foreign students like me.

コメント 20 : 詳しい方なら使いやすいコード、とかでしょうか。ほかの labo で詳しいひとは R でふつうのグラフなんかも書くのがらく、とかいってましたが普通の wet やさんにはそういう発想はないので

コメント 21 : やや時代の流れに逆行しますが、門田先生のような方の「ライブ」という本や論文に乗ってこない情報・状況を体感することで技術や思考過程を強烈に焼き付ける=本を読むだけより効率が良い、というところに講習会の意義があると思います。Web はどこの国や大学もやっていることですので、全く逆に、講習会に参加することでしか得られない内容を充実する方面で考えを進められてはいかがでしょうか？少数意見になるかとは思いますが、ご参考になれば幸いです。

コメント 22 : R などに用いられる (よく使われる) 関数を日本語でまとめたものがあると助かります。

コメント 23 : 大量に得られた NGS データのサーバーやホームページ運用術

コメント 24~30 : 特になし / 特にありません。 / 今のところなし / 思い当たりません。 等 7 件