

東京大学大学院新領域創成科学研究科
メディカル情報生命専攻

平成 28 年度

修士論文

熱帯熱マラリア原虫 *Plasmodium falciparum*
におけるシングルセルトランスクリプトーム解析

2017 年 1 月 27 日提出
指導教員 鈴木 穰 教授

救仁郷 圭祐

目次

| | |
|---|-----|
| 第1章 序論 | 2 |
| 1.1. マラリア | 2 |
| 1.2. 熱帯熱マラリア原虫 <i>Plasmodium falciparum</i> | 4 |
| 1.3. シングルセルトランスクリプトーム解析 | 8 |
| 1.4. 本研究の目的 | 9 |
| 第2章 材料・手法 | 10 |
| 2.1. 熱帯熱マラリア原虫の調製 | 10 |
| 2.2. 1 赤血球 RNA-Seq | 16 |
| 2.3 生育ステージの同定 | 36 |
| 2.4 トランスクリプトーム解析 | 41 |
| 第3章 結果・考察 | 46 |
| 3.1. 実験結果 | 46 |
| 3.2. シーケンスデータの信頼性と再現性..... | 55 |
| 3.3. 生育ステージ分類法の検討 | 59 |
| 3.4. 1 赤血球トランスクリプトーム解析 | 79 |
| 第4章 総括 | 108 |
| 引用文献 | 111 |
| 謝辞 | 122 |

第 1 章 序論

1.1. マラリア

マラリアは世界三大感染症の一つである。全世界で年間約 2.4 億人が発症し、アフリカの小児を中心に毎年約 44 万人の命を奪っている原虫感染症である (WHO Report 2015)。

マラリア原虫は、アピコンプレックス門に属する寄生性原生生物であり、ハマダラカ *Anopheles* を介して中間宿主に感染する。

ハマダラカを介して伝搬することから、ハマダラカが生息する熱帯や亜熱帯地域で多く感染が報告されている(図 1)。

マラリア原虫の中間宿主となるのは、ほ乳類、鳥類、爬虫類である。

ヒトに感染すると、マラリア原虫は血流に乗って肝臓へと到達し、肝細胞に感染する。肝細胞内で大量増殖し、肝細胞を破裂させて血中に流れ出る。血中のマラリア原虫は赤血球へ感染して増殖し、赤血球を破裂させる。再び血中に流れ出たマラリア原虫は赤血球に感染して、増殖、破裂、感染を繰り返す(図 2)。

マラリアは高熱、吐き気、頭痛、悪寒、貧血等の症状を引き起こすが、この内の高熱はマラリア原虫が赤血球に感染する際に引き起こされる (Rouzine and McKenzie 2003)。

マラリアには感染する原虫の種類によって三日熱マラリア、四日熱マラリア、熱帯熱マラリアなどの種類があるが、その中で熱帯熱マラリアは迅速かつ適切な処置を行わないかた場合、短期間で重症化して腎臓や脳障害等の症状を表し、最悪の場合死亡に至る危険性を持つ寄生虫感染症である。熱帯熱マラリアは熱帯熱マラリア原虫 *Plasmodium falciparum* によって引き起こされる。

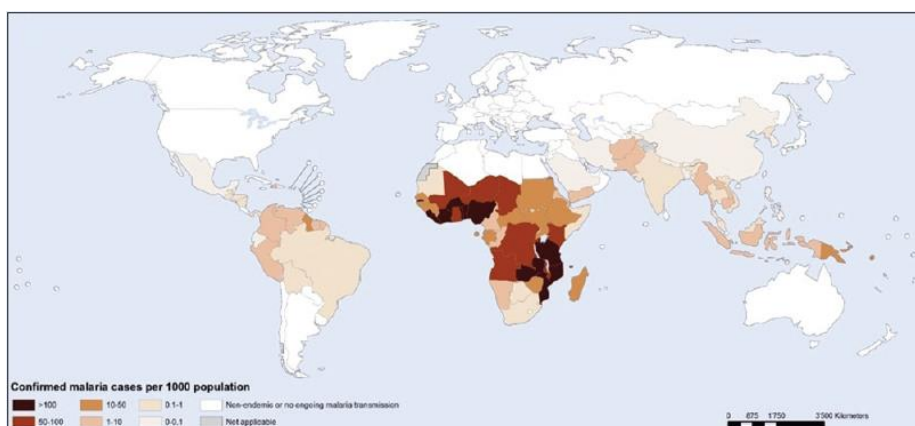


図 1. マラリア発症地域の分布図

(Figure 1, Antony and Parija 2016 *Tropical Parasitology* 6(1): 30–41 doi: 10.4103/2229-5070.175081)

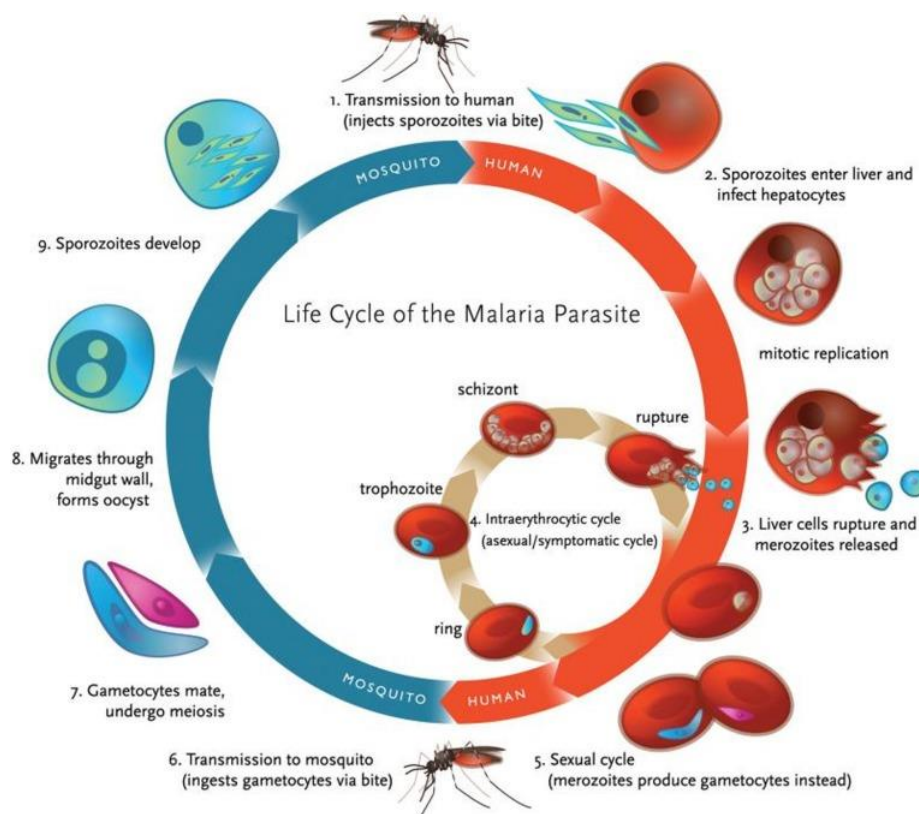


図 2. マラリア原虫の生活環.

(Figure 1, Vale *et. al.* 2014 *frontiers in pharmacology* 5: 275 doi: 10.3389/fphar.2014.00275)

1.2. 熱帯熱マラリア原虫 *Plasmodium falciparum*

ヒトに感染するマラリア原虫は、熱帯熱マラリア原虫 *Plasmodium falciparum*、卵型マラリア原虫 *Plasmodium ovale*、三日熱マラリア原虫 *Plasmodium malariae*、四日熱マラリア原虫 *Plasmodium vivax* の4種である。この内、熱帯熱マラリア原虫は最も病原性が高く、高熱等を引き起こす。

熱帯熱マラリア原虫は、血液ステージでヒトに対する病原性を示す。それぞれのステージで生育速度にはバラつきがあるが、血液ステージの周期は36-48時間とされる。血液ステージにおいて、原虫は一倍体であり、性別を持たない無性世代である。

熱帯熱マラリア原虫について、そのゲノム配列は解読されている (Gardner *et.al.* 2002)。ゲノムサイズは 2.3×10^7 bp、染色体とミトコンドリア、アピコプラストの総遺伝子は5777個、全体のGC%は22.9%である。

近年、熱帯熱マラリア原虫では、抗マラリア薬に対する耐性株が出現しており、大きな問題となっている。

抗マラリア薬は、天然物であるキニーネが最も古い。後年、キニーネは比較的副作用が強いことから、その代替品となるクロロキンが開発された。クロロキンは副作用も少なく、安価であることから世界中で使用されていた。しかし、1990年代から、その耐性株が出現し、耐性株が世界中で拡大したことから現在、多くの地域では、抗マラリア薬としての地位を失っている(図3)。

2000年に入り、新たに見出されたアルテミシニンが強力な抗マラリア効果を有することが見出され、広く使用されるようになった。現在のところほとんどの地域で、マラリアに対しては、アルテミシニン投与を中核とした治療が行われ効果を発揮している。しかし、近年このアルテミシニンに対しても、その耐

性株の出現が方向された(図 4)。タイ-カンボジア国境付近で発生した耐性株はその後、急速な拡散を見せているという。

クロロキン、アルテミシニン耐性株の出現は、マラリア原虫ゲノム中の薬剤耐性関連遺伝に変異が導入されることに起因することが明らかになっている。

マラリア原虫において代表的なクロロキン耐性遺伝子、PFCRT(*Plasmodium falciparum* Chloroquine resistance transporter)は K76T の変異によって薬剤排出効率を向上させ薬剤耐性獲得が起こる(図 5,6)。

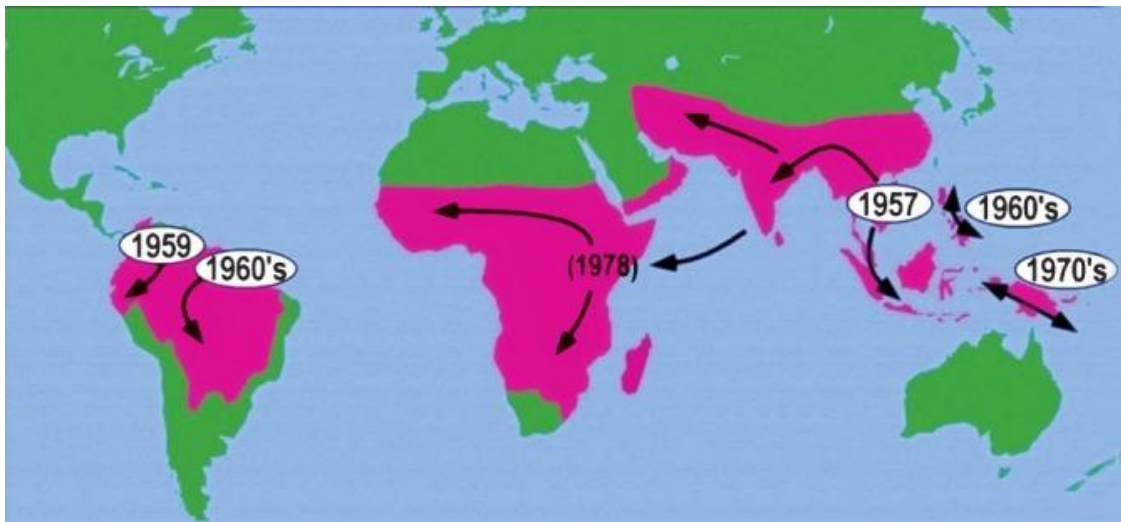


図 3. 熱帯熱マラリアにおけるクロロキン耐性株出現と拡大

タイ-カンボジア国境付近より現れ、世界中に伝播した.

(Figure 6, V. P. Sharma , 2012 , *Indian Journal of Medical Research* 136(6))

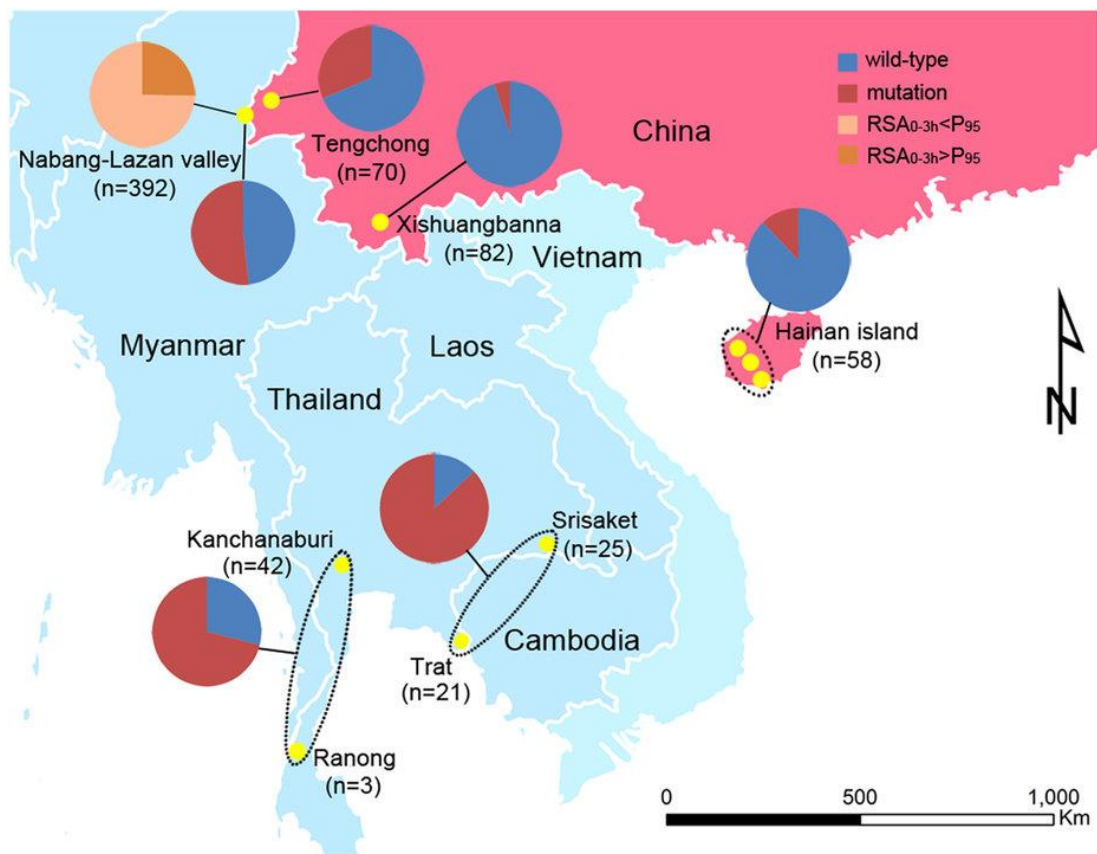
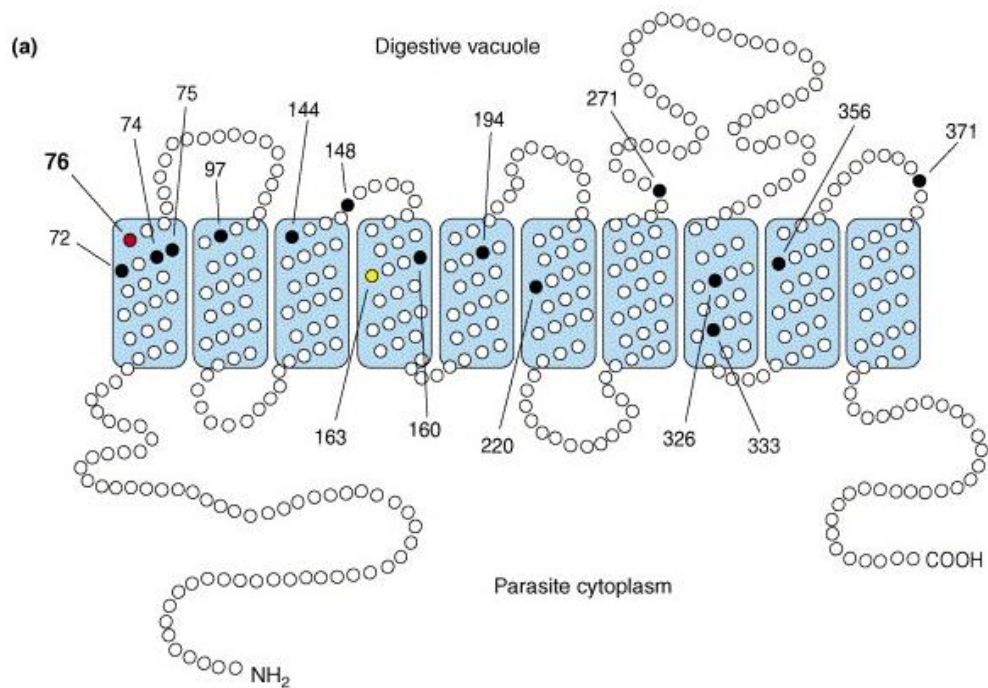


図 4. 東南アジア諸国における *Kelch13* 遺伝子に変異を持つ熱帯熱マラリア原虫の分布

(Figure 1, Run Ye *et. al.* , 2016 , *Scientific Reports* doi:10.1038/srep20100)



(b)

pfcrT allelic variants identified in *Plasmodium falciparum* field isolates and laboratory-adapted lines

| Region | Reference line (origin) | PfCRT position and encoded amino acid | | | | | | | | | | | | | | |
|---------------------------|---------------------------|---------------------------------------|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 72 | 74 | 75 | 76 | 97 | 144 | 148 | 160 | 194 | 220 | 271 | 326 | 333 | 356 | 371 |
| Chloroquine-sensitive | | | | | | | | | | | | | | | | |
| All | Wild type (HB3, Honduras) | C | M | N | K | H | A | L | L | I | A | Q | N | T | I | R |
| Chloroquine-resistant | | | | | | | | | | | | | | | | |
| Asia and Africa | Dd2 (Indochina) | C | I | E | T | H | A | L | L | I | S | E | S | T | T | I |
| Southeast Asia | 734 (Cambodia) | C | I | D | T | H | F | I | L | T | S | E | N | S | I | R |
| Pacific region | 2300 (Indonesian Papua) | C | I | K | T | H | A | L | L | I | S | E | S | T | I | I |
| South America and Pacific | PH2 (Philippines) | S | M | N | T | H | T | L | Y | I | A | Q | D | T | I | R |
| | 7G8 (Brazil) | S | M | N | T | H | A | L | L | I | S | Q | D | T | L | R |

Gray shading indicates residues that differ from the wild-type allele.

図 5. *P. falciparum* chloroquine resistance transporter の予測される構造と代表的な表現型

(Figure 1, Valderramos and Fidock, 2006, *Trends in Pharmacological Sciences* doi: DOI: <http://dx.doi.org/10.1016/j.tips.2006.09.005>)

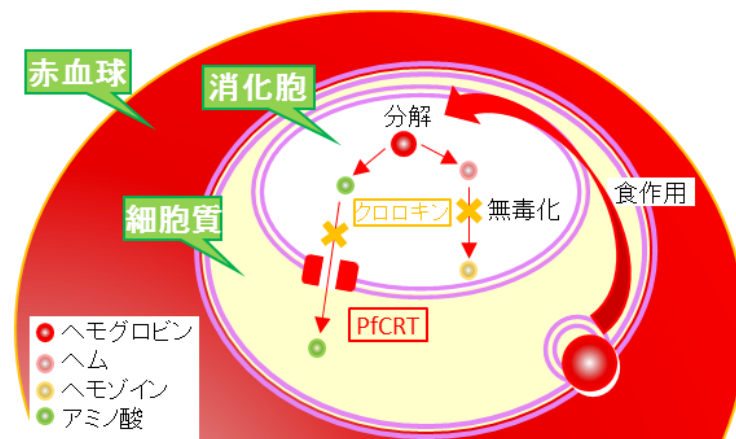


図 6. クロロキンの作用機序

ヘムからヘモゾインへの分解反応を阻害し、マラリア原虫にとって有毒であるヘムを蓄積させる。また、消化胞から細胞質へのアミノ酸の輸送を阻害し、マラリア原虫に餓死を引き起こさせる(Valderramos and Fidock 2006)。

1.3. シングルセルトランスクリプトーム解析

ひろく遺伝子発現解析は、遺伝子自体の機能およびその発現制御様式の解明あるいはその他様々な目的のために行われてきた。2000年以降、多くの生物種でDNAマイクロアレイが開発され、網羅的な遺伝子発現解析、いわゆるトランスクリプトーム解析ができるようになってきている。さらに近年、次世代シーケンサーの登場によって、いわゆるRNA-Seqにより、以前より高い精度でトランスクリプトーム解析が行うことができるようになった (Otto *et.al.* 2010)。

RNASeq解析に用いる鋳型調整法の進展も著しい。特に、SMART-Seqの登場により、1細胞レベルの微量サンプルからのRT-PCRが可能となった。これによりシングルセルトランスクリプトーム解析が可能となっている (Ramsköld *et.al.* 2012)。

シングルセルトランスクリプトーム解析の用途は様々である。ES細胞やiPS細胞においては、その分化効率の検証や、分化系を時系列で観測した際にどの

ようなトランスクリプトームの収束、または発散が起こるかを観測することができる。また、がん細胞のようなヘテロジェナイティーを持つ細胞集団がどのような細胞で構成されているかを検証することで疾患の理解を深める際等、近年、急速にその応用途が拡大している。さらに、1細胞レベルでトランスクリプトームを観測することで、バルク RNA-Seq 以上の解像度のデータを得ることが期待できる。特に、均一であることが保証されない細胞集団に対して、その内在する細胞種の区分、さらには同一細胞種内にも存在するトランスクリプトーム多様性を観測するには不可欠の手法である（後述）。

1.4. 本研究の目的

上述のように、これまでの熱帯熱マラリア原虫の薬剤耐性株の出現では、最終的には薬剤耐性遺伝子にゲノム変異が導入されることによって引き起こされる。しかし、薬剤耐性株の出現には、その前段階としてトランスクリプトーム多様性が寄与している可能性があるのではないかと私は考えた。最も簡単には、トランスクリプトームの多様性によって薬剤耐性関連遺伝子の発現量が高い株が出現、薬剤に対して弱い抵抗を示し、一定の確率である時間まで生き残ることで、その間に変異が導入されることによって耐性株が出現する可能性である。その他にもトランスクリプトーム多様性が薬剤耐性のリザーバーとして寄与する、より複雑な分子機序を仮定することも可能である。

本研究では、マラリア原虫が感染した 1 赤血球 RNA-Seq データの取得を試みた。得られた 1 赤血球 RNA-Seq を基に、トランスクリプトーム解析を行い、トランスクリプトーム多様性と薬剤耐性獲得の関連性について検証した。さらにその多様性が抗マラリア薬の投与でどのように変化するのか解析を試みた。

第2章 材料・方法

2.1. 熱帯熱マラリア原虫の調製

RNA-Seq データ取得に用いる熱帯熱マラリア原虫を *in vitro* で調製した。

2.1.1. 熱帯熱マラリア原虫の培養

クロロキン感受性株である 3D7 株、HB3 株、クロロキン耐性株である 7G8、Dd2 株、K1 株を培養した(これら全ての株はアルテミシニン感受性株)。上記の 5 株は、American Type Culture Collection より入手した。

クライオチューブに保存されていた、冷凍サンプル(-80℃)から細胞シードを培養した。クライオチューブを 37℃ に設定した Water bath にてチューブ内の冷凍サンプルが溶け切らない程度に溶解させ、Water bath から引き上げた。

溶解させたサンプル 500 μ l をクライオチューブから 15ml チューブに移し、12%NaCl 100 μ l(サンプルの 0.2 倍量)を加え、タッピングでよく混和し、室温で 5 分間静置した。1.6% NaCl 5ml(サンプルの 10 倍量)を少し入れて振ってを繰り返しながら分注した。

新しい 15ml チューブに RBC(赤血球) 1ml を移し、RPMI GPF Medium(表 2) 9ml を分注してピペッティングした。サンプル溶液と RBC 溶液それぞれを、24℃、600 \times g、6min で遠心した。

サンプル溶液は 1ml 程残して上清を除き、残った 1ml をピペッティングで混和し、RPMI GPF Medium を入れて 10ml までメスアップした。RBC 溶液も上清を除き、RPMI GPF Medium を入れて 10ml までメスアップした。サンプル溶液と RBC 溶液それぞれを、24℃、600 \times g、6min で遠心した。

RBC 溶液の上清を除き、再度 RPMI GPF Medium を入れて 10ml までメスアップした。RBC 溶液を、24°C、600×g、6min で遠心した。

サンプル溶液と RBC 溶液それぞれの上清を取り除いた。サンプル溶液はマイクロピペッターを用いて、残りが無いように上清を取り除いた。

新しい 15ml チューブに RPMI Complete Medium(表 4) 10ml を加えて、ペレットと同量の RBC を加えてピペティングした。溶液を Cell Culture Flask(50ml)に移し、37°C、CO₂ 濃度 5% でインキュベートした。

細胞シードを起こした際の原因の増殖速度は急激に増加する場合や、その逆に増加しない場合があり安定しなかった。そのため細胞シードを起こし培養してから 24 時間毎に培地交換及びギムザ染色を行って、原虫感染率を確認することで安定的に継代し続けることができた。

表 1. RPMI Basal Medium の組成

| | |
|--|--------|
| RPMI 1640 Medium (Gibco #11875-093) | 500ml |
| HEPES 1M pH7.4 | 12.5ml |
| Gentamicinin 10mg/μl (Sigma #G1272-10ml) | 200μl |

表 2. RPMI GPF Medium の組成

| | |
|------------------------------|---------|
| RPMI Basal Medium | 512.7ml |
| Daigo GF21 (Wako #634-25725) | 50ml |

表 3. Hypoxanthine (10mM)の組成

| | |
|--------------------------------|--------|
| Hypoxanthine (Sigma #H9377-5G) | 54.4mg |
| milliQ | 40ml |
| 2N HCl | 600μl |

※ 上記の組成で調製してスターラーで 1 時間混和した後、フィルター滅菌する

表 4. RPMI Complete Medium の組成

| | |
|---------------------|-------|
| RPMI Gpf Medium | 100ml |
| Hypoxanthine (10mM) | 1.5ml |

※ 使用する直前に Hypoxanthine (10mM)を解凍して調製する

2.1.2. 熱帯熱マラリア原虫の継代

以下の手法で原虫の継代を行った。継代は、72～96 時間に一度行った。これにより、*in vitro* でマラリア原虫を培養する場合にしばし問題となる、培養液中にマラリア原虫由来の成分が蓄積することで原虫が死滅する問題を回避した。

まず初めに、インキュベーターから原虫培養液が入った Cell Culture Flask を取り出し、クリーンベンチに移動させた。この際、Cell Culture Flask の蓋に培養液が付かないよう注意して移動した。Cell Culture Flask の底に溜まっている赤血球を揺すって混和した。Cell Culture Flask 内の培養液を 15ml チューブへ移した。新しい 15ml チューブに RBC 1ml を移し、RPMI GPF Medium 9ml を分注してピペッティングした。培養液と RBC 溶液それぞれを、24℃、600×g、6min で遠心して、培養液と RBC 溶液の上清を除いた。

続いて、ギムザ染色を行って原虫感染率を測定した。上清を除いた培養液をタッピングで少し混和し、4μl をスライドグラスに乗せ、もう一枚のスライドグラスを用いて培養液でスメアを形成した。スライドグラスを素早く乾かし、50ml チューブ内でメタノールに 1 分間浸した。ピンセットを用いてスライドをメタノールから引き上げ、キムワイプに押し付けてよく乾かした。ギムザ染色液をスライド上に乗せて 15 分間静置した(表 5)。

スライド上に直接かけないように気を付けながら、ギムザ染色液を水で洗い流した。キムワイプに押し付けてスライドよく乾かした後でアニソールを乗せ、対物レンズを 100 倍に合わせて光学顕微鏡で赤血球を観察した。1 つの視野の中の RBC と iRBC(原虫感染赤血球)を数えて、感染率を測定した。

培養液と RBC 溶液に RPMI GPF Medium を入れて 10ml までメスアップした。原虫溶液と RBC 溶液それぞれを、24℃、600×g、6min で遠心した。培養

液と RBC 溶液の上清を除き、再び RPMI GPF Medium を入れて 10ml までメスアップした。培養液と RBC 溶液を、24°C、600×g、6min で遠心し、培養液と RBC 溶液の上清を除いた。

新しい 15ml のチューブに RPMI Complete Medium 10ml と、RBC の総量が 200 μ l になるように培養液と RBC 溶液の沈殿を加えて継代した(原虫感染率 5% の培養液を 0.5%にして継代する場合、培養液の沈殿 20 μ l と RBC 溶液の沈殿 180 μ l を加えた)。15ml チューブの培養液を、ピペットを用いて Cell Culture Flask に移し、37°C、CO₂ 濃度 5%でインキュベートした。

原虫感染率を 0.5%になるよう留意して継代することで、安定的に継代を維持することができた。

表 5. ギムザ染色液の組成

| | |
|----------------------------------|-----------|
| 1×PBS | 2~3ml |
| ギムザ染色液 (Merck Millipore #109204) | 0.2~0.3ml |

2.1.3. 熱帯熱マラリア原虫の同調培養

同調培養株は、杏林大学医学部の朝日博子博士から提供を受けた同調培養サンプルを使用した。

マラリア原虫培養株(FCR3株)に対して0.5%D-ソルビトール(Sigma-Aldrich)処理を行い、リングステージ以外の感染赤血球を不活性化し、感染率を0.3~0.5%程度に合わせた。その後、37°C、CO₂濃度5%で培養して46時間後、再度0.5%ソルビトール処理を行った。続けて、63% isopycnic Percoll(GEヘルスケア バイオサイエンス)を用いた密度勾配遠心を行い、リングステージ以外の感染赤血球を取り除いた。その後、37°C、CO₂濃度5%で培養した(Asahi *et.al.* 2011)。

以上の処理をにより得られた、FCR3株(クロロキン感受性株)のリング同調培養株、トロホゾイド同調培養株、シizont同調培養株から1赤血球RNA-Seqデータ取得を試みた。

2.2.1 赤血球 RNA-Seq

1 赤血球レベルで RNA-Seq を行うことで、原虫集団の保持するトランスクリプトームの多様性の存在や、従来のバルクシーケンス以上の解像度を持ったデータを取得できることが期待される(Shalek *et.al.* 2014)。

しかし、マラリア原虫感染赤血球を 1 赤血球レベルで RNA-Seq したデータは、今まで存在しなかった。そのため、本研究ではマラリア原虫感染赤血球の 1 赤血球 RNA-Seq データを取得する実験段階から行った。

また、薬剤添加時にトランスクリプトームはどのように変化するか、多様性が薬剤耐性獲得初期に影響するかについて検証することができると考えた。

本研究では、次世代シーエンサー HiSeq 2500(illumina)を用いた RNA-Seq を行い、熱帯熱マラリア原虫 *Plasmodium falciparums* の 1 赤血球トランスクリプトームを取得した(図 7)。

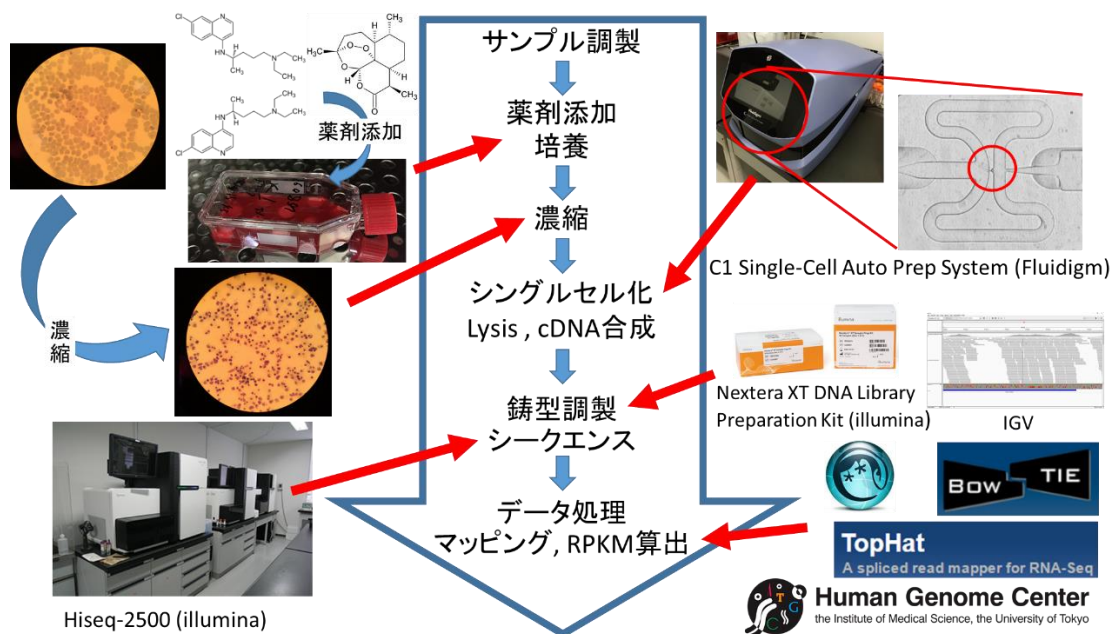


図 7. 実験概要図

2.2.1. 熱帯熱マラリア原虫の調製

本研究では、抗マラリア薬（クロロキン、ジヒドロアルテミシニン）の添加により、原虫トランスクリプトームが1赤血球レベルでどのように変動するかを観測することが目的の一つである。トランスクリプトームの変化を観測するための各実験系の比較には、実験条件を一定に揃える必要がある。

本研究では、原虫感染率5%に調製した培養液に、抗マラリア薬（最終濃度：クロロキン 10nM(Reynolds *et.al.* 2007)、またはジヒドロアルテミシニン 700nM(Witkowski *et.al.* 2013)) を添加して、6 時間後、12 時間後、24 時間後に濃縮を行った。

2.2.2. 熱帯熱マラリア原虫の濃縮

本研究では、C1 Single-cell Auto prep system (Fluidigm)を用いてシングルセル化を行った。実験に使用した C1 System のフローセル C1 IFC(Fluidigm)は、最大 96 細胞までしかキャプチャーすることができず、赤血球の場合では 40-60 細胞程度しか 1 細胞でキャプチャーできない。

熱帯熱マラリア原虫は、*in vitro* で培養する場合、原虫感染率を約 10%までしか上げることができない。原虫感染率 10%の溶液を C1 IFC にロードして、50 細胞がキャプチャーされた場合、原虫感染赤血球はわずか 5 細胞になる。

そのため本研究では、以下の手法で原虫の濃縮を行って原虫感染率を高めて、選択的に原虫感染赤血球を C1 IFC にキャプチャーさせる手法を取った(Mata-Cantero *et.al.* 2014)。

まず初めにインキュベーターから Cell Culture Flask を取り出し、クリーンベンチに移動させた。この際、Cell Culture Flask の蓋に培養液が付かないよう注意して移動した。Cell Culture Flask の底に溜まっている赤血球を揺すって混和させた。培養液を 15ml ファルコンチューブへ移した。原虫培養液を、24℃、600×g、6min で遠心して、培養液の上清を除いた。

上清を除いた培養液でギムザ染色を行った(p13 参照)。

スライド上に直接かけないように気を付けながら、ギムザ染色液を水で洗い流した。キムワイプに押し付けてスライドよく乾かした後でアニソールを乗せ、対物レンズを 100 倍に合わせて光学顕微鏡で赤血球を観察した。1 つの視野の中の RBC と iRBC を数えて、感染率を測定した。

培養液に RPMI GPF Medium を入れて 10ml までメスアップした。培養液を、24℃、600×g、6min で遠心した。培養液の上清を除き、RPMI GPF Medium を入れて 10ml までメスアップした。再び、24℃、600×g、6min で遠心して、培養液の上清を除いた。

MACS Multistand(Miltenyi Biotec)に MACS Separation column(Miltenyi Biotec)をセットした。Separation column の下に受け皿として、50ml チューブをセットし、ウォーターバスにて 37℃に温めた RPMI Basal Medium 4ml を Separation column にアプライした(図 8)。上清を取り除いたマラリア原虫培養液に RPMI Basal Medium(37℃) 4ml を加えてピペッティングし、RPMI Basal Medium が落ち切ったところで Separation column にアプライした。溶液が落ち切った後で、RPMI Basal Medium(37℃) 10ml を Separation column にアプライした。溶液が落ち切った後で、Separation column を MACS Multistand から取り外し、15ml ファルコンチューブ上にセットして RPMI Basal Medium(37℃) 4ml を加えた。カラムを通った溶液を 24℃、600×g、6min で

遠心した。上清を少しだけ残す形で取り除き、ペレット 1 μ l を取り出して、ギムザ染色を行い、濃縮が正しく実行できたかを確認した(図 9)。

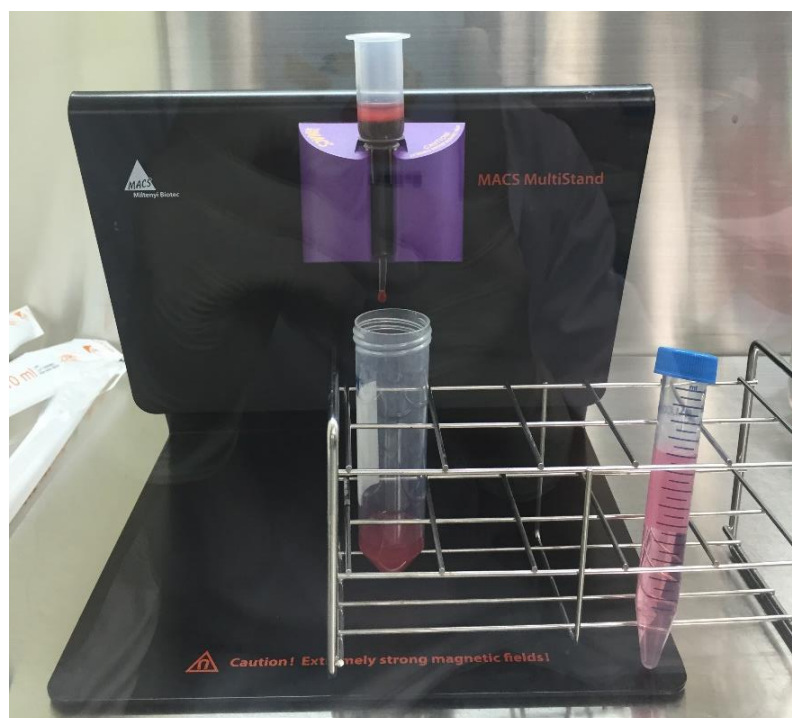


図 8. Separation column を用いたマラリア原虫の濃縮

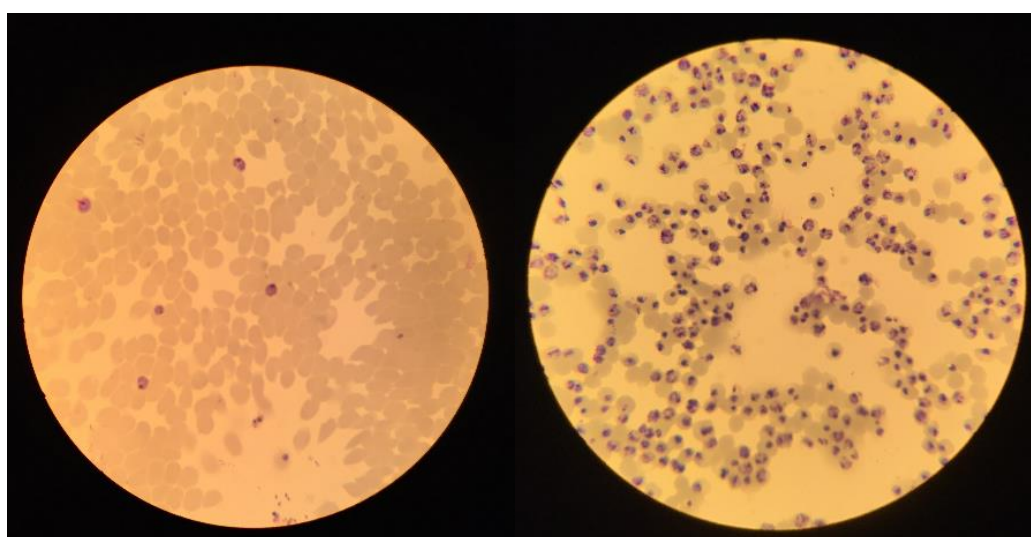


図 9. 正しく濃縮されたマラリア原虫

2.2.3. 原虫感染赤血球のシングルセル化・cDNA 合成

本研究では、C1™ Single-Cell mRNA Seq IFC, 5–10 µm(Fluidigm) を用いて C1 System にてシングルセル化を行い、SMARTer Ultra Low RNA Kit for the Fluidigm C1 System(Takara)を用いた SMART-Seq によって RT-PCR を行った(図 10)。

まず初めに、細胞の Lysis、RT、PCR を行うための試薬調製を行った。Diluted RNA Spikes mixture(表 6)、MixA(Lysis) (表 7)、MixB(RT) (表 8)、MixC(PCR) (表 9)、MixD(Cell) (表 10)を各種試薬を調製して作成した。MixD を作成する際に用いた赤血球溶液は、濃縮を行った赤血球溶液を血球計算盤を用いて細胞数を計測して 170-200K cell/ml になるように RPMI Basal Medium で希釈して調製した。

続いて C1 IFC に Cell load するための Priming を行った。C1 Harvest Reagent 20(or 200)µl、C1 Preloading Reagent 20µl、C1 Blocking Reagent 15µl、C1 Cell Wash Buffer 20µl を C1 IFC 上の所定のウェルにアプライした(図 11)。C1 System の電源を入れ、Open を選択し、C1 IFC を台にセットして Close を選択。mRNA Seq : Prime (1771x)を選択して、Start を選択して Priming を行った。Priming 終了後、Eject を選択して C1 IFC を回収した。

続いて C1 IFC にて Cell Load を行った。C1 IFC にアプライした C1 Blocking Reagent をピペットで取り出して廃棄した(図 12)。C1 Cell Wash Buffer 20µl、Cell Mix D 10µl を C1 IFC 上の所定のウェルにアプライした(図 12)。C1 System で Open を選択し、C1 IFC を台にセットして Close を選択。mRNA Seq : Cell Load (1771x)を選択して、Start を選択して Cell Load を行った。Cell Load 終了後、Eject を選択して C1 IFC を回収した。

C1 IFC 中央部の細胞がキャプチャーされる 96 部屋を全て、顕微鏡で確認し各部屋にいくつの細胞がキャプチャーされたかを記録した(図 13)。キャプチャーされた細胞数が少ない場合、C1 IFC にアプライした Cell MixD をピペットで取り出して廃棄し、再度 Cell Load をやり直した。

続いて、C1 IFC にて Lysis、RT、PCR を行った。C1 Harvest Reagent 180 μ l、Lysis MixA 9 μ l、RT MixB 9 μ l、PCR MixC 24 μ l を C1 IFC 上の所定のウェルにアプライした(図 14)。C1 System で Open を選択し、C1 IFC を台にセットして Close を選択。mRNA Seq : RT&Amp (1771x)を選択して、Start を選択して Lysis、RT、PCR を行った。RT-PCR 終了後、Eject を選択して C1 IFC を回収した。

1 赤血球 RNA-Seq と並行してバルク RNA-Seq(ポジティブコントロール)とネガティブコントロールの作成を行った。PCR チューブに赤血球溶液(170-200K cell/ml) 1 μ l に Mix A 2 μ l をピペッティングで混和して加えた。ネガティブコントロールは、C1 Cell Wash Buffer 1 μ l に Mix A 2 μ l をピペッティングで混和して加えた(表 11)。サーマルサイクラーにて、72 $^{\circ}$ C 3 分に続き、4 $^{\circ}$ C 10 分、最後に 25 $^{\circ}$ C 1 分で処理して Lysis を行った。

PCR チューブを回収し、それぞれにボルテクスで混和した MixB 4 μ l を加えて蓋をし、PCR チューブをボルテクスして卓上遠心機で遠心した(表 12)。サーマルサイクラーにて、42 $^{\circ}$ C 90 分に続き、70 $^{\circ}$ C 10 分で処理して逆転写を行った。

新しい PCR チューブ 2 本にボルテクスで混和した MixC 9 μ l を加え、それぞれのチューブにバルク RNA-Seq とネガティブコントロールの逆転写産物 1 μ l を加えて、PCR チューブをボルテクスして卓上遠心機で遠心した(表 13)。サーマルサイクラーにて、熱変性 95 $^{\circ}$ C 1 分、続いて 5 サイクルの反応(1 サイクル :

熱変性 95°C 20 秒、アニーリング 58°C 4 分、伸長反応 68°C 6 分)を行い、9 サイクルの反応(1 サイクル：熱変性 95°C 20 秒、アニーリング 64°C 30 秒、伸長反応 68°C 6 分)を行い、続いて 7 サイクルの反応(1 サイクル：熱変性 95°C 30 秒、アニーリング 64°C 30 秒、伸長反応 68°C 7 分)を行い、最後に伸長反応を 72°C 10 分で処理した。

続いて、C1 IFC から RT-PCR 産物を回収した。MicroAmp® Optical 96-Well Reaction Plate (Life Technologies)の全ウェルに C1 DNA Dilution Reagent (Fluidigm) 10µl をアプライした。

C1 IFC についているシールを剥がし、RT-PCR 産物が保存されている所定のウェルから 8 連ピペットを 5µl に合わせて RT-PCR 産物を吸い出し、C1 DNA Dilution Reagent を加えた 96 穴プレートに移した(図 15)。

回収した RT-PCR 産物を 2100 バイオアナライザ(Agilent)を用いて定量した (p29 参照)。

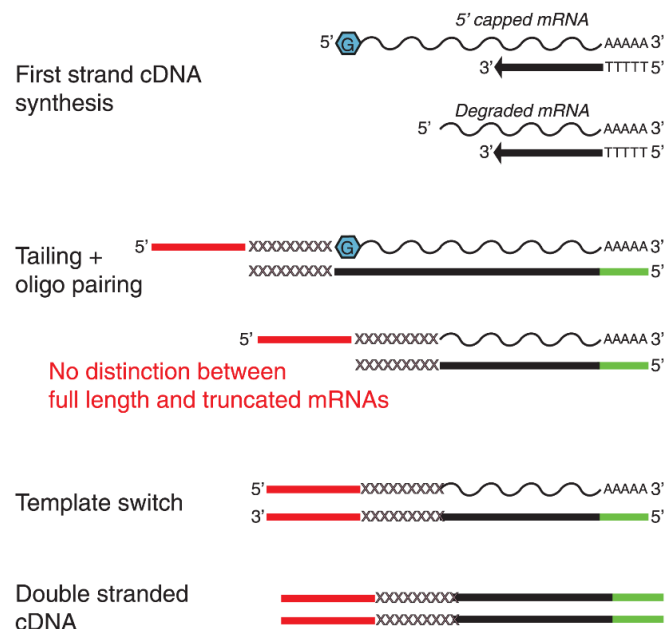


図 10. Smarter-Seq の手順

(Figure 1A, Maria Cartolano *et.al.*, 2016, *PLoSone*, 11(6): e0157779.

doi:10.1371/journal.pone.0157779)

表 6. Diluted RNA Spikes mixture の組成

| | |
|-------------------------------|--------------|
| RNA Spikes mixture (Ambion) | 1.0 μ l |
| C1 Loading Reagent (Fluidigm) | 99.0 μ l |

表 7. Lysis Mix (Mix A)

| | |
|---|--------------|
| Diluted RNA Spikes mixture | 1.0 μ l |
| RNase Inhibitor (Clontech) | 0.5 μ l |
| 3' SMART CDS Primer II A (Clontech) (stored at -20 °C) | 7.0 μ l |
| Clontech Dilution Buffer (brown bottle) (Do not vortex) | 11.5 μ l |

表 8. RT Reaction Mix (Mix B)

| | |
|---|--------------|
| C1 Loading Reagent (Fluidigm) | 1.2 μ l |
| 5X First-Strand Buffer (RNase-free) (Clontech) | 11.2 μ l |
| Dithiothreitol (Clontech) | 1.4 μ l |
| dNTP Mix (dATP, dCTP, dGTP, dTTP, each at 10 mM) (Clontech) | 5.6 μ l |
| SMARTer II A Oligonucleotide (Clontech) (stored at -80 °C) | 5.6 μ l |
| RNase Inhibitor (Clontech) | 1.4 μ l |
| SMARTScribe™ Reverse Transcriptase (Clontech) | 5.6 μ l |

表 9. PCR Mix (Mix C)

| | |
|--|--------------|
| PCR Water (Advantage 2 Kit) | 63.5 μ l |
| 10X Advantage 2 PCR Buffer (not SA, short amplicon) (Advantage 2 Kit) | 10.0 μ l |
| 50X dNTP Mix (Advantage 2 Kit) | 4.0 μ l |
| IS PCR primer (Clontech SMARTer) | 4.0 μ l |
| 50X Advantage 2 Polymerase Mix (Advantage 2 Kit) | 4.0 μ l |
| C1 Loading Reagent (Fluidigm) | 4.5 μ l |

表 10. Cell Mix (Mix D)

| | |
|---------------------------------------|------------|
| Cells 170-200K/mL | 12 μ l |
| C1 Cell Suspension Reagent (Fluidigm) | 8 μ l |

表 11. バルク細胞の Lysis

| Tube 1 – Positive | Control Tube 2 - NTC | |
|-------------------|----------------------|-------------|
| Cells 170-200K/mL | C1 Cell Wash Buffer | 1.0 μ l |
| Lysis Final Mix A | Lysis Final Mix A | 2.0 μ l |

表 12. バルク細胞の RT

| Tube 1 – Positive | Control Tube 2 - NTC | |
|-------------------|----------------------|-------------|
| Cell Lysis Mix | Cell Lysis Mix | 3.0 μ l |
| RT Final Mix B | RT Final Mix B | 4.0 μ l |

表 13. バルク細胞の PCR

| Tube 1 – Positive | Control Tube 2 - NTC | |
|-------------------|----------------------|-------------|
| PCR Mix C | PCR Mix C | 9.0 μ l |
| RT Reaction | RT Reaction | 1.0 μ l |

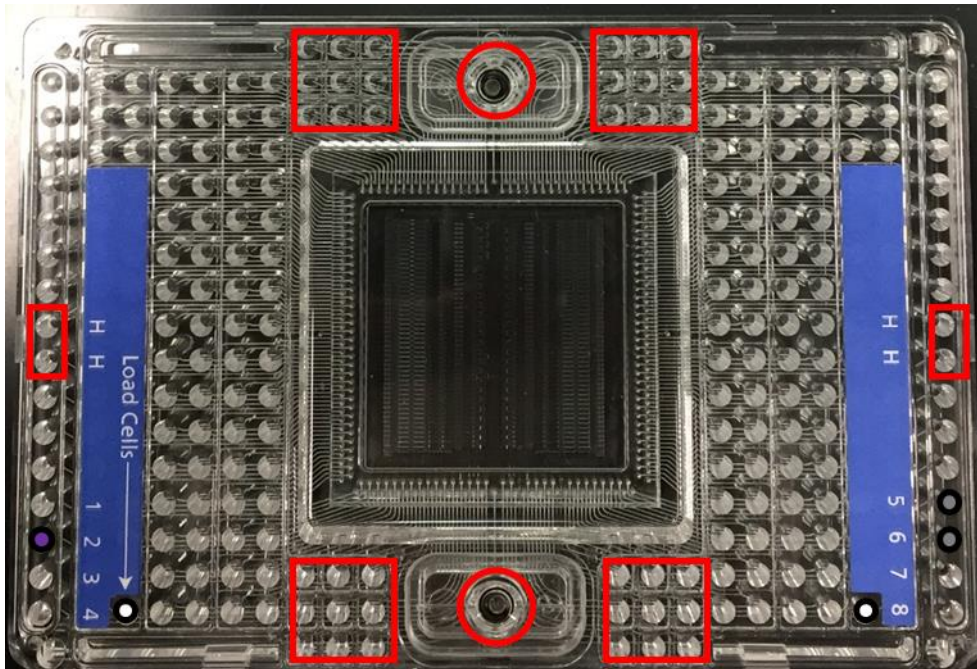


図 11. Prime 時に試薬をアプライした C1 IFC のウェル

□で囲んだウェルに Harvest Reagent 20 μ l, ○で囲んだウェルは蓋をピペットチップで押して隙間を作り, Harvest Reagent 200 μ l をアプライした. 白塗りの○に Blicking Reagent 15 μ l, #2 のウェルに Preloading Reagent 20 μ l, #5, #6 のウェルに Cell Wash Buffer 20 μ l を加えた.

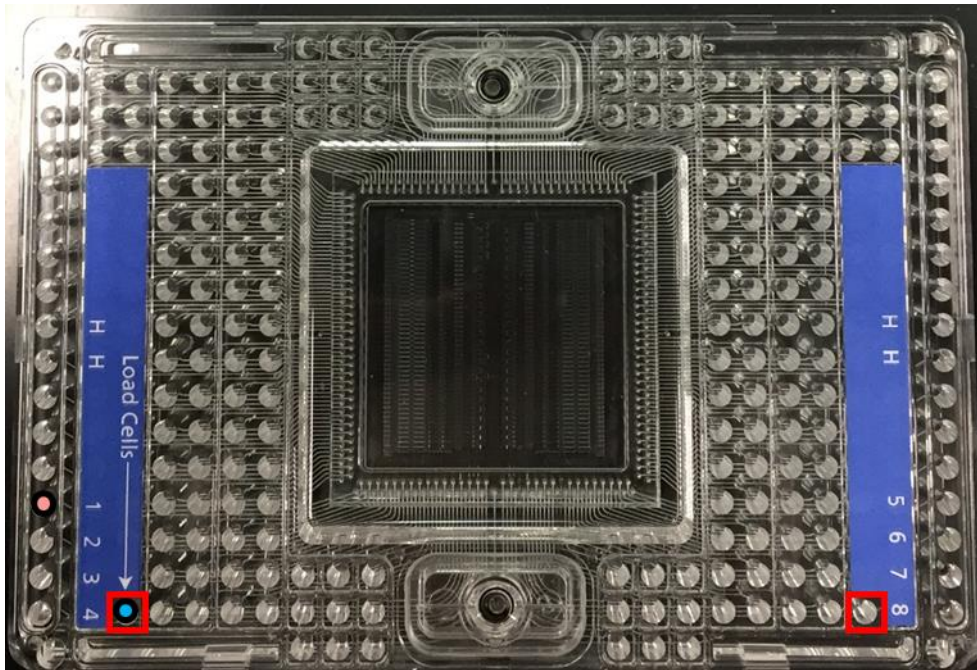


図 12. Cell Load 時に試薬をアプライした C1 IFC のウェル

□で囲んだウェルから Blocking Reagent を取り除き，#1 のウェルに Cell Wash Buffer 20 μ l， Load Cells のウェルに Cell Mix D 10 μ l をアプライした．

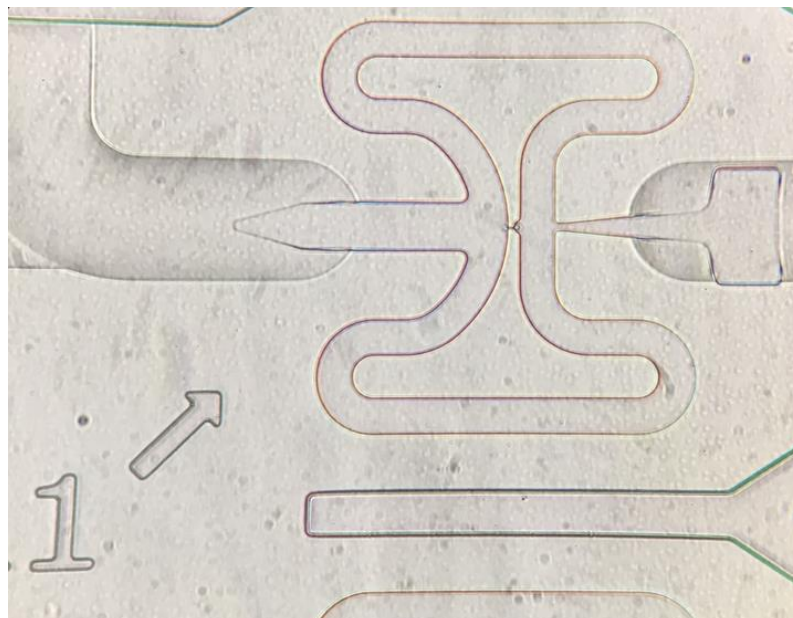


図 13. C1 IFC 上の細胞がキャプチャーされる各部屋の顕微鏡写真

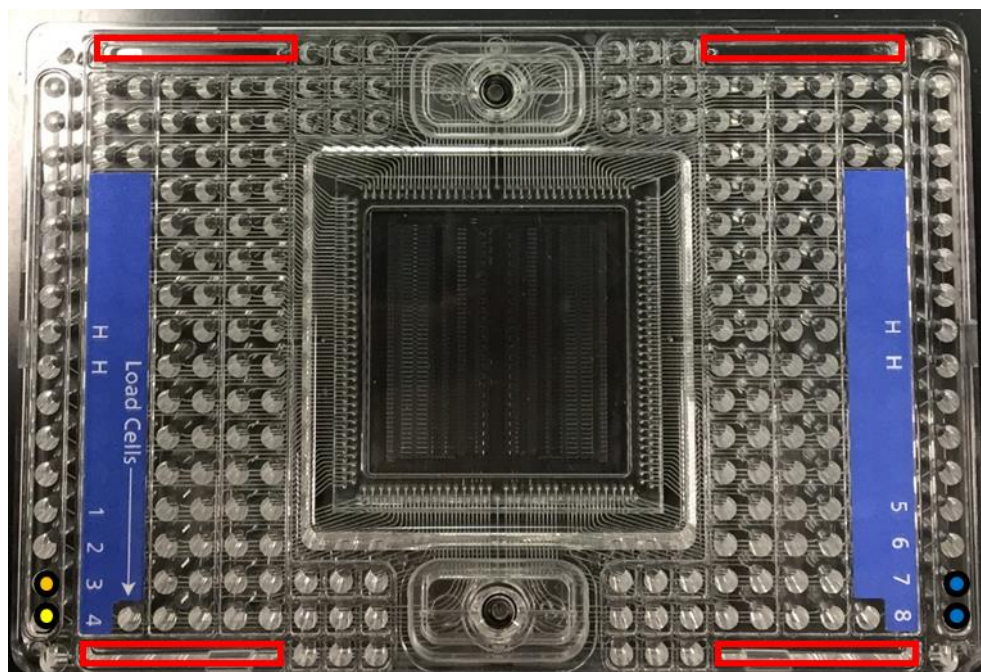
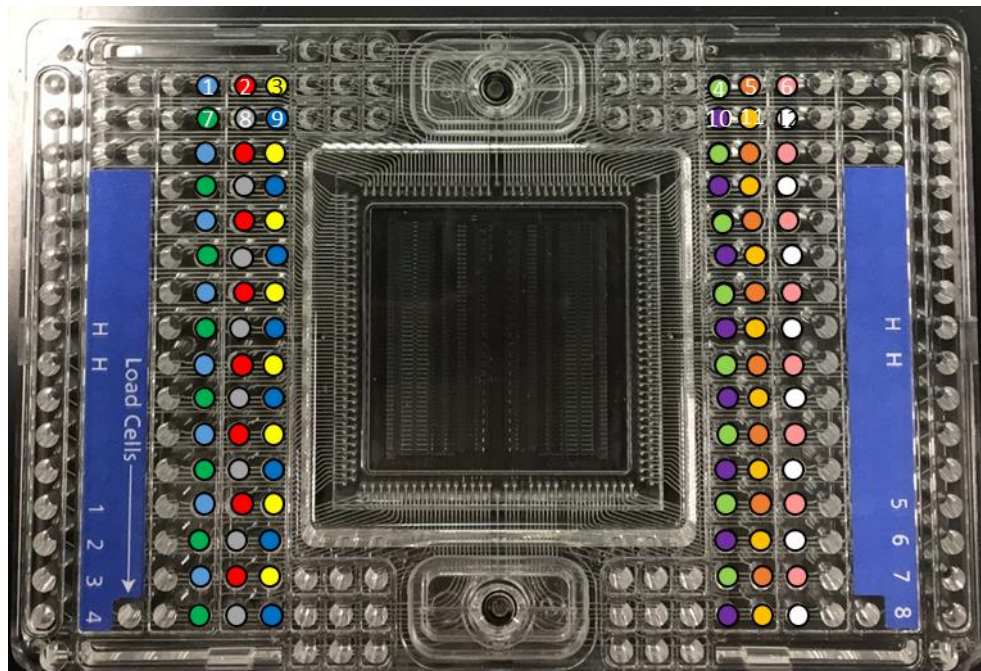


図 14. RT-PCR 時に試薬をアプライする C1 IFC のウェル

□で囲ったウェルに Harvest Reagent 180 μ l をアプライし, #3 のウェルに Lysis Mix A 9 μ l, #4 のウェルに RT MixB 9 μ l, #7, #8 のウェルに PCR MixC 24 μ l をアプライした.



| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|------|-----|--------|-------|--------|------|-------|------|------|--------|--------|-------|
| A | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |
| B | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |
| C | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |
| D | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |
| E | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |
| F | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |
| G | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |
| H | Blue | Red | Yellow | Green | Orange | Pink | Green | Grey | Blue | Purple | Yellow | White |

図 15. RT-PCR 産物の回収手順

8連ピペットを用いて96穴プレートに1列ずつ移動させた。順番は番号の順
に行った。

2.2.4. cDNA 量の定量

cDNA 量の定量は、2100 バイオアナライザを用いて、High Sensitivity DNA キット(Agilent)を調製して行った。

High Sensitivity DNA キットを 30 分間、室温で静置した。DNA gel matrix チューブに、DNA dye 溶液 15 μ l を加えてボルテクスした。2240 \times g \pm 20% で 10 分間遠心して gel-dye Mix を作成した。High Sensitivity DNA チップ (Agilent)の G(白抜き)マークに gel-dye Mix 9 μ l をアプライした。

Chip Priming Station を組み立てる。クリップのレバーを一番下の位置にセットしシリンジに装着した。Chip Priming Station から保護キャップを取り、シリンジのプランジャーを 1ml の目盛りに合わせ、Chip Priming Station に装着した。

High Sensitivity DNA チップをセットして、Chip Priming Station の蓋を閉じる。クリップに引っかかるまでプランジャーを下げ、1 分間静置した。クリップを引いて 5 秒間静置し、プランジャーを 1ml の目盛りまで慎重に引き上げた。Chip Priming Station の蓋を開けて High Sensitivity DNA チップを取り出した。

High Sensitivity DNA チップの G マークに gel-dye Mix 9 μ l をアプライした。続いて、番号が割り振られているウェルとラダーマークのウェルに Marker 5 μ l をアプライした。ラダーマークのウェルに High Sensitivity DNA ラダー 1 μ l をアプライした。定量する DNA 溶液 1 μ l を番号が割り振られているウェルにアプライした。DNA チップ専用のボルテクスミキサーで 2400rpm、1 分間でボルテクスした。電極クリーナーに Rnase-free water 350 μ l をアプライして 2100 バイオアナライザにセットして蓋を閉める。10 秒間静置して蓋を開け電極クリーナーを取り出した。電極を風乾した後、High Sensitivity DNA チップをセット

した。

コンピューターで 2100 Expert Software(Agilent)を立ち上げた。Assay Selection から dsDNA を選択して、High Sensitivity DNA.xsy を選択した。Run sample に定量するサンプル数を入力し、Start をクリックして定量を開始した。

定量後、High Sensitivity DNA チップを取り出し、電極クリーナーに Rnase-free water 350 μ l をアプライして 2100 バイオアナライザにセットして蓋を閉める。10 秒間静置して蓋を開け電極クリーナーを取り出し、電極を風乾した。

2.2.5. cDNA の鋳型調製

HiSeq 2500 によるシーケンスを実行するために、Nextera XT DNA Library Preparation Kit(illumina)を用いて cDNA のライブラリー調製を行った。

cDNA にインデックス配列を付加するために、タグメンテーションを行った。cDNA 溶液を 100-300pg/ μ l になるように RNase-free water で希釈し、Diluted Sample とした。

Tagment DNA Buffer(illumina)と Amplification Tagment Mix(illumina)を 2:1 の割合で 1.5ml チューブに移し、ボルテクスミキサーで混和して Pre-Mix とした。

8 連ピペットを用いて、96 穴プレートに Pre-Mix 3.75 μ l を分注し、Diluted Sample 1.25 μ l を各ウェルに分注した(表 14)。

PlateLoc Thermal Microplate Sealer (Agilent)を用いて 96 穴プレートをシールし、ボルテクスミキサーで混和し、1000rpm ,1min で遠心した。

サーマルサイクラーにて 55°C 10 分に続き、10°C ∞分で反応させた。続いて、96 穴プレートに NT Buffer 1.25µl を加えた(表 15)。

PlateLoc Thermal Microplate Sealer を用いて 96 穴プレートをシールし、ボルテクスミキサーで混和し、1000rpm ,1min で遠心した。

NPM 3.75µl と各 index プライマー 1.25µl(forward と reverse で計 2.5µl)を加えた(図 16,表 16)。PlateLoc Thermal Microplate Sealer を用いて 96 穴プレートをシールし、ボルテクスミキサーで混和し、1000rpm ,1min で遠心した。

PCR のサイクルは、伸長反応 72°C 3 分に続き、熱変性 95°C 30 秒、続いて 12 サイクルの反応(1 サイクル:熱変性 95°C 10 秒、アニーリング 55°C 30 秒、伸長反応 72°C 1 分)を行い、最後に伸長反応を 72°C 5 分で行った。

得られた PCR 産物を Ampure XP beads (beckmancoulter)を用いて磁気でビーズ精製した。96 穴プレートの各ウェルから PCR 産物 1µl を取り出し、1 つの 1.5ml チューブに集めた。ボルテクスして混和した Ampure XP beads を 1.5ml チューブに集めた PCR 産物の 90%量を加えた。5 回ほどピペッティングし、5 分間室温で静置した。

磁気スタンドにセットして 2 分間静置した後で、スタンドにセットしたままの状態の上清を取り除いた。70%エタノール 180µl を加えて 30 秒間静置した。エタノールをピペットで取り除き、もう一度 70%エタノール 180µl を加えて 30 秒間静置した。エタノールをピペットで取り除き、磁気スタンドから取り外して 5~10 分間風乾した。

1.5ml チューブに集めた PCR 産物 + 1µl 量の RNase-free water を加えて、20 秒間ボルテクスして 2 分間静置した。磁気スタンドにセットして、2 分間静置した。新しい 1.5ml チューブにビーズが混入しないように気を付けながら上清を回収した(この時、1.5ml チューブに集めた PCR 産物と同量の上清を回収した)。

もう一度、回収した上清でビーズ精製を行った。ボルテクスして混和した Ampure XP beads を 1.5ml チューブに回収した溶液の 90%量を加えた。5 回ほどピペッティングし、5 分間室温で静置した。

磁気スタンドにセットして 2 分間静置した後で、スタンドにセットしたままの状態の上清を取り除いた。70%エタノール 180 μ l を加えて 30 秒間静置した。エタノールをピペットで取り除き、もう一度 70%エタノール 180 μ l を加えて 30 秒間静置した。エタノールをピペットで取り除き、磁気スタンドから取り外して 5~10 分間風乾した。

RNase-free water 31 μ l を加えて、20 秒間ボルテクスして 2 分間静置した。磁気スタンドにセットして、2 分間静置した。新しい 1.5ml チューブにビーズが混入しないように気を付けながら上清 30 μ l を回収した。

バイオアナライザーによる Nextera 産物の定量を行った(p29 参照)。

表 14. Tagmentation reaction

| | |
|---------------------------|--------------|
| Tagment DNA Buffer | 2.5 μ l |
| Amplification Tagment Mix | 1.25 μ l |
| Diluted Sample | 1.25 μ l |

表 15. NT buffer addition

| | |
|--------------------|--------------|
| Library Prep Plate | 5.0 μ l |
| NT Buffer | 1.25 μ l |

表 16. Tagment reaction plus NPM

| | |
|--------------------|--------------|
| Library Prep Plate | 6.25 μ l |
| NPM | 3.75 μ l |

| | N701 | N702 | N703 | N704 | N705 | N706 | N707 | N708 | N709 | N710 | N711 | N712 | |
|-------------|----------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| S517 | A | S517/ N701 | S517/ N702 | S517/ N703 | S517/ N704 | S517/ N705 | S517/ N706 | S517/ N707 | S517/ N708 | S517/ N709 | S517/ N710 | S517/ N711 | S517/ N712 |
| S502 | B | S502/ N701 | S502/ N702 | S502/ N703 | S502/ N704 | S502/ N705 | S502/ N706 | S502/ N707 | S502/ N708 | S502/ N709 | S502/ N710 | S502/ N711 | S502/ N712 |
| S503 | C | S503/ N701 | S503/ N702 | S503/ N703 | S503/ N704 | S503/ N705 | S503/ N706 | S503/ N707 | S503/ N708 | S503/ N709 | S503/ N710 | S503/ N711 | S503/ N712 |
| S504 | D | S504/ N701 | S504/ N702 | S504/ N703 | S504/ N704 | S504/ N705 | S504/ N706 | S504/ N707 | S504/ N708 | S504/ N709 | S504/ N710 | S504/ N711 | S504/ N712 |
| S505 | E | S505/ N701 | S505/ N702 | S505/ N703 | S505/ N704 | S505/ N705 | S505/ N706 | S505/ N707 | S505/ N708 | S505/ N709 | S505/ N710 | S505/ N711 | S505/ N712 |
| S506 | F | S506/ N701 | S506/ N702 | S506/ N703 | S506/ N704 | S506/ N705 | S506/ N706 | S506/ N707 | S506/ N708 | S506/ N709 | S506/ N710 | S506/ N711 | S506/ N712 |
| S507 | G | S507/ N701 | S507/ N702 | S507/ N703 | S507/ N704 | S507/ N705 | S507/ N706 | S507/ N707 | S507/ N708 | S507/ N709 | S507/ N710 | S507/ N711 | S507/ N712 |
| S508 | H | S508/ N701 | S508/ N702 | S508/ N703 | S508/ N704 | S508/ N705 | S508/ N706 | S508/ N707 | S508/ N708 | S508/ N709 | S508/ N710 | S508/ N711 | S508/ N712 |

図 16. 分注したインデックスプライマーの位置と組み合わせ

2.2.6. 遺伝子発現量の算出

本研究では、HiSeq 2500 を用いて RNA-Seq を行った。ビーズ精製を施した cDNA ライブラリーをシングルリード、36bp でシーケンスした。

主にデータ処理は、東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピュータを利用して行った。

生の RNA-Seq データからクオリティが低いリードとマッピングクオリティの低いリードを取り除いた上で遺伝子発現量の指標である RPKM を算出した。

HiSeq 2500 より得られたデータから CASAVA を用いて FASTQ ファイルを作成した。CASAVA によって悪いリードと判定されたリード(Y)を Perl で作成したスクリプトを用いてフィルターした。

フィルターした FASTQ ファイルを TopHat2(Version 2.0.14) (Kim *et.al.* 2013)によって、熱帯熱マラリア原虫のリファレンスゲノムにマッピングした。熱帯熱マラリア原虫リファレンスゲノムは、PlasmoDB(Aurrecoechea C *et.al.* 2009)より PlasmoDB-11.0_Pfalciparum3D7.gff を入手して作成した。

```
--segment-length 18 --segment-mismatches 1 --no-novel-juncs
```

の条件でマッピングを実行した。

マッピングした BAM ファイルを SAMtools(Version 1.2)(Li *et.al.* 2009)で SAM ファイルに変換した。Perl で作成したスクリプトを用いて、SAM ファイルからマッピングクオリティを低いリード(FLG=0 または 16 以外、MAPQ<40)を除去した。

続いてタグカウントを算出し、RPKM を算出した。RPKM は、 $10^9 \times \text{Mapped Sequence reads} / \text{Gene length} \times \text{Total Malaria Parasite Mapped Sequence reads}$ (熱帯熱マラリア原虫にマッピングされた総リード)で割った。

得られた各 RNA-Seq データは熱帯熱マラリア原虫ゲノムに 50 万 read 以上

がマッピングされたのみを解析に用いるデータとした。

50 万 read という一定の基準以上のリード数があり、RPKM を算出した際に数値がノーマライズされることから、GAPDH 等のハウスキーピング遺伝子を用いたノーマライズは行わなかった。また、マラリア原虫はハウスキーピング遺伝子であっても発現量が生育ステージによって変化することから、発現量が一定ではないため使用するべきではないと考えた。

続いて、cDNA 合成を行う際に添加した Spike-in Mix のタグカウントから RPKM を算出した。CASAVA によって悪いリードと判定されたリード(Y)をフィルターした FASTQ ファイルを、Spike1、Spike4、Spike7 配列に bowtie2(Version 2.2.5) (Langmead & Salzberg 2012)を用いてマッピングした。

マッピングした BAM ファイルを SAMtools で SAM ファイルに変換した。Perl で作成したスクリプトを用いて、SAM ファイルからマッピングクオリティを低いリード(FLG=0 または 16 以外、MAPQ<40)を除去した。

タグカウントを算出し、各 Spike の RPKM を算出した。1 赤血球 RNA-Seq データセット中の Spike-in Mix(Spike1、Spike4、Spike7)の平均 RPKM と標準偏差(SD)を算出し、Spike1、Spike4、Spike7 のそれぞれの RPKM が平均 RPKM \pm 2SD の範囲内にある 1 赤血球 RNA-Seq データを増幅の偏りが少ない正しく cDNA 合成ができているデータとして、以後の解析に用いることとした(p56 参照)。

2.3. 生育ステージの同定

マラリア原虫の生活環は媒介蚊、宿主赤外期、宿主赤内期の3つに分かれる。赤内型マラリア原虫(宿主赤内期)には、メロゾイド、リング、トロホゾイド、シズントの生育ステージが存在する(図17)。

赤内型マラリア原虫は、その生育ステージによって大きくトランスクリプトームを変動させることが知られている(Bozdech *et. al.* 2003)。そのため、1赤血球 RNA-Seq データを利用した各種解析を行う際には、この生育ステージを考慮する必要性が有る。

本研究で得られた1赤血球 RNA-Seq データの生育ステージ同定を試みた。

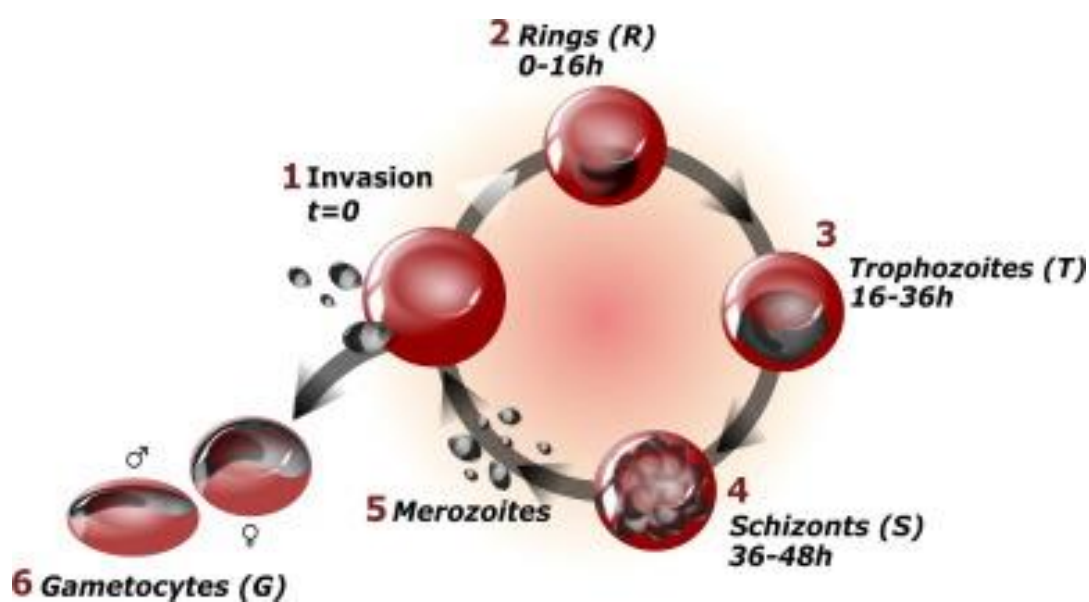


図17. 血液ステージのマラリア原虫

(Figure 1, Dekel *et. al.*, 2017, *Method*, 112(1):157-166.

<http://dx.doi.org/10.1016/j.ymeth.2016.06.021>)

2.3.1. 主成分分析

主成分分析とは、多変量データを統合して新しい指標を作り出すための手法である(Pearson 1901)。複数あるデータの要素をまとめて次元削減し、元のデータの情報量をなるべく多く維持しながら 2 次元または 3 次元で可視化することを可能とする。

この手法を、本研究で取得した同調培養株の 1 赤血球 RNA-Seq データ (トロホゾイド同調培養株 40 株、シズント同調培養株 26 株) (p54 参照) に対し利用して解析を行った。

同調培養株はその生育ステージが比較的均一であることから、異なる時間点の同調培養株(トロホゾイド同調培養株、シズント同調培養株)をまとめて主成分分析を行うことで、虫体毎(トロホゾイドとシズント)に分かれてクラスターを形成することが考えられる。この仮説を検証し、マラリア 1 赤血球 RNA-Seq データを主成分分析でどの程度解析することができるかについて検討した。

RStudio(Version 0.99)上で R 言語(R x64 3.3.1)を用いて解析を行った。トロホゾイド同調培養株 40、シズント同調培養株 26 の RNA-Seq データを $\text{Log}(\text{RPKM} + 1)$ で処理して対数化し、主成分分析を行った。パッケージ cluster を用いて PC1(第一主成分)と PC2(第二主成分)の 2 次元平面で k-means Clustering を行った。K = 2 の教師を与え、各 1 赤血球 RNA-Seq データの主成分がどのような分布でクラスタリングされるかを確認した。

また、PC1 と PC2 の因子負荷量を調べて、PC1 と PC2 にどのような機能を持った遺伝子が大きく寄与しているかを観測し、PC1 と PC2 が何を説明しているのかについて検討した。

また主成分分析と同様に tSNE(t-distributed stochastic neighbor embedding)

を実行し、どのような分布を示すかを確認した(Maaten and Hinton 2008)。パッケージ Rtsne を用いて、set.seed = 1、perplexity = 5 に設定して tSNE を行い、tSNE1 と tSNE2 の 2 次元平面における 1 赤血球 RNA-Seq データの分布を確認した。

2.3.2. 生育ステージ特異的遺伝子

DNA マイクロアレイを用いた先行研究により、バルク単位の発現解析により赤内型マラリアの後期ステージのトロホゾイドとシズントでそれぞれ特異的に発現しているマーカー遺伝子が同定されている(Bozdech *et. al.* 2003)。

本研究は、磁気による濃縮によって赤内型マラリアの後期ステージ(トロホゾイド、シズント)のみからデータを取得している。そのため、先行研究で示されていたマーカー遺伝子を用いることで、生育ステージの同定を行うことができることが考えられる。

マーカー遺伝子の発現量を確認することで、主成分分析によって得られた結果と比較して、主成分分析から予測された生育ステージの根拠を強めることができる考えた。

また本研究で取得したトロホゾイド同調培養株とシズント同調培養株の 1 赤血球 RNA-Seq データを用いて、生育ステージの各マーカー遺伝子の発現量がどのように表されるかは興味深い。

そのため、Bozdech らの先行研究で明らかになった各ステージのマーカー遺伝子の中から 10 遺伝子、計 20 遺伝子を取り上げ、RPKM を Z 変換した上でヒートマップを作成した(表 17) (Bozdech *et. al.* 2003)。

表 17. ヒートマップ作成の用いた各ステージのマーカー遺伝子

| Gene | Symbol | Product | Stage |
|---------------|---------------|--|-------------|
| PfEF2 | PF3D7_1451100 | elongation factor 2 | Trophozoite |
| RPS26 | PF3D7_0217800 | 40S ribosomal protein S26e, putative | Trophozoite |
| PfHSP86 | PF3D7_0708400 | heat shock protein 90 (HSP90) | Trophozoite |
| Enolase | PF3D7_1015900 | enolase (ENO) | Trophozoite |
| PfTPI(TIM) | PF3D7_1439900 | triosephosphate isomerase (TIM) | Trophozoite |
| PfLDH | PF3D7_1324900 | L-lactate dehydrogenase (LDH) | Trophozoite |
| PfHGPRT | PF3D7_1012400 | hypoxanthine-guanine phosphoribosyltransferase (HGPRT) | Trophozoite |
| Falcilysin | PF3D7_1360800 | falcilysin (FLN) | Trophozoite |
| plasmepsin II | PF3D7_1408000 | plasmepsin II | Trophozoite |
| PfHSP70 | PF3D7_0818900 | heat shock protein 70 (HSP70) | Trophozoite |
| PfPKAc | PF3D7_0934800 | cAMP-dependent protein kinase catalytic subunit (PKAc) | Schizont |
| RAP1 | PF3D7_1410400 | rhoptry-associated protein 1 (RAP1) | Schizont |
| RAP2 | PF3D7_0501600 | rhoptry-associated protein 2 (RAP2) | Schizont |
| MSP1 | PF3D7_0620400 | merozoite surface protein 1 (MSP1) | Schizont |
| MSP6 | PF3D7_1035500 | merozoite surface protein 6 (MSP6) | Schizont |
| EBA-175 | PF3D7_0731500 | erythrocyte binding antigen-175 (EBA175) | Schizont |
| EBL1 | PF3D7_1371600 | erythrocyte binding like protein 1, pseudogene (EBL1) | Schizont |
| AMA1 | PF3D7_1133400 | apical membrane antigen 1 (AMA1) | Schizont |
| PfSUB1 | PF3D7_0507500 | subtilisin-like protease 1 (SUB1) | Schizont |
| PfSUB2 | PF3D7_1136900 | subtilisin-like protease 2 (SUB2) | Schizont |

トロホゾイドは RNA・タンパク質合成が盛んに行われ、その関連遺伝子が多く発現する。シズントは次の RBC の感染する際に必要なメロゾイドの構造体に関連する遺伝子を多く発現している(Bozdech *et. al.* 2003).

2.3.3. 公共データセットを用いた生育ステージの同定

Bartfai らを含めた複数のグループが、血液ステージの熱帯熱マラリア原虫同調培養株から複数の時間点(hpi(hours post-infection))でバルク RNA-Seq を行い、FASTQ ファイルをインターネット上で公開している(Bartfai *et. al.* 2010)。

このバルク RNA-Seq の公共データセット(FASTQ ファイル)をダウンロードして RPKM を算出することで、生育段階が比較的均一であるバルク同調培養株の複数の時間点(hpi)と実験データを比較することができる。

公共データセットと各 1 赤血球 RNA-Seq データから相関係数を算出することで、各 1 赤血球 RNA-Seq データの生育ステージを予測できると考えた。

公共データセットの中で、血液ステージの熱帯熱マラリア原虫同調培養株から多くの時間点(hpi)で RNA-Seq を行ったデータセットを探し、GSE23865 と PRJEB2015 が使用に適していると判断した。

GSE23865、PRJEB2015 の FASTQ ファイルをダウンロードし、RPKM を算出し、公共データセットと各実験データの $\log(\text{RPKM}+1)$ からピアソン相関係数を算出した。PRJEB2015 はペアエンドシーケンスであったため、片側のみを利用した。

RPKM は各種実験データから RPKM 算出した際と同様の条件で算出した (p34,35 参照)。

2.4. トランスクリプトーム解析

本研究にて、5つの培養株で薬剤無添加サンプルと抗マラリア薬(クロロキン、ジヒドロアルテミシニン)添加サンプルの複数の薬剤添加後時間点から1赤血球 RNA-Seq データを取得した。

取得した1赤血球 RNA-Seq データから、各データにおけるトランスクリプトーム多様性や薬剤添加後の多様性変動等を観測できることが予想される。1赤血球トランスクリプトーム解析から薬剤耐性に対してトランスクリプトーム多様性が寄与しているかについて検証できると大変興味深い。この可能性を検証すべく、1赤血球 RNA-Seq データを用いた各種解析に取り組んだ。

2.4.1. 遺伝子発現数と分散

1赤血球 RNA-Seq データのそれぞれが、どの程度発現しているか、またどの程度のバラつきが存在するかは興味深い内容である。

この内容を調べるため、遺伝子発現数と分散を算出した。各データの遺伝子発現数と分散は、生育ステージや薬剤刺激によって変動する可能性が考えられる。この可能性を加味した上で、遺伝子発現数と分散を以下の手法で観測した。

まず初めに、生育ステージを予測しやすい同調培養株の RNA-Seq データを用いて、生育ステージと遺伝子発現数と分散の関係性を調べた。

データ中の RPKM > 5 以上の遺伝子を発現している遺伝子と定義して、各 RNA-Seq データ毎にカウントして集計した。同調培養株の主成分分析結果(PC1、PC2)と遺伝子発現数を対応させて、生育ステージとの関係性を調べた。また、

PC1 と PC2 の二次元平面上で K=3 の教師を与えて K-means Clustering を行い、Cluster1(トロホゾイド)、Cluster2(前期シズント)、Cluster3(後期シズント)にクラスターを分け、遺伝子発現数でヒストグラムを作成した。遺伝子発現数と生育ステージの関係性を検証した。

続いて、Cluster1(トロホゾイド)、Cluster2(前期シズント)、Cluster3(後期シズント)のそれぞれで分散を算出した。各クラスター中で 2 細胞以上が発現している遺伝子(RPKM \geq 5)に限定して分散を算出し、ヒストグラムを作成した。

2.4.2. 遺伝子発現変動

薬剤添加後反応を 1 赤血球レベルのトランスクリプトームから観測した際に、どのような変動を示されるかを各 1 赤血球 RNA-Seq データの生育ステージを加味した上で観測した。

公共データセットを用いた生育ステージ同定法(p40 参照)を用いて生育段階を後期トロホゾイド、前期シズント、後期シズントの三つに分類した(図 18)。

1 赤血球 RNA-Seq データの生育ステージの分類法は、公共データセットとの相関の最高値から判断して行った。1 赤血球 RNA-Seq データと GSE23865(5~40hpi)の相関の最高値が 35hpi かつ PRJEB2015(0~48hpi)の相関の最高値が 32hpi のデータを後期トロホゾイド、1 赤血球 RNA-Seq データと GSE23865 の相関の最高値が 40hpi かつ PRJEB2015 の相関の最高値が 40hpi のデータを前期シズント、1 赤血球 RNA-Seq データと GSE23865 の相関の最高値が 5hpi かつ PRJEB2015 の相関の最高値が 48hpi のデータを後期シズントと定義した(図 19)。

1 赤血球 RNA-Seq データの内、後期トロホゾイド、前期シズント、後期シズ

ントに分類されたデータでそれぞれ主成分分析を行った。薬剤無添加サンプルと抗マラリア薬添加後サンプルでどのように変動するかを確認するために、PC1 と PC2 から二次元平面上で可視化した。

全遺伝子 5777 を変数として、 $\text{Log}(\text{RPKM}+1)$ で主成分分析を実行した。

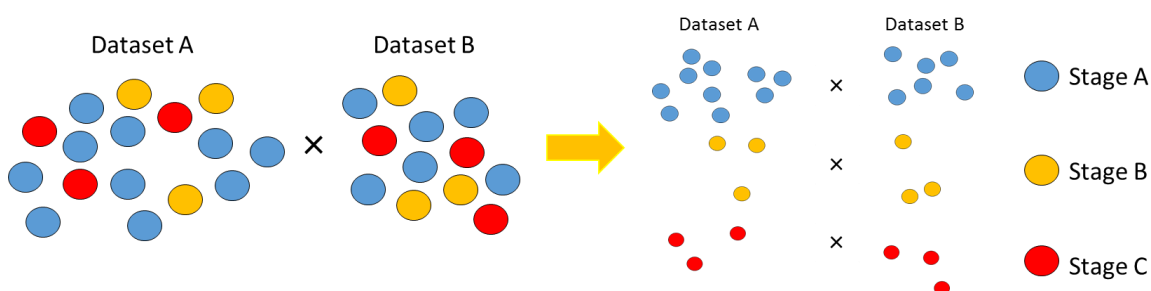


図 18. 生育ステージ分類の概略図

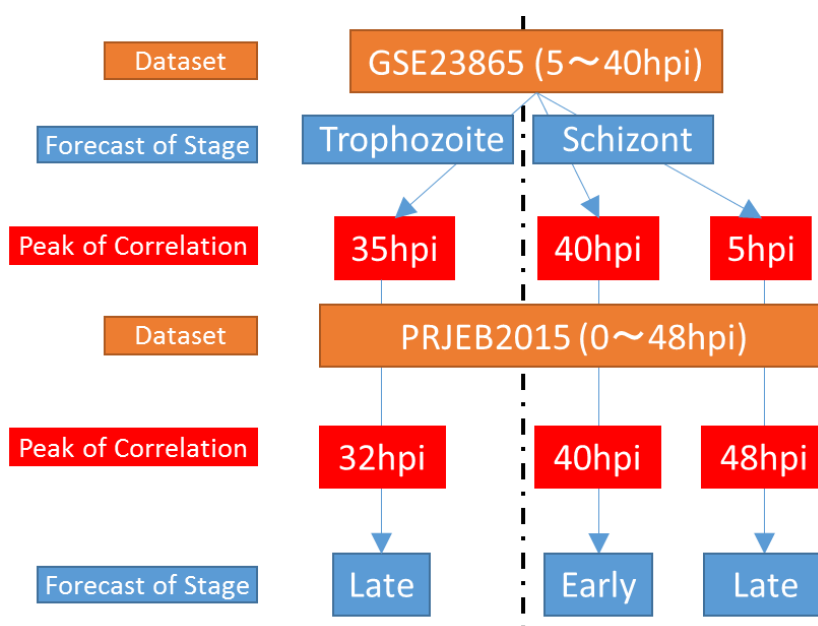


図 19. 生育ステージ分類法

1 赤血球 RNA-Seq データと GSE23865(5~40hpi)の相関の最高値(5~40hpi, 8 つのデータの中で最も相関係数が高かったデータの hpi)と PRJEB2015(0~48hpi, 7 つのデータ)の相関の最高値を規準に生育ステージを分類した。

2.4.3. 薬剤耐性関連遺伝子の発現量観測

熱帯熱マラリア原虫で、薬剤耐性関連遺伝子として考えられている遺伝子が複数存在する(表 18)。CRT に代表される薬剤耐性関連遺伝子は、変異により薬剤耐性が獲得されることが知られている(図 5)。

これら薬剤耐性関連遺伝子の発現多様性を、公共データセットを用いた生育ステージ同定法(p40 参照)を用いて各 1 赤血球 RNA-Seq データの生育ステージを推測した上で、薬剤耐性関連遺伝子の発現量を観測した。

各実験系データで、それぞれの薬剤耐性関連遺伝子の RPKM を基にストリップチャートを作成した。

表 18. 熱帯熱マラリア原虫の薬剤耐性関連遺伝子

| Gene | Symbol | Product |
|-----------------|---------------|---|
| Apocytochrome b | mal_mito_3 | apocytochrome b (cyb) |
| PfATPase6 | PF3D7_0106300 | calcium-transporting ATPase (ATP6) |
| MRP1 | PF3D7_0112200 | multidrug resistance-associated protein 1 (MRP1) |
| DHFR-TS | PF3D7_0417200 | bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS) |
| TCTP | PF3D7_0511000 | translationally-controlled tumor protein homolog (TCTP) |
| MDR1 | PF3D7_0523000 | multidrug resistance protein (MDR1) |
| CRT | PF3D7_0709000 | chloroquine resistance transporter (CRT) |
| DHPS | PF3D7_0810800 | dihydropteroate synthetase (DHPS) |
| ABC Transporter | PF3D7_1229100 | ABC transporter, (CT family) (MRP2) |
| K13-propeller | PF3D7_1343700 | kelch protein, putative |

クロロキン耐性関連遺伝子：CRT(Valderramos and Fidock 2006), アルテミシニン耐性関連遺伝子(または、耐性への関連が示唆される遺伝子)：PfATPase6 (Uhlemann *et.al.* 2012), K13-propeller(Ariey *et.al.* 2014), (TCTP(Eichhorna *et.al.* 2013)), ピリメタミン耐性関連遺伝子：DHFR-TS(Plowe *et.al.* 1997), アトバコン耐性関連遺伝子：Apocytochrome b(Fisher *et.al.* 2012), サルフアドキシニン耐性関連遺伝子：DHPS(Wang *et.al.* 1995), 多剤耐性関連遺伝子：MRP1(Dahlström *et.al.* 2009), MDR1(Reed *et.al.* 2000), ABC Transporter

2.4.4. GO 解析

薬剤添加前と薬剤添加後データを比較して薬剤添加によって遺伝子発現が Up-regulate または Down-regulate される遺伝子にどのような傾向があるかを観測するために、Gene ontology を用いた GO 解析を実行した。

マラリア原虫は生育ステージによってトランスクリプトーム解析を大きく変動させる。そのため、データを比較する際にはその生育ステージを加味する必要がある。生育ステージを分類し、薬剤添加前、添加後データで同一の生育ステージの 1 赤血球データで平均 RPKM を算出して比較することとした。

後期トロホゾイド、前期シズント、後期シズントの 3 種類に該当するデータを選択して、バーチャルバルク(平均 RPKM)を作成した。生育ステージの分類法は、公共データセットとの相関から判断することとした。

薬剤添加前と添加後で後期トロホゾイド、前期シズント、後期シズントの 3 種類に該当するデータを選択して、バーチャルバルク(平均 RPKM)を作成した。

算出した平均 RPKM に+1 して、薬剤添加後で RPKM5 以上かつ、薬剤添加後 RPKM / 薬剤添加前 RPKM が 3 以上の遺伝子を Up-regulate(薬剤添加後 3 倍以上の増加)、薬剤添加前で RPKM5 以上かつ、薬剤添加前 RPKM / 薬剤添加後 RPKM が 3 以上の遺伝子を Down-regulate(薬剤添加後 1/3 倍以下の減少)された遺伝子と定義した。

Up-regulate された遺伝子群と Down-regulate された遺伝子群にどのような傾向が見られるかを GO 解析によって検証した。

R 言語の phyper 関数を用いて、超幾何分布から p-value を算出した。

第3章 結果・考察

3.1. 実験結果

マラリア原虫を *in vitro* で調製し、SMART-Seq を行って 1 赤血球 RNA-Seq を実行した。サンプル調製や SMART-Seq 等それぞれの実験段階で条件検討を行う場面があった。

各実験の結果と、それぞれの実験段階で生じた問題点とそれに対するアプローチを記した(後述)。

3.1.1. サンプル調製

マラリア原虫の濃縮を行う際、血液ステージの後期に当たるトロホゾイドとシズントが少ない場合、濃縮の効率が落ちることが判明した。

感受性株に薬剤を添加して感染率が大きく減少した場合や、前期ステージ(リング)の割合が多いバッチから濃縮を行った際に、MACS Separation column から回収される RBC の総量が少なくなり、感染赤血球率が少なくなる傾向があったことからこのように判断した。

濃縮を行う段階で後期ステージの感染赤血球が多く存在すると考えられるタイミングで、バッチを調製して薬剤添加を行うことで、濃縮の効率を維持することができた。

また抵抗性株に薬剤添加を行う場合、感染率が上昇し過ぎる恐れがある。マラリア原虫は *in vitro* では、感染率 10%以上で死滅し始める。そのため、抵抗性株の感染率を 5%に調製し薬剤を添加する際には、薬剤添加後の培養期間でシ

ゾントが破裂して感染赤血球がなるべく増えないよう、リングステージの iRBC の割合が多いバッチに薬剤を添加した。

生育ステージを意識した上でサンプル調製を行うことで実験の成功率を高めることができた。

3.1.2. 同調培養

赤内型熱帯熱マラリア原虫の生育ステージはその周期が 36~48 時間であり、生育速度が不均一である(Dekel *et.al.* 2017)。

本研究で得られた RNA-Seq データは、様々なステージが存在する状態で濃縮を行って cDNA 合成を行っている。そのため、このデータセットを用いた各種解析を行う際には、生育ステージを判別する必要があった。

この生育ステージは同調培養を行うことで、限界はあるものの一定の精度で揃えることができる(Lambros and Vanderberg 1979)。

この同調培養株から 1 赤血球 RNA-Seq データを取得することができれば、生育ステージがある程度把握できている 1 赤血球 RNA-Seq データを得ることができる。そのため、生育ステージを同定する手法を検討する上で大きく利用できると考えられた。

また、同調培養株の 1 赤血球 RNA-Seq を行った先行研究は存在しなかった。そのため同調培養株に対して、本研究で用いた手法で 1 赤血球 RNA-Seq を行い、同調培養株の 1 赤血球レベルの観測と生育ステージの同定法の検討への利用を試みた。

杏林大学医学部の朝日博子博士より提供を受けた同調培養株(リング、トロホゾイド、シゾント)の濃縮前、濃縮後でギムザ染色を行い(p13 参照)、顕微鏡観察を行った (図 20-22)。

提供を受けた FCR3 株は血液ステージの周期が約 36 時間である。リング同調培養株は 4hpi(hours post-infection)で Lysis、トロホゾイド同調培養株では 16hpi で濃縮を開始して高い割合で同調していることを確認した(図 21)。

シズント同調培養株は 28hpi で濃縮を開始した。同調培養には限界があり、血液ステージが進むにつれ同調している割合が低くなる。シズント同調培養株は、シズントの中で生育段階に多少のバラつきが生じたことが考えられる(図 22)。

リング同調培養株は濃縮を行うことができないためバルク RNA-Seq のみ、トロホゾイド同調培養株とシズント同調培養株はバルク RNA-Seq と 1 赤血球 RNA-Seq の両方を SMARTer 法で実行した(p20 参照)。鋳型調製を施して RNA-Seq を実行し、バルク及び 1 赤血球 RNA-Seq データを取得した(後述)。

また 0.5%D-ソルビトール処理後に、クロロキン(最終濃度:200nM)を添加したトロホゾイド同調培養株で、バルク RNA-Seq と 1 赤血球 RNA-Seq の両方を SMARTer 法で実行した。鋳型調製を施して RNA-Seq を実行したが、クロロキン添加に問題があり薬剤刺激が入っていない可能性が示唆されたため、解析には使用しなかった。

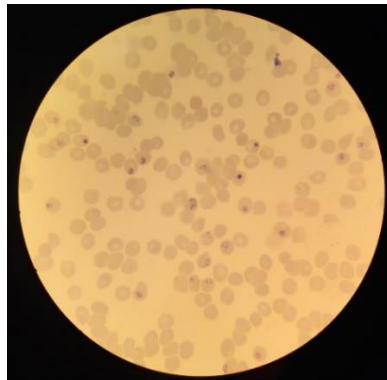


図 20.リング同調培養株の顕微鏡写真

リングは Separation による濃縮を行えないため、濃縮は行わなかった。

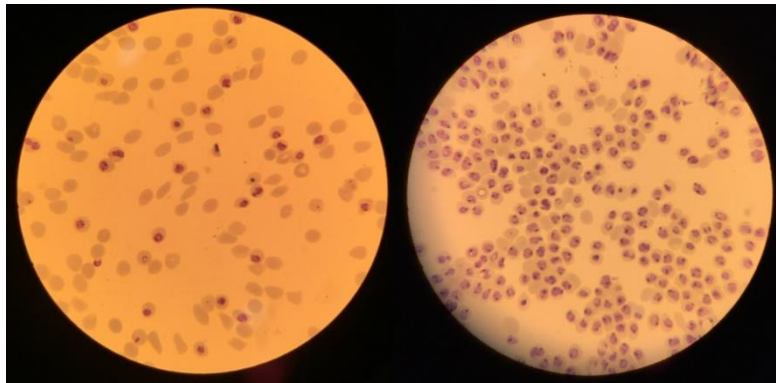


図 21.トロホゾイド同調培養株の顕微鏡写真

(左)濃縮前, (右)濃縮後.

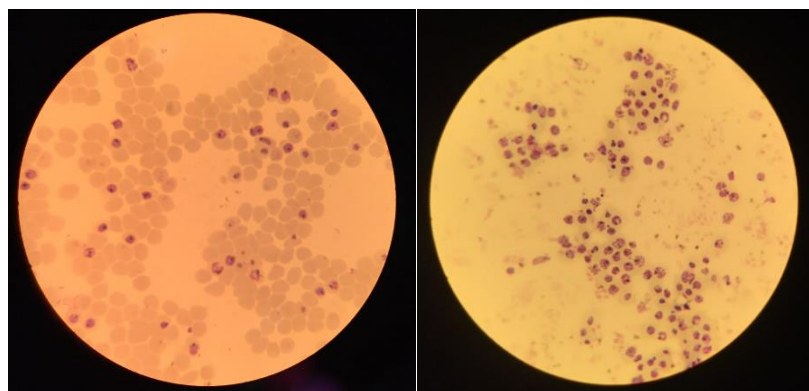


図 22.シズント同調培養株の顕微鏡写真

(左)濃縮前, (右)濃縮後.

3.1.3. SMART-Seq

C1 より回収した cDNA 溶液をバイオアナライザーによって定量し、正しく cDNA 合成できているかを検証した。

その結果、感受性株への薬剤添加を行った場合、得られる cDNA 量が少なくなることが明らかとなった。このことから、薬剤刺激によってトランスクリプトームの総量に影響が出ている可能性が考えられた。

また、正しく cDNA 合成が行われている場合と行われていない場合が存在した。この理由として、SMARTer Ultra Low RNA Kit for the Fluidigm C1 System のロットによる RNA 検出感度の差が原因ではないかと考えた。

本研究は、マラリア原虫が感染した赤血球をシングルセル化して Lysis、RT-PCR を行っている。そのため、赤血球内に残存する RNA とマラリア原虫が保持する RNA と添加した Spike-in RNA から cDNA 合成を行う。しかし、マラリア原虫は赤血球内に存在するため、赤血球の Lysis がうまく行かなかった場合、マラリア原虫 RNA を取り出せなくなることが考えられる。これにより RNA 総量が低くなり、SMART-Seq による増幅が正しく行われな可能性が考えられた。

RNA 総量が低い場合に問題となってくるのが、SMARTer Ultra Low RNA Kit のロットによる差であると考えた。通常のシングルセル RNA-Seq で問題にならない程度のロット差が、マラリア原虫を 1 赤血球レベルで Lysis、RT、PCR する場合にいずれかの段階で問題となり、増幅効率が落ちてしまった可能性が考えられた。

本研究で当初用いていた正しく cDNA 合成が行えるロット(SMARTer Ultra Low RNA Kit for the Fluidigm C1 System)をタカラバイオ株式会社の協力の元探し求めたが見つかることができなかった。

そのため、現存する限られたロットの中で比較的増幅が可能なロットを選出し、選出したロットでプロトコールの一部改変して増幅した cDNA に鋳型調製を施した。

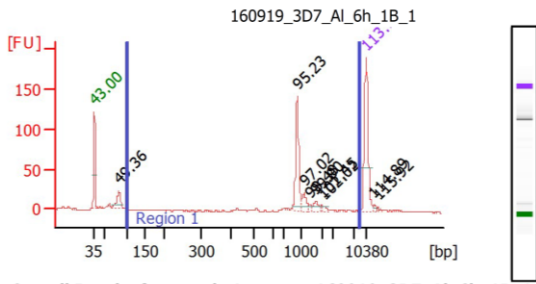
C1 System より回収した cDNA は、通常のプロトコールでは C1 Dilution Buffer 10 μ l による希釈で濃度調製を行ってからバイオアナライザーにより cDNA 量を定量する(図 22)。この Dilution Buffer による希釈の工程を省くことにより、Nextera によるタグメンテーションが可能である濃度(100-300pg/ μ l)に達するかを検証した。その結果、鋳型調製に用いることができるだけの濃度を持った cDNA 溶液を得ることができた。

しかし、マラリア原虫の cDNA よりはるかに多い Spike-in 由来の cDNA(約 950bp)が合成されてしまった(図 23)。図 23 の実験系から得られた cDNA から RNA-Seq を実行したが、マラリア原虫ゲノムへのマップ率は 16.4%程度であり、Spike-in 配列へのマップ率は 66.5%となった。

本研究の当初使用していた、増幅効率に問題のないロットのキットから取得したデータでは、マラリア原虫ゲノムへのマップ率は 34.2%、Spike-in 配列へのマップ率は 52%であった。このことから、使用できるロットに通常量の Spike-in RNA を添加場合、Spike-in 由来の cDNA が多く合成されてしまうことが明らかとなった。

このことを受け、Spike-in RNA の添加量を 2.5 倍、4 倍、5 倍、10 倍希釈して同様の実験を行った。その結果、Spike-in RNA 添加量を 2.5~5 倍に希釈することで、解析に使用できるだけのマラリア原虫由来と Spike-in RNA のリード数を得ることができた(図 25)。2.5 倍希釈した場合、マラリア原虫ゲノムへのマップ率は 28.8%、Spike-in へのマップ率は 57.1%であった。

Electropherogram Summary



Overall Results for sample 1 : 160919 3D7 AI 6h 1B 1

Number of peaks found: 9
 Noise: 0.4
 Corr. Area 1: 244.4

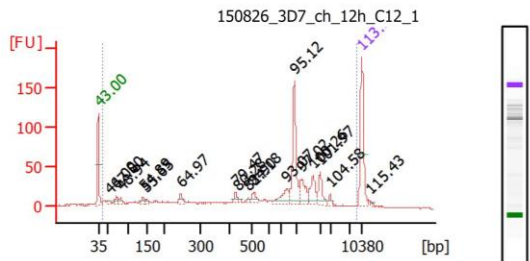
Region table for sample 1 : 160919 3D7 AI 6h 1B 1

| From [bp] | To [bp] | Corr. Area | % of Total | Average Size [bp] | Size distribution in CV [%] | Conc. [pg/μl] | Molarity [pmol/l] | Co lo |
|-----------|---------|------------|------------|-------------------|-----------------------------|---------------|-------------------|-------|
| 103 | 8,570 | 244.4 | 66 | 1,275 | 73.9 | 113.41 | 168.2 | ■ |

Peak table for sample 1 : 160919 3D7 AI 6h 1B 1

| Peak | Size [bp] | Conc. [pg/μl] | Molarity [pmol/l] | Observations |
|------|-----------|---------------|-------------------|--------------|
| 1 | 35 | 125.00 | 5,411.3 | Lower Marker |
| 2 | 84 | 46.29 | 835.6 | |
| 3 | 943 | 69.72 | 112.0 | |
| 4 | 1,173 | 13.29 | 17.2 | |
| 5 | 1,431 | 2.61 | 2.8 | |
| 6 | 1,680 | 8.04 | 7.2 | |
| 7 | 1,955 | 1.64 | 1.3 | |
| 8 | 2,108 | 2.90 | 2.1 | |
| 9 | 10,380 | 75.00 | 10.9 | Upper Marker |
| 10 | 12,269 | 0.00 | 0.0 | |
| 11 | 13,308 | 0.00 | 0.0 | |

図 23. マラリア由来の cDNA 合成効率が低い SMARTer Ultra Low RNA Kit を Spike-in 濃度を変更せずに使用した 1 赤血球由来の cDNA の定量結果



Overall Results for sample 1 : 150826 3D7 ch 12h C12 1

Number of peaks found: 17
 Noise: 0.5
 Corr. Area 1: 944.3

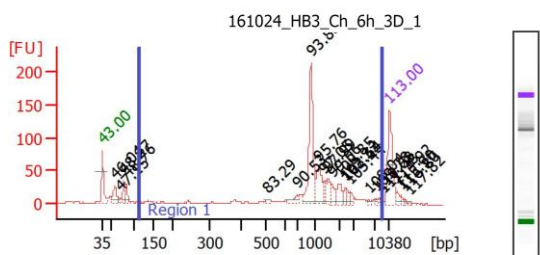
Region table for sample 1 : 150826 3D7 ch 12h C12 1

| From [bp] | To [bp] | Corr. Area | % of Total | Average Size [bp] | Size distribution in CV [%] | Conc. [pg/μl] | Molarity [pmol/l] | Co lo |
|-----------|---------|------------|------------|-------------------|-----------------------------|---------------|-------------------|-------|
| 41 | 8,697 | 944.3 | 96 | 1,013 | 95.4 | 581.62 | 3,806.0 | ■ |

Peak table for sample 1 : 150826 3D7 ch 12h C12 1

| Peak | Size [bp] | Conc. [pg/μl] | Molarity [pmol/l] | Observations |
|------|-----------|---------------|-------------------|--------------|
| 1 | 35 | 125.00 | 5,411.3 | Lower Marker |
| 2 | 55 | 10.58 | 290.6 | |
| 3 | 71 | 17.02 | 363.9 | |
| 4 | 81 | 15.46 | 290.0 | |
| 5 | 139 | 13.49 | 147.3 | |
| 6 | 147 | 6.69 | 69.1 | |
| 7 | 246 | 13.58 | 83.6 | |
| 8 | 426 | 7.98 | 28.4 | |
| 9 | 440 | 3.45 | 11.9 | |
| 10 | 486 | 2.44 | 7.6 | |
| 11 | 511 | 12.89 | 38.3 | |
| 12 | 814 | 27.99 | 52.1 | |
| 13 | 961 | 91.14 | 143.6 | |
| 14 | 1,222 | 30.01 | 37.2 | |
| 15 | 1,745 | 27.65 | 24.0 | |
| 16 | 2,045 | 22.48 | 16.7 | |
| 17 | 2,899 | 5.17 | 2.7 | |
| 18 | 10,380 | 75.00 | 10.9 | Upper Marker |
| 19 | 13,015 | 0.00 | 0.0 | |

図 24. マラリア由来の cDNA 合成効率に問題がない SMARTer Ultra Low RNA Kit を使用した 1 赤血球由来の cDNA の定量結果



Overall Results for sample 2 : 161024 HB3 Ch 6h 3D 1

Number of peaks found: 23
 Noise: 0.4
 Corr. Area 1: 811.0

Region table for sample 2 : 161024 HB3 Ch 6h 3D 1

| From [bp] | To [bp] | Corr. Area | % of Total | Average Size [bp] | Size distribution in CV [%] | Conc. [pg/μl] | Molarity [pmol/l] | Color |
|-----------|---------|------------|------------|-------------------|-----------------------------|---------------|-------------------|-------|
| 114 | 8,679 | 811.0 | 76 | 1,525 | 92.2 | 412.04 | 900.0 | Blue |

Peak table for sample 2 : 161024 HB3 Ch 6h 3D 1

| Peak | Size [bp] | Conc. [pg/μl] | Molarity [pmol/l] | Observations |
|------|-----------|---------------|-------------------|--------------|
| 1 | 35 | 125.00 | 5,411.3 | Lower Marker |
| 2 | 57 | 51.41 | 1,377.8 | |
| 3 | 67 | 9.10 | 206.6 | |
| 4 | 77 | 47.40 | 934.5 | |
| 5 | 81 | 29.52 | 550.0 | |
| 6 | 514 | 6.75 | 19.9 | |
| 7 | 700 | 15.12 | 32.7 | |
| 8 | 943 | 139.52 | 224.2 | |
| 9 | 1,175 | 41.72 | 53.8 | |
| 10 | 1,381 | 10.86 | 11.9 | |
| 11 | 1,531 | 24.97 | 24.7 | |
| 12 | 1,702 | 12.44 | 11.1 | |
| 13 | 1,987 | 20.02 | 15.3 | |
| 14 | 2,438 | 7.50 | 4.7 | |
| 15 | 2,536 | 7.75 | 4.6 | |
| 16 | 2,789 | 7.91 | 4.3 | |
| 17 | 6,001 | 2.39 | 0.6 | |
| 18 | 7,402 | 1.99 | 0.4 | |
| 19 | 8,229 | 1.87 | 0.3 | |
| 20 | 8,974 | 1.89 | 0.3 | |
| 21 | 10,380 | 75.00 | 10.9 | Upper Marker |
| 22 | 12,159 | 0.00 | 0.0 | |
| 23 | 13,069 | 0.00 | 0.0 | |
| 24 | 13,896 | 0.00 | 0.0 | |

図25. マラリア由来のcDNA合成効率が低いSMARTer Ultra Low RNA Kit を Spike-in濃度を変更(2.5倍希釈)して使用した1赤血球由来のcDNAの定量結果

3.1.4. シーケンスデータの取得

取得データの内、Spike-in の増幅に外れ値が存在する 1 赤血球 RNA-Seq データ(p56 参照)、マラリア原虫にマッピングされたリード数が 50 万以下の 1 赤血球 RNA-Seq データを除いて取得した 1 赤血球 RNA-Seq データ数を計測した

結果、本研究で 27 の実験系から計 897 細胞のデータを取得した。また同調培養株から 2 つの実験系(トロホゾイド同調培養株、シizont同調培養株)から計 66 細胞のデータを所得した。取得データの内訳は、下記の表 19、20 に記す。

3D7 株の薬剤無添加サンプルは 3 回、クロロキン添加後 12 時間サンプルは 3 回、クロロキン添加後 24 時間サンプルは 2 回の実験からデータを取得した。

その他の実験系では、一度の実験よりデータを取得した。

表 19. 本研究で得られた実験データセット

| | | Number of Cell | | | | | | | Mean of Malaria mapped tag | | | | | | |
|-------|-----|----------------|----|-----|-------------|----|-----|-------------|----------------------------|---------|---------|-------------|--------|---------|-------------|
| | | Free | | | Chloroquine | | | Artemisinin | Free | | | Chloroquine | | | Artemisinin |
| 添加後時間 | | 0H | 6H | 12H | 24H | 6H | 12H | 24H | 0H | 6H | 12H | 24H | 6H | 12H | 24H |
| 株 | 3D7 | 78 | 39 | 35 | 30 | 12 | 21 | 9 | 3168329 | 1170353 | 874254 | 860387 | 802218 | 793512 | 765610 |
| | HB3 | 18 | 25 | 24 | 15 | 19 | 24 | | 1999500 | 672318 | 901191 | 1445157 | 690264 | 1065879 | |
| | 7G8 | 41 | 39 | 20 | 28 | 25 | | | 1531454 | 946856 | 932584 | 756653 | 607080 | | |
| | K1 | 35 | 46 | 28 | 24 | | | | 2722759 | 2575461 | 1193620 | 865639 | | | |
| | DD2 | 40 | 21 | 29 | 24 | 33 | | | 1021500 | 657571 | 815036 | 940504 | 931041 | | |

表 20. 同調培養株から得られたデータセット

| Stage | Number of Cell | | Mean of Malaria mapped tag | |
|-------|----------------|----------|----------------------------|----------|
| | Trophozoite | Schizont | Trophozoite | Schizont |
| FCR3 | 40 | 26 | 831406 | 917940 |

3.2. シーケンスデータの信頼性と再現性

得られた 1 赤血球 RNA-Seq データの信頼性と再現性を、ポジティブコントロール(バルク RNA-Seq データ)との比較及び Spike-in の増幅から検証した。

3.2.1. シーケンスデータの信頼性

Spike-in Mix の増幅量からデータの信頼性を検証するとした。

cDNA 合成時に添加した Spike-in Mix が正しく増幅したかを検証した。

Spike-in Mix は、通常 Spike1 > Spike4 > Spike7 の順に多く増幅されるよう設計されている。

この Spike-in Mix が正しく増幅しているかを各実験系で確認した。1 赤血球 RNA-Seq の平均 RPKM \pm 2SD に範囲内のデータのみを抽出した。各 Spike1、Spike4、Spike7 の内 1 つでも平均 RPKM \pm 2SD から外れた場合、その 1 赤血球 RNA-Seq データは取り除いた。その結果、約 10%程度の 1 赤血球 RNA-Seq データが除かれた。

すべての実験系で Spike-in の増幅量を確認した結果、Spike1 > Spike4 > Spike7 の順の増幅していた(図 26)。

続いて、得られた各データセットのバルク RNA-Seq データの RPKM とシングルセル RNA-Seq の平均 RPKM から相関係数を算出した。全遺伝子(5777 遺伝子)を変数としてピアソン相関係数を算出し、 $\log_{10}(\text{RPKM}+1)$ で散布図を作成して各遺伝子の相関を検証した。

その結果、3D7 株、薬剤無添加の実験系では、RPKM 同士で相関係数 0.96 という値を得ることができた(図 26)。この値は、本研究の解析で用いる上では

十分な値であると考えた。

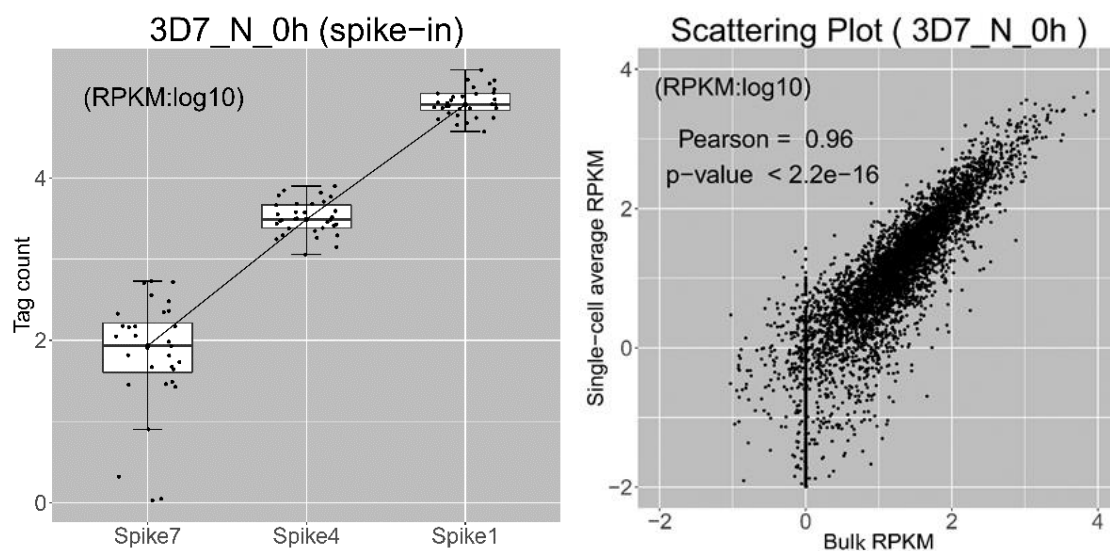


図 26. 3D7 株、薬剤無添加の実験系におけるデータの信頼性を検証.

(左) Spike-in Mix の増幅量, (右)全遺伝子(5777)における Bulk RPKM とシングルセル平均 RPKM に散布図.

3.2.2. シークエンスデータの再現性

3D7 株、薬剤無添加の実験系で再現実験を行い、同一条件のデータセット間の相関係数を調べた。3つの Duplicate を作成し、それぞれのデータセット間の相関係数を算出した(表 21)。

Duplicate 間でバルク対バルク RPKM、シングルセル平均対シングルセル平均 RPKM でピアソン相関係数を算出し、 $\log_{10}(\text{RPKM})$ で散布図を作成した。

その結果、バルク対バルク RPKM で相関係数 0.88-0.96、シングルセル平均対シングルセル平均 RPKM で相関係数 0.93-0.97 という十分であると考えられる値を得ることができた(図 27)。

これらの結果から、本研究の 1 赤血球 RNA-Seq は一定の再現性を持ったデータを取得できると考えた。

表 21. 3D7 株、薬剤無添加のデータセットの詳細

| Dataset | Number of Cell | Mean of Malaria Mapped tag | Malaria Mapped tag (Bulk) |
|----------|----------------|----------------------------|---------------------------|
| Dataset1 | 32 | 5463395 | 882483 |
| Dataset2 | 7 | 572889 | 177795 |
| Dataset3 | 39 | 1751047 | 905291 |

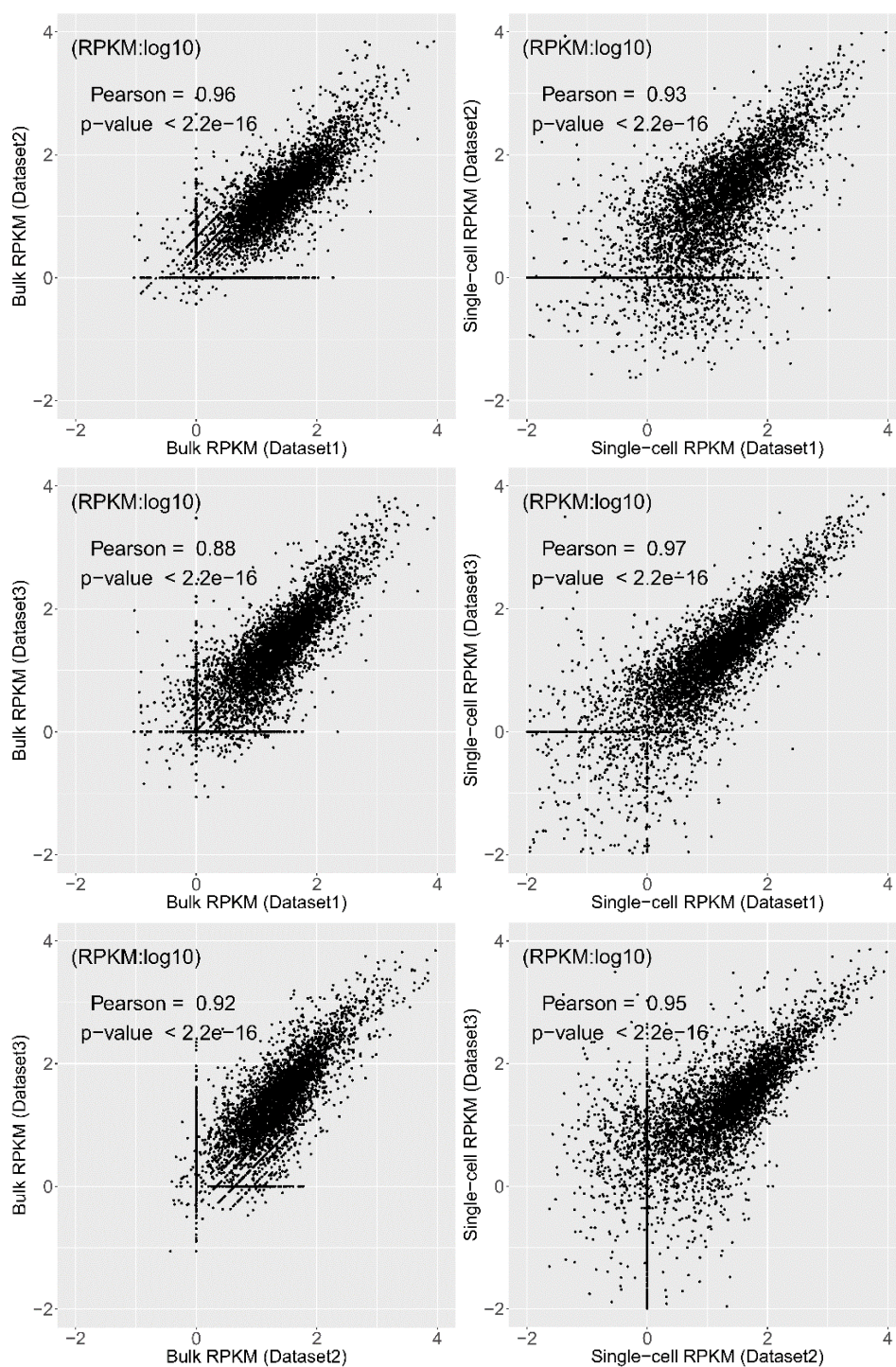


図 27. Duplicate 同士の散布図(3D7 株、薬剤無添加の実験系)

(左) 同一条件である実験系のバルク RPKM 同士の散布図,

(右) 同一条件である実験系の平均シングルセル RPKM 同士の散布図,

(上段) Duplicate1 対 2, (中段) Duplicate1 対 3, (下段) Duplicate2 対 3.

3.3. 生育ステージ分類法の検討

生育ステージ同定のため、以下の 3 つの手法を考案した。これらの手法を取得されたデータに適用し、その妥当性およびそれぞれの手法が適応可能な制約条件について考察を行った。

詳細は後述するが、要約すると、第一の主成分分析を用いた方法では、その生育ステージと各細胞のトランスクリプトーム類似度を可視化することができた。しかし、生育ステージの分類はデータセット中に各生育ステージの細胞が一定数存在する場合のみであった。特定のステージが大多数を占めた場合にクラスタリングできないといった欠点があった。

また、クラスタリング結果のみから各クラスターの生育ステージを同定することは不可能であり、因子負荷量を調べた上でその値の絶対値が高い遺伝子を調べることで初めて推測できる程度であった。加えて、主成分分析結果単体では、その結果を十分信用できるだけの根拠を得られないといった欠点もあった。これらの問題点を補うべく、次の手法を考案した。

第二のマーカー遺伝子を用いた方法は、生育ステージの分類にデータセット中に各生育ステージの細胞存在する必要は必ずしもなかった。各生育ステージのマーカー遺伝子の発現の有無によって、その結果の裏付けも存在することから主成分分析の欠点を補うことができる手法であることが評価できた。

しかし、主成分分析と併用することで各データの生育ステージを同定して分類することが可能となったが、生育ステージの境界付近に位置する細胞を分類する際に明確に判断できないといった欠点があった。生育ステージを分類する明確な基準を定めることが難いため、各種解析で大量の 1 赤血球 RNA-Seq データに生育ステージを分類する際に、この手法を利用することが現実的でないこ

とが考えられた。また、特定の遺伝子に絞って生育ステージを分類する手法は、1 赤血球 RNA-Seq データを処理する上では、変数選択が難しいという欠点もある。そこで、変数を全遺伝子でも設定でき、かつ生育ステージを分類する基準を作成しやすく、簡便に大量の 1 赤血球 RNA-Seq データを生育ステージ毎に分類できる手法を次に考案した。

第三の公共データセットとの相関から生育ステージを同定する手法であるが、こちらは各種解析に用いる際に最も使用しやすい手法であった。公共データセットを用いていることから、得られる結果に各生育ステージの裏付けが取れているからである。また、生育ステージを分類する明確な基準を定めることができるためである。そのため、各種解析に用いる際に、各 1 赤血球 RNA-Seq データを 1 細胞からでも同定することができる他、数値で区切ることができるため分類が難しい、ステージの境界付近に位置する細胞であっても分類することが可能である。

この分類の精度がどこまで信用することができるかは、今後更なる検討をする必要があるが、第一、第二の手法と結果を照らし合わせた結果、大多数の細胞で同様の結果が得られたことから、この第三の手法を主として利用することとした。

全遺伝子を変数として、解析する第一と第三の手法は、薬剤添加等の条件によって、そのトランスクリプトームが大きく変動した場合にその精度が問題になる可能性が考えられる。

今後、各手法の更なる条件検討を行い、それぞれがより良い手法となるよう研究を試みていきたい。

それぞれの手法における具体的な解析結果、考察を下記に記した。

3.3.1. 主成分分析

血液ステージにおける虫体が均一に近いと思われる同調培養株（トロホゾイド同調培養株とシズント同調培養株）について、1 赤血球 RNA-Seq データを用いて主成分分析を行った。主成分分析を行う際、熱帯熱マラリア原虫が保持する全遺伝子(5777 遺伝子)を変数とした。得られた第一主成分(PC1)と第二主成分(PC2)を用いて K-means Clustering を行い、2 つの虫体（トロホゾイドとシズント）に 1 赤血球 RNA-Seq データがクラスタリングされるかについて検証した。

トロホゾイド同調株の 1 赤血球 RNA-Seq データ 40 細胞と、シズント同調株の 1 赤血球 RNA-Seq データが 26 細胞で主成分分析を行った。

第一主成分(PC1)と第二主成分(PC2)で K-means clustering を行って 2 群に分けた結果、トロホゾイドとシズントに分類できたと考えられた(図 28)。トロホゾイドと考えられる群には 35 細胞が分類され、トロホゾイド同調株のみが分類された。一方で、シズントと考えられる群にはシズント同調株 26 細胞とトロホゾイド同調株 5 細胞が分類された。

続いて、K-means clustering を行って 3 群に分けた結果、2 群で分けた際のシズントと考えられたクラスターがさらに 2 つに分かれた。分かれたシズントの 2 つのクラスターは、どちらか一方が前期シズントであり、もう一方が後期シズントであることが考えられた(図 28)。

同調培養株の主成分分析と K-means Clustering の結果に対する解釈が正しいかを検証するため、因子負荷量が高い遺伝子を調べた。第一主成分、第二主成分の因子負荷量で、正の因子負荷量が高い遺伝子と負の因子負荷量が高い遺伝子、それぞれ 50 遺伝子を調べた。

その結果、PC1 の正の因子負荷量ではシズント特異的に発現するとされる遺

伝子が多く存在することが明らかとなった。Merozoite Surface Protein、cAMP-dependent protein kinase catalytic、rhoptry-associated protein 等が存在した(表 22)。

一方、PC1 の負の因子負荷量ではトロホゾイドで特異的に発現するとされる遺伝子が多く存在することが明らかとなった。トロホゾイドは RNA 合成やタンパク質合成が盛んに行われているとされ、それに関連すると考えられる遺伝子が多く存在することが明らかとなった。トロホゾイドで多く発現するといわれている、ribosomal protein、triosephosphate isomerase、heat shock protein 70、hypoxanthine-guanine phosphoribosyltransferase 等が存在した(表 23)。

次に、PC2 の正負の大きい因子負荷量を調べた。その結果、どちらもシズント特異的な遺伝子が存在することが明らかとなった。rhoptry-associated protein、serine repeat antigen、subtilisin-like protease、erythrocyte binding antigen-175、myosin A 等が含まれていた(表 24,25)。以上の結果から、PC2 ではシズントにおける前期段階と後期段階を説明していることが考えられた。

ここで得られた PC1 と PC2 の正と負の因子負荷量が多い各 50 遺伝子を用いることで、生育ステージを分類することができるのではないかと考えた。また、同調培養株から得た因子負荷量の大きい遺伝子を別のデータセットで観測することで、同調培養株から得た因子負荷量が多い遺伝子と生育ステージの関連性を検証できると考えた。

PC1 と PC2 の正と負の因子負荷量が多いそれぞれ 50 遺伝子を用いてヒートマップを作製した(図 29)。ヒートマップは、3D7 株、薬剤無添加サンプル、3 つの duplicate を合わせた計 78 細胞のデータを用いて行った(p57 参照)。

その結果、PC1 と PC2 の両方で正の因子負荷量の遺伝子群と負の因子負荷量の遺伝子群で発現量が拮抗している細胞が存在した。このことから、同調培養

株から得た因子負荷量が大きい遺伝子と生育ステージと関連している可能性が存在することが明らかとなった。その他の生育ステージ分類法と併用して検証し、この可能性を強める根拠を探った(後述)。

同調培養株において、主成分分析を用いることで生育ステージを分類することができると考えられた。同調培養を行っていない実験データに、この手法が応用できるかを確かめるべく、同調培養株で行った際の変数条件(全遺伝子)を対応させ主成分分析を行った。

3D7 株、薬剤無添加サンプル、3つの duplicate を合わせた計 78 細胞のデータを用いて生育段階に応じて分布するかを検証した。その結果、一定の細胞数を持つクラスターを形成した(図 30)。しかし、この結果のみからは生育ステージによってクラスターが分かれているか判断することができなかった。また、バッチ差による影響が考えられる。

そのため、1つバッチのみ計 32 細胞で主成分分析と K-means Clustering を実行した。その結果、細胞数 1 のクラスターが現れた。これは、このバッチ中に同様のステージの細胞が偏って多く存在していたことが考えられる。このような条件下では、主成分分析による生育ステージの推測は行うことができないことが明らかになった。このことから主成分分析による生育ステージの分類は、その他分類手法と併用して判断するべきであると考えた。

続いて、同調培養株で tSNE を行い tSNE1 と tSNE2 でどのように各細胞が分布するかを検証した結果、主成分分析同様に 3 つの群に分かれることが明らかとなった(図 31)。主成分分析と比べ、明確に分類された。しかし、set.seed や perplexity といった、値の設定により結果が前後することから本研究では、主成分分析を優先して使用した。

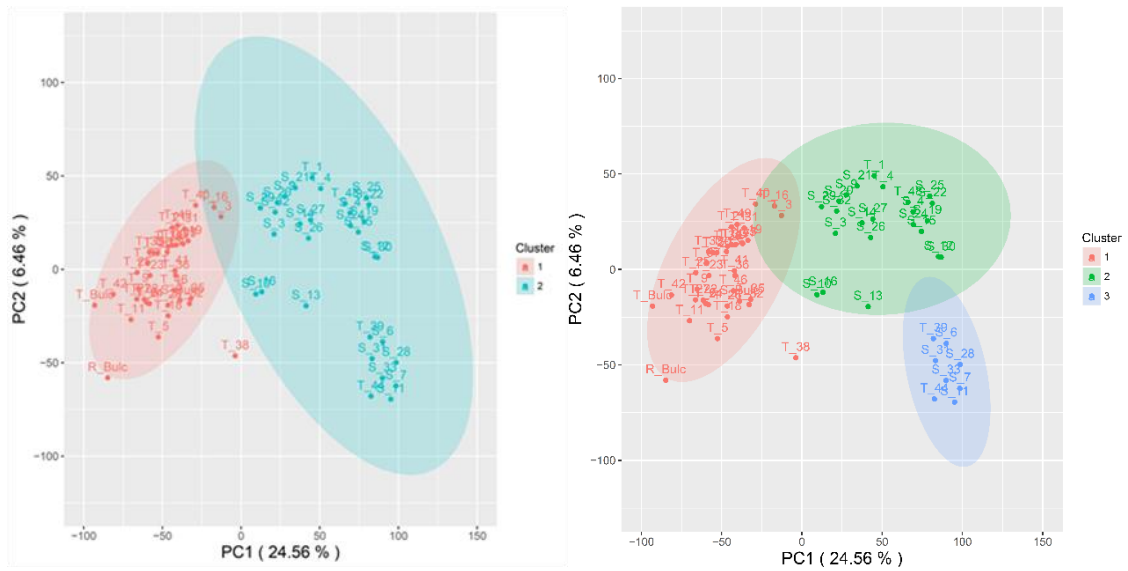


図 28. 同調培養株の主成分分析(変数：全遺伝子(5777))

各細胞の名称は、T_○がトロホゾイド同調培養株，S_○がシズント同調培養株である。

表 22. PC1 で正の因子負荷量が大きい遺伝子（上位 50 から一部を抜粋）

| Positive correlation | Product | Loadings |
|----------------------|--|----------|
| PF3D7_1035300 | glutamate-rich protein (GLURP) | 2.31 |
| PF3D7_1003600 | membrane skeletal protein IMC1-related (ALV5) | 2.14 |
| PF3D7_0217500 | calcium-dependent protein kinase 1 (CDPK1) | 2.08 |
| PF3D7_1468400 | zinc finger protein, putative (D13) | 1.86 |
| PF3D7_0404700 | dipeptidyl peptidase 3 (DPAP3) | 1.66 |
| PF3D7_1423300 | serine/threonine protein phosphatase (PP7) | 1.59 |
| PF3D7_0628100 | HECT-domain (ubiquitin-transferase), putative | 1.58 |
| PF3D7_0707300 | rhostry-associated membrane antigen (RAMA) | 1.53 |
| PF3D7_1410400 | rhostry-associated protein 1 (RAP1) | 1.52 |
| PF3D7_1351600 | glycerol kinase (GK) | 1.50 |
| PF3D7_0722200 | rhostry-associated leucine zipper-like protein 1 (RALP1) | 1.48 |
| PF3D7_1035400 | merozoite surface protein 3 (MSP3) | 1.47 |
| PF3D7_0934800 | cAMP-dependent protein kinase catalytic subunit (PKAc) | 1.34 |
| PF3D7_1035200 | S-antigen | 1.33 |
| PF3D7_0302500 | cytoadherence linked asexual protein 3.1 (CLAG3.1) | 1.30 |
| PF3D7_0424100 | reticulocyte binding protein homologue 5 (RH5) | 1.27 |
| PF3D7_1116000 | rhostry neck protein 4 (RON4) | 1.26 |
| PF3D7_1035500 | merozoite surface protein 6 (MSP6) | 1.25 |
| PF3D7_1335100 | merozoite surface protein 7 (MSP7) | 1.22 |
| PF3D7_0717500 | calcium-dependent protein kinase 4 (CDPK4) | 1.21 |
| PF3D7_1472600 | protein disulfide isomerase (PDI-14) | 1.21 |
| PF3D7_1460600 | inner membrane complex protein, putative | 1.19 |
| PF3D7_0414900 | armadillo-domain containing rhostry protein (ARO) | 1.17 |
| PF3D7_1251200 | coronin | 1.17 |
| PF3D7_0501600 | rhostry-associated protein 2 (RAP2) | 1.15 |

表 23. PC1 で負の因子負荷量が大きい遺伝子（上位 50 から一部を抜粋）

| Negative correlation | product | Loadings |
|----------------------|---|----------|
| PF3D7_1408100 | plasmepsin III,histo-aspartic protease (HAP) | -2.95 |
| PF3D7_0516900 | 60S ribosomal protein L8, putative | -2.92 |
| PF3D7_1130200 | 60S ribosomal protein P0 (PfP0) | -2.81 |
| PF3D7_0532100 | early transcribed membrane protein 5 (ETRAMP5) | -2.79 |
| PF3D7_0814000 | 60S ribosomal protein L13-2, putative | -2.78 |
| PF3D7_0917900 | heat shock protein 70 (HSP70-2) | -2.75 |
| PF3D7_1462800 | glyceraldehyde-3-phosphate dehydrogenase (GAPDH) | -2.74 |
| PF3D7_0513800 | Rab GTPase 1a (RAB1a) | -2.72 |
| PF3D7_1130100 | 60S ribosomal protein L38e, putative | -2.66 |
| PF3D7_1420000 | splicing factor 3B subunit 4, putative (SF3B4) | -2.64 |
| PF3D7_0915100 | ubiquitin conjugating enzyme, putative | -2.61 |
| PF3D7_1022400 | pre-mRNA splicing factor, putative | -2.57 |
| PF3D7_1325100 | phosphoribosylpyrophosphate synthetase | -2.55 |
| PF3D7_0322000 | peptidyl-prolyl cis-trans isomerase (CYP19A) | -2.53 |
| PF3D7_0814200 | DNA/RNA-binding protein Alba 1 (ALBA1) | -2.48 |
| PF3D7_0511000 | translationally-controlled tumor protein homolog (TCTP) | -2.47 |
| PF3D7_1119300 | U2 snRNP auxiliary factor, small subunit, putative | -2.47 |
| PF3D7_1033700 | bromodomain protein, putative | -2.46 |
| PF3D7_1016300 | glycophorin binding protein (GBP) | -2.45 |
| PF3D7_1012400 | hypoxanthine-guanine phosphoribosyltransferase (HGPRT) | -2.44 |
| PF3D7_1103100 | 60S acidic ribosomal protein P1, putative | -2.40 |
| PF3D7_0826700 | receptor for activated c kinase (RACK) | -2.40 |
| PF3D7_0309600 | 60S acidic ribosomal protein P2 (PfP2) | -2.39 |
| PF3D7_0103200 | nucleoside transporter 4 (NT4) | -2.39 |
| PF3D7_1439900 | triosephosphate isomerase (TIM) | -2.37 |

表 24. PC2 で正の因子負荷量が大きい遺伝子（上位 50 から一部を抜粋）

| Positive correlation | Product | Loadings |
|----------------------|---|----------|
| PF3D7_0517900 | zinc finger protein, putative | 2.01 |
| PF3D7_0929400 | high molecular weight rhoptry protein 2 (RhopH2) | 1.96 |
| PF3D7_0508100 | SET domain protein, putative (SET9) | 1.94 |
| PF3D7_0905400 | high molecular weight rhoptry protein 3 (RhopH3) | 1.89 |
| PF3D7_0604100 | transcription factor with AP2 domain(s),SPE2-interacting protein (SIP2) | 1.88 |
| PF3D7_1035100 | probable protein, unknown function | 1.79 |
| PF3D7_1410400 | rhoptry-associated protein 1 (RAP1) | 1.78 |
| PF3D7_1252100 | rhoptry neck protein 3 (RON3) | 1.76 |
| PF3D7_1228600 | merozoite surface protein 9 (MSP9) | 1.72 |
| PF3D7_0207700 | serine repeat antigen 4 (SERA4) | 1.68 |
| PF3D7_0207600 | serine repeat antigen 5 (SERA5) | 1.64 |
| PF3D7_0522600 | inner membrane complex protein | 1.63 |
| PF3D7_1361100 | Sec24 subunit a (SEC24a) | 1.62 |
| PF3D7_0423500 | glideosome associated protein with multiple membrane spans 2 (GAPM2) | 1.59 |
| PF3D7_0806300 | ferlin like protein, putative | 1.55 |
| PF3D7_0828500 | translation initiation factor EIF-2b alpha subunit, putative | 1.55 |
| PF3D7_0919900 | regulator of chromosome condensation, putative | 1.53 |
| PF3D7_1020900 | ADP-ribosylation factor (ARF1) | 1.47 |
| PF3D7_0707300 | rhoptry-associated membrane antigen (RAMA) | 1.47 |
| PF3D7_0807800 | proteasome subunit alpha type 5, putative | 1.43 |
| PF3D7_0302500 | cytoadherence linked asexual protein 3.1 (CLAG3.1) | 1.39 |
| PF3D7_1364100 | cysteine-rich surface protein (Pf92) | 1.38 |
| PF3D7_0722200 | rhoptry-associated leucine zipper-like protein 1 (RALP1) | 1.38 |
| PF3D7_0612200 | leucine-rich repeat protein (LRR6) | 1.37 |
| PF3D7_0918000 | secreted acid phosphatase (GAP50) | 1.36 |

表 25. PC2 で負の因子負荷量が大きい遺伝子（上位 50 から一部を抜粋）

| Negative correlation | product | Loadings |
|----------------------|--|----------|
| PF3D7_1342600 | myosin A (MyoA) | -1.87 |
| PF3D7_1302100 | gamete antigen 27/25 (Pfg27) | -1.86 |
| PF3D7_0212600 | secreted protein altered thrombospondin repeat protein (SPATR) | -1.82 |
| PF3D7_0418500 | Bet3 transport protein, putative (BET3p) | -1.78 |
| PF3D7_0724900 | kinesin-like protein, putative | -1.68 |
| PF3D7_1143100 | transcription factor with AP2 domain(s) (AP2-O) | -1.67 |
| PF3D7_1251200 | coronin | -1.64 |
| PF3D7_1136900 | subtilisin-like protease 2 (SUB2) | -1.63 |
| PF3D7_0525800 | membrane skeletal protein IMC1-related | -1.63 |
| PF3D7_1003600 | membrane skeletal protein IMC1-related (ALV5) | -1.61 |
| PF3D7_1231400 | amino acid transporter, putative | -1.59 |
| PF3D7_0731500 | erythrocyte binding antigen-175 (EBA175) | -1.55 |
| PF3D7_1245800 | clathrin coat assembly protein AP180, putative | -1.52 |
| PF3D7_0308200 | TCP-1/cpn60 chaperonin family, putative | -1.51 |
| PF3D7_1351600 | glycerol kinase (GK) | -1.50 |
| PF3D7_0414900 | armadillo-domain containing rhoptry protein (ARO) | -1.49 |
| PF3D7_0102500 | erythrocyte binding antigen-181 (EBA181) | -1.48 |
| PF3D7_1035500 | merozoite surface protein 6 (MSP6) | -1.47 |
| PF3D7_1218000 | thrombospondin-related apical membrane protein (TRAMP) | -1.46 |
| PF3D7_1125700 | kelch protein, putative | -1.44 |
| PF3D7_0220100 | DnaJ protein, putative | -1.43 |
| PF3D7_0424100 | reticulocyte binding protein homologue 5 (RH5) | -1.41 |
| PF3D7_0507500 | subtilisin-like protease 1 (SUB1) | -1.36 |
| PF3D7_0206900.2 | merozoite surface protein 5 (MSP5) | -1.35 |
| PF3D7_0402300 | reticulocyte binding protein homologue 1,normocyte binding protein 1 (RH1) | -1.34 |

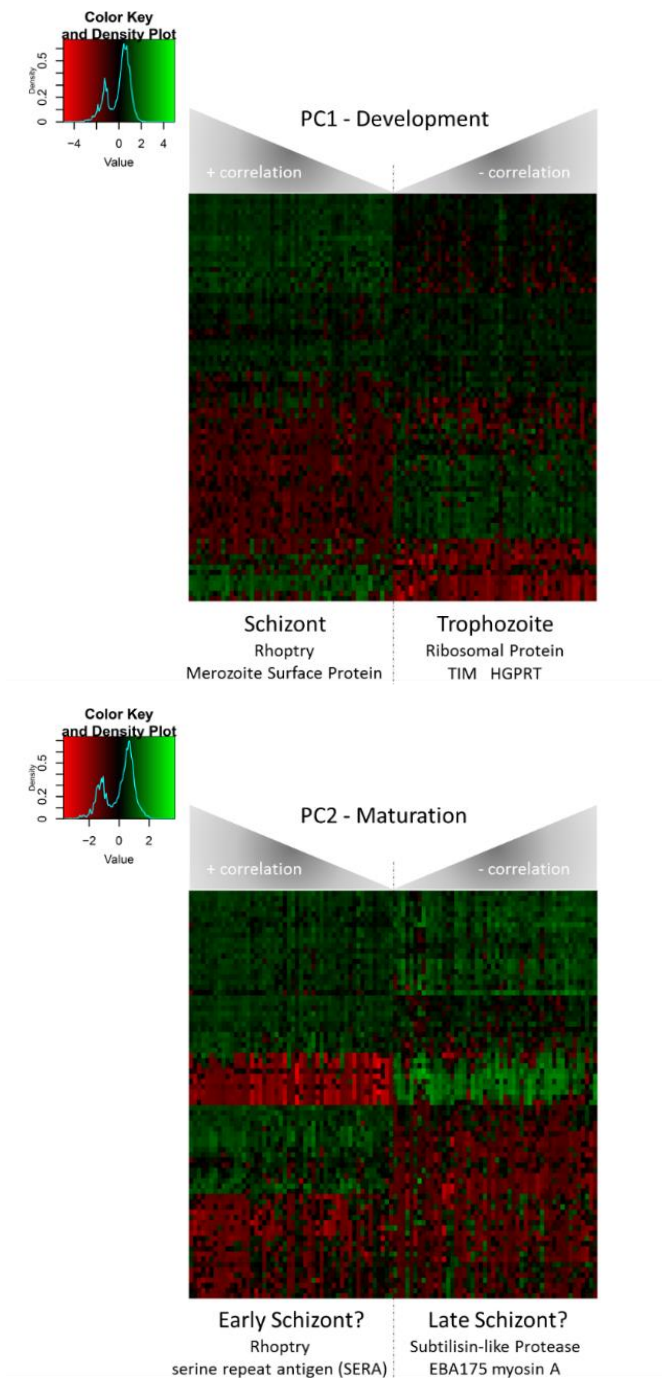


図 29. PC1、PC2 で因子負荷量が高い遺伝子で作成したヒートマップ

3D7 株，薬剤無添加のデータセットを用いて作成した．ヒートマップの両端が抽出した遺伝子(因子負荷量の絶対値が大きい遺伝子)の中で最も因子負荷量が高い遺伝子，中央が小さい遺伝子になるよう配置した(X軸)．Y軸が各細胞(計 78 細胞)でクラスタリングを行った．



図 30. 3D7 株、薬剤無添加データの主成分分析(変数：全遺伝子(5777))

各細胞の名称は，T_○がトロホゾイド同調培養株，S_○がシゾン同調培養株である。

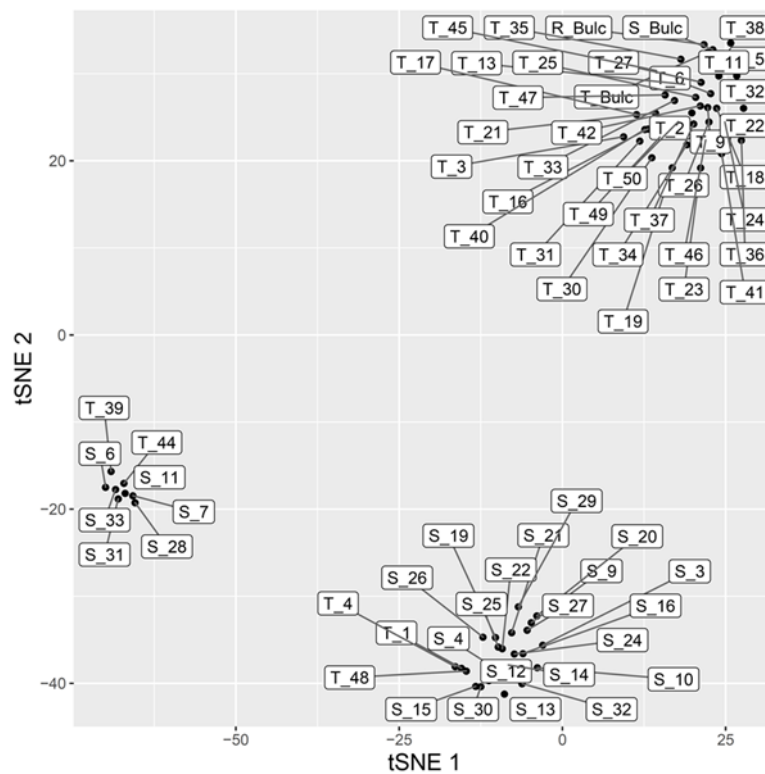


図 31. 同調培養株の tSNE 結果 (変数：全遺伝子(5777)).

3.3.2. マーカー遺伝子

トロホゾイド特異的遺伝子、シゾン特異的遺伝子それぞれ 10 遺伝子を選択してヒートマップを作成し、それぞれの細胞をクラスタリングした。

クラスターは、一部の細胞を除いて殆どがトロホゾイド同調培養株とシゾン同調培養株で分かれた (図 32)。また、シゾンのクラスターはさらに二つに分かれており、PfPKAc、RAP1、RAP2 が高発現しているクラスターと MSP6、EBA-175、EBL1、AMA1、PfSUB1、PfSUB2 が高発現しているクラスターに分かれた。これは一方が前期シゾン、もう一方が後期シゾンで分かれていることが考えられた。また、後期シゾンと考えられるクラスター(T_39 から S_28 の 8 細胞)と、主成分分析の Cluster3 は一致した(p64 参照)。

続いて、同調培養を行っていない培養株で同様のヒートマップを作製した(図 33)。培養株のデータは 3D7 株、薬剤無添加データ、3 回の実験から得た計 78 細胞を用いた。クラスターは同調培養株と同様に大きく分けて 3 つのクラスターで分かれたと考えられた。また、この結果は 3D7 株、薬剤無添加サンプル、3 つの duplicate を合わせた計 78 細胞のデータを用いた主成分分析を行った際のクラスタリング結果と似たものとなった(p68 参照)。このことから、Bozdech らがステージ特異的遺伝子として主張していた 20 遺伝子から生育ステージの推定を行うことができた(Bozdech *et. al.* 2003)。

しかし、生育ステージを分類する明確な基準を定めることが難しい。各遺伝子の Z-Score で規準を定めるとすると、データセット毎に定める必要があり、また規準を作る際の線引きを決める際等に恣意的になってしまう恐れがある。マーカー遺伝子によるステージ分類法は、その他分類法の裏付けとして使用する際に用いることとした。

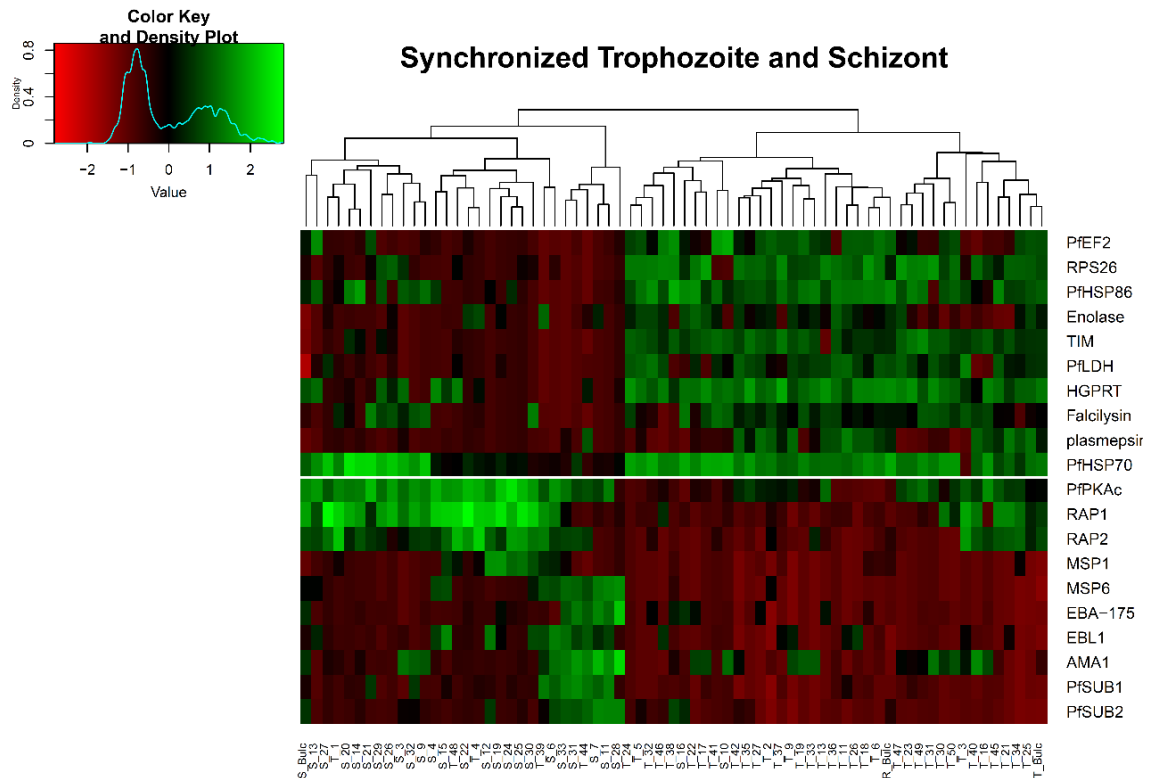


図 32. 同調培養株のマーカー遺伝子のヒートマップ

上から 10 遺伝子がトロホゾイド特異的遺伝子であり，下から 10 遺伝子がシズント特異的遺伝子である．各細胞の名称は，T_○がトロホゾイド同調培養株，S_○がシズント同調培養株である．

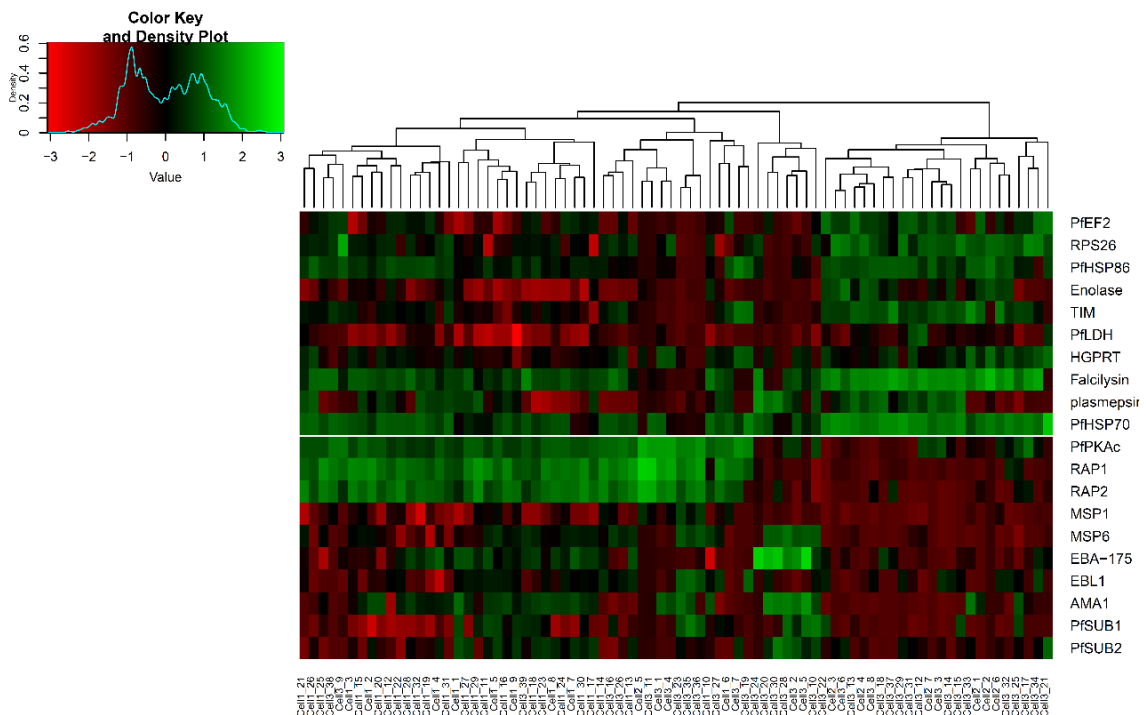


図 33. 3D7 株、薬剤無添加のマーカー遺伝子のヒートマップ

上から 10 遺伝子がトロホゾイド特異的遺伝子であり，下から 10 遺伝子がシズ
 ント特異的遺伝子である．各細胞の名称は，Cell1_○が 3D7 株、薬剤無添加
 1(duplicate1)，Cell2_○が 3D7 株、薬剤無添加 2(duplicate2)，Cell3_○が 3D7
 株、薬剤無添加 3(duplicate3)．

3.3.3. 公共データセットを用いた生育ステージの同定

血液ステージの熱帯熱マラリア原虫同調培養株から、複数の hpi でバルク RNA-Seq を行って取得した公共データセットの FASTQ ファイルをインターネット上より入手した。公共データの中で、血液ステージの熱帯熱マラリア原虫同調培養株から多くの時間点(hpi)で RNA-Seq を行っていた GSE23865 と PRJEB2015 を本研究に用いた (表 26,27)。

このバルク RNA-Seq の公共データセット(FASTQ ファイル)からそれぞれの RPKM を算出して、複数の hpi と各 1 赤血球 RNA-Seq データの相関係数を算出することで、各 1 赤血球 RNA-Seq データの生育ステージの予測を試みた。

GSE23865、PRJEB2015 の FASTQ ファイルをダウンロードし、RPKM を算出し、公共データセットと各実験データの全遺伝子(5777 遺伝子)を変数として、 $\log(\text{RPKM}+1)$ からピアソン相関係数を算出した。

続いて、本研究で使用した公共データセット同士が相関を持つかについて検証した。2つの公共データセット(PRJEB2015、GSE23865)で全遺伝子 5777 を変数として算出した RPKM 同士でピアソン相関係数を算出した。その結果、hpi が近いデータセット同士では高い相関を示し、一方で hpi が遠いデータセット同士では低い相関を示した(図 34)。このことから、2つの公共データセット間には強い相関が存在することが示された。よって、この 2つの公共データセット(PRJEB2015、GSE23865)を併用して各種解析に使用することとした。

続いて、実験データと公共データセットの相関を算出することで生育段階を明らかにすることができるかを検証した。

本研究で実験より取得した同調培養株(リング、トロホゾイド、シゾン)のバルク RNA-Seq データと公共データセットの全遺伝子(5777 遺伝子)を変数とし

て、 $\log(\text{RPKM}+1)$ からピアソン相関係数を算出した。

その結果、各同調培養株のバルク RNA-Seq データセットと最も高い相関を示す公共データセットの hpi が生育ステージに準じて変化することを確認した(図 35)。

GSE23865 のデータセットと比較した結果、リング同調培養株との相関の最高値は 25hpi、トロホゾイド同調培養株との相関の最高値は 40hpi、シズント同調培養株との相関の最高値は 40hpi であった。また、PRJEB2015 のデータセットと比較した結果、リング同調培養株との相関の最高値は 16hpi、トロホゾイド同調培養株との相関の最高値は 32hpi、シズント同調培養株との相関の最高値は 40hpi であった。

生育ステージがリング、トロホゾイド、シズントと進むにつれ、そのステージとの相関の最高値(hpi)が後期段階に向けて進んでいることから、実験データと公共データセットとの相関を見ることで、生育ステージの同定を行うことができる可能性が示唆された。

トロホゾイド及びシズント同調培養株、計 66 細胞を用いて主成分分析によるステージ分類法を検討した際、同調培養株の 1 赤血球 RNA-Seq データを用いて行った。その結果の PC1 と PC2 で K-means Clustering を行い 3 つのクラスターに分類した(p64 参照)。

さらに、マーカー遺伝子から 3 つに分類したそれぞれのクラスターがトロホゾイド、初期シズント、後期シズントで分かれていることが仮定された(p70 参照)。この結果と、公共データセットと 1 赤血球 RNA-Seq データの相関から生育ステージを推測した際に、主成分分析の K-means Clustering 結果と一致するかを検証した。

初めに、実験データと GSE23865 の各データを比較した。実験データと

GSE23865 の各データ(hpi)の相関係数の最高値が 30 または 35hpi の 1 赤血球 RNA-Seq データをトロホゾイド、実験データと GSE23865 の相関の最高値が 40hpi の 1 赤血球 RNA-Seq データを前期シズント、実験データと GSE23865 の相関の最高値が 5hpi の 1 赤血球 RNA-Seq データを後期シズントと定義した。ここで定義した生育ステージと主成分分析(PC1 と PC2)の K-means Clustering(K=3)結果がどの程度一致するかを検証した。

その結果、全体で約 66%(44/66)の結果が一致した。Cluster 1 (トロホゾイド)では約 51%(18/35)、Cluster 2 (前期シズント)では、約 78%(18/23)、Cluster 3 (後期シズント)では 100%(8/8)が一致していた。

続いて、実験データと PRJEB2015 の各データを比較した。実験データと PRJEB2015 の各データ(hpi)の相関係数の最高値が 24 または 32hpi の 1 赤血球 RNA-Seq データをトロホゾイド、実験データと PRJEB2015 の相関の最高値が 40hpi の 1 赤血球 RNA-Seq データを前期シズント、実験データと PRJEB2015 の相関の最高値が 48hpi の 1 赤血球 RNA-Seq データを後期シズントと定義した。ここで定義した生育ステージと主成分分析(PC1 と PC2)の K-means Clustering(K=3)結果がどの程度一致するかを検証した。

その結果、全体で約 80%(53/66)の結果が一致した。Cluster 1 (トロホゾイド)では約 71%(25/35)、Cluster 2 (前期シズント)では、約 91%(21/23)、Cluster 3 (後期シズント)では約 88%(7/8)が一致していた。結果が一致しなかった 1 赤血球 RNA-Seq データは、全てトロホゾイドとシズントの境界に存在する細胞であった。

主成分分析と公共データセット(PRJEB2015)との相関を用いた手法で約 80%の結果が一致したことから、公共データセットを用いたこの手法によって生育ステージの分類を行うことが可能であると考えられた。

この手法を用いて、同調培養を行っていない実験データを使用して、生育ステージを推定することが可能であるか検証した。

78 個の 3D7 株、薬剤無添加の 1 赤血球 RNA-Seq データと公共データセット (GSE23865 と PRJEB2015) の相関係数を算出し、1 赤血球 RNA-Seq データ全体の可視化を行うために Z 変換してヒートマップを作成した(図 36,37)。

その結果、ヒートマップ上で各相関は生育ステージの最高値とその前後で比較的高い値を示すことが明らかとなった。また、どちらの公共データセット (GSE23865 と PRJEB2015) を用いても、非常に近いクラスタリング結果となった。このことから、同調培養を行っていない薬剤無添加株の実験データに対しても、この手法が利用できることが示唆された。

当初、GSE23865 の hpi40 と PRJEB2015 の hpi40 の RPKM データが入れ替わる誤りが生じた。この際、実験データと GSE23865 の各データ(hpi)の相関係数の最高値が 35hpi の 1 赤血球 RNA-Seq データをトロホゾイド、実験データと GSE23865 の相関の最高値が 40hpi かつ PRJEB2015 との相関の最高値が 40hpi の 1 赤血球 RNA-Seq データを前期シズント、実験データと GSE23865 の相関の最高値が 40hpi かつ PRJEB2015 との相関の最高値が 48hpi の 1 赤血球 RNA-Seq データ、または実験データと GSE23865 の相関の最高値が 5hpi の 1 赤血球 RNA-Seq データを後期シズントと定義した。ここで定義した生育ステージと主成分分析(PC1 と PC2)の K-means Clustering(K=3)結果がどの程度一致するかを検証した。

その結果、全体で約 90%(59/66)の結果が一致していた。Cluster 1 (トロホゾイド)では約 83%(29/35)、Cluster 2 (前期シズント)では、約 96%(22/23)、Cluster 3 (後期シズント)では 100%(8/8)が一致していた。

解析条件を調整することによって、より正しく生育ステージの分類が行える

ことが考えられるため、変数選択や相関係数の扱いについて検討していきたい。

表 26. 解析に使用した GSE23865 の詳細 (Bartfai *et. al.* 2010)

| Data | Point (hpi) | Strain | Seqencer | Contributors |
|-----------|-------------|--------|------------------------------------|------------------------|
| GSM588466 | 5 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |
| GSM588467 | 10 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |
| GSM588468 | 15 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |
| GSM588469 | 20 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |
| GSM588470 | 25 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |
| GSM588471 | 30 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |
| GSM588472 | 35 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |
| GSM588473 | 40 hpi | 3D7 | Illumina Genome Analyzer II (76bp) | Bartfai <i>et. al.</i> |

表 27. 解析に使用した PRJEB2015 の詳細 (Otto *et. al.* 2010)

| Data | Point (hpi) | Strain | Seqencer | Contributors |
|-------------|-------------|--------|---------------------------------|---------------------|
| ERR006180_1 | 0 hpi | 3D7 | Illumina Genome Analyzer (36bp) | Otto <i>et. al.</i> |
| ERR006179_1 | 8 hpi | 3D7 | Illumina Genome Analyzer (36bp) | Otto <i>et. al.</i> |
| ERR006178_1 | 16 hpi | 3D7 | Illumina Genome Analyzer (36bp) | Otto <i>et. al.</i> |
| ERR006177_1 | 24 hpi | 3D7 | Illumina Genome Analyzer (36bp) | Otto <i>et. al.</i> |
| ERR006181_1 | 32 hpi | 3D7 | Illumina Genome Analyzer (36bp) | Otto <i>et. al.</i> |
| ERR006182_1 | 40 hpi | 3D7 | Illumina Genome Analyzer (36bp) | Otto <i>et. al.</i> |
| ERR006183_1 | 48 hpi | 3D7 | Illumina Genome Analyzer (36bp) | Otto <i>et. al.</i> |

| | hpi 0 | hpi 8 | hpi 16 | hpi 24 | hpi 32 | hpi 40 | hpi 48 |
|--------|-------|-------|--------|--------|--------|--------|--------|
| hpi 5 | 0.820 | 0.804 | 0.691 | 0.593 | 0.531 | 0.655 | 0.711 |
| hpi 10 | 0.835 | 0.878 | 0.824 | 0.700 | 0.563 | 0.570 | 0.635 |
| hpi 15 | 0.787 | 0.864 | 0.873 | 0.750 | 0.572 | 0.490 | 0.545 |
| hpi 20 | 0.750 | 0.826 | 0.882 | 0.809 | 0.619 | 0.485 | 0.519 |
| hpi 25 | 0.719 | 0.762 | 0.838 | 0.860 | 0.714 | 0.533 | 0.526 |
| hpi 30 | 0.717 | 0.720 | 0.776 | 0.851 | 0.794 | 0.612 | 0.568 |
| hpi 35 | 0.732 | 0.698 | 0.723 | 0.799 | 0.825 | 0.698 | 0.614 |
| hpi 40 | 0.760 | 0.694 | 0.662 | 0.722 | 0.810 | 0.814 | 0.721 |

図 34. 公共データセット同士(GSE23865 と PRJEB2015)の相関係数
X 軸が PRJEB2015 の hpi 0~48 の RNA-Seq データ, Y 軸が GSE23865 の hpi
5~40 の RNA-Seq データ.

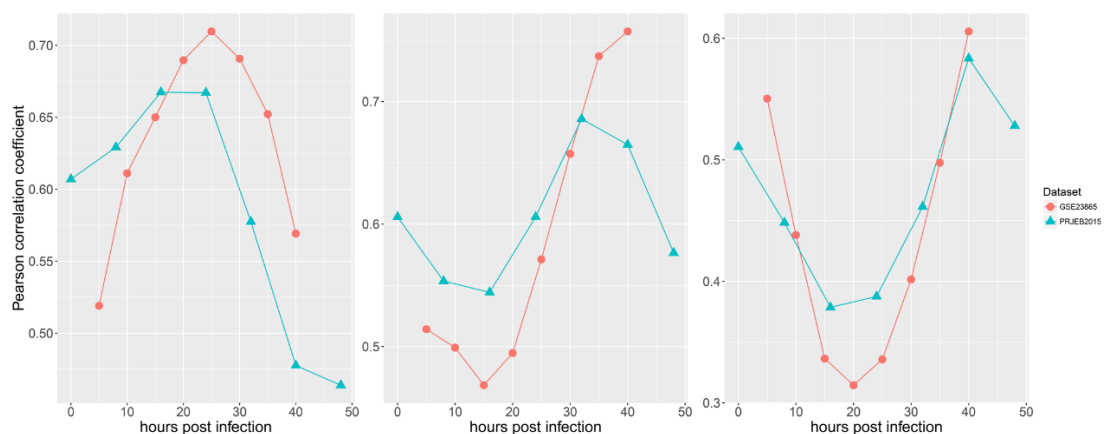


図 35.同調培養株と各公共データセットの相関係数

X 軸が hpi(hour post invation), Y 軸がピアソン相関係数. 赤が GSE23865 と実験データの相関, 青が PRJEB2015 と実験データの相関. (左)リング同調培養株, (中) トロホゾイド同調培養株, (右)シizont同調培養株.

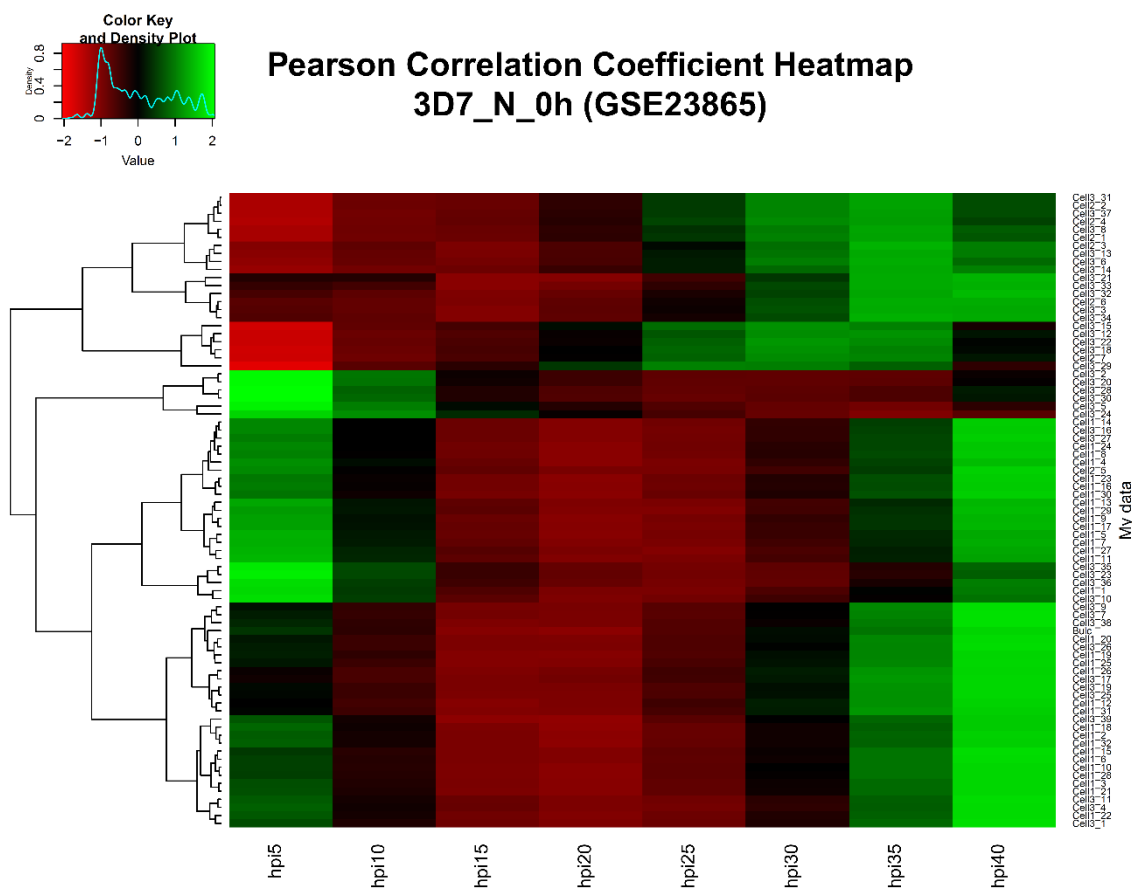


図 36. 実験データと公共データセット(GSE23865)の相関係数のヒートマップ

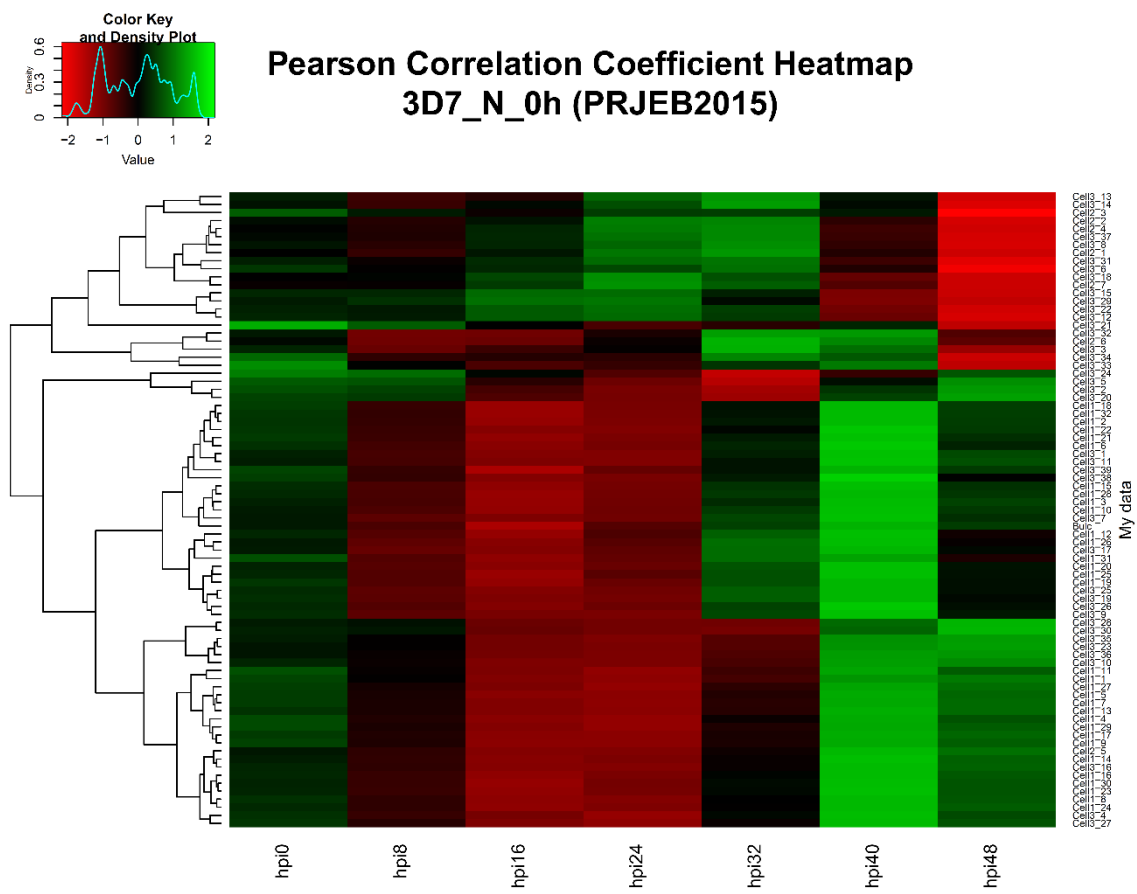


図 37. 実験データと公共データセット(PRJEB2015)の相関係数のヒートマップ

3.4.1 赤血球トランスクリプトーム解析

取得した 1 赤血球 RNA-Seq データから、各データにおけるトランスクリプトーム多様性や薬剤添加後の多様性変動等の観測を試みた。

各データセット中の発現変動の指標の一つとして、遺伝子発現数と分散を利用できると考え、値を算出した上で考察を行った。その結果、遺伝子発現数はその生育ステージによって大きく変動する可能性が示唆された。この結果を受け更なる考察を行った。

続いて、遺伝子発現変動を 1 赤血球レベルの観測を試みた。生育ステージを分類した上で、主成分分析から遺伝子発現変動を観測できるかを検証した。

薬剤耐性関連遺伝子とされている遺伝子に注目して、その変動を観測し考察を行った。

3.4.1. 遺伝子発現数と分散

各データセット中の発現変動の指標の一つとして、遺伝子発現数と分散を利用できると考えた。しかし、トランスクリプトームの変動に生育ステージが大きく寄与することが明らかとなっている(Bozdech *et. al.* 2003)。そのため、まず初めに生育ステージと遺伝子発現数の関係性を解析した。

本研究で取得したトロホゾイドとシズント同調培養株の 1 赤血球 RNA-Seq データ(及びバルク RNA-Seq データ)の PC1 と PC2 の散布図に遺伝子発現数を対応させた(図 38)。その結果、各クラスターの遺伝子発現数は Cluster1(トロホゾイド)では多く、Cluster3(後期シズント)では少ないことが明らかとなった。

遺伝子発現数は Cluster1(トロホゾイド)で 1600~2746 遺伝子、Cluster2(前

期シズント)で 684~1303 遺伝子、Cluster3(後期シズント)では 341~994 であった。一方、バルク RNA-Seq ではトロホゾイドは 3907、シズントは 3394 であった(図 39)。

Cluster1(トロホゾイド)と Cluster3(後期シズント)では、その遺伝子発現数に大きな開きがあり、Cluster2(前期シズント)はその中間に位置することが明らかとなった。トロホゾイドでは遺伝子発現数が多く、後期シズントでは遺伝子発現数が少ないという結果になった。シーケンスのデプスの差による影響も考えられたが、どのステージの株もタグ数に大きな開きはなく一定であった。

この結果から、後期ステージ(トロホゾイド~シズント)では、生育ステージが進むにつれて発現が特定の遺伝子に集中することが考えられた。

しかし、トロホゾイド同調培養株とシズント同調培養株間のバッチによる影響が懸念された。そのため、その他のデータセットで同様のことがいえるかを検証した。その結果、殆どのデータセットで同様の結果を示した(表 28)。薬剤無添加株とクロロキン添加株(図 40,41)、一部のジヒドロアルテミシニン添加株では生育ステージが進むにつれ、遺伝子発現数の減少が起きていた(図 42,43)。

続いて、公共データセットから算出した RPKM から、従来のバルク RNA-Seq における遺伝子発現数を計測した。その結果、遺伝子発現数と生育ステージの関連性を示すと考えることができるデータを得ることはできなかった(表 29,30)。

このことから、3つの可能性が考えられた。1つ目の可能性は後期シズントの遺伝子発現は一定の遺伝子に集中し、その発現は赤血球単位で多様であるという可能性、2つ目の可能性は後期シズントでは一定の遺伝子に発現が集中し、その発現は後期シズント中に細かく変動し続けるという可能性である。そして、3つ目の可能性は、シズントステージの細胞から Lysis、RT、PCR を行った際に実験的な問題が生じている可能性である。

上記で挙げた内、3つ目の可能性を検証するために、ヒトゲノムにマッピングされたリード数とヒトゲノムにマップされたリード数から、マラリア由来のリード RNA の絶対量を検定して差があるかについて検証した。

トロホゾイド同調培養株(トロホゾイド 35 株、前期シズント 3 株、後期シズント 2 株)の各ステージでヒトゲノムにマップされたリード/マラリアにマップされたリードで割合を算出した。その結果、各ステージの割合はトロホゾイドで 2.9%、前期シズントで 2.8%、後期シズントで 3.0%であり、ほぼ違いがないことが明らかとなった。よって、生育ステージによって抽出できるマラリア由来のリードに差が出る可能性は低いと考えた。

また、IGV(Integrative Genomics Viewer)(Robinson *et.al.* 2011)で SAM ファイルを可視化して、PCR Sister 存在について観測することとした。各データの中でマップされたリード数の多い、histone H2B (PF3D7_1105100)にマッピングされたシークエンスタグを IGV にて可視化して確認した。その結果、どの生育ステージの株も、遺伝子配列の中で広域にマッピングされていることが明らかとなった(図 44)。このことから、PCR によるバイアスが影響している可能性は考えにくい。

続いて、各クラスターの分散を調べた。その結果として、後期シズントの Cluster3 では分散が大きい遺伝子が多く存在することが明らかとなった(図 46)。しかし、Cluster3 は遺伝子発現数が少ないことから、発現している遺伝子の発現量が大きくなる傾向があり、分散が振れやすい可能性が懸念された。

そのため、変動係数(標準偏差/平均 RPKM)を算出して、ヒストグラムを作成した(図 47)。その結果、Cluster1 では変動係数が大きい遺伝子から小さい遺伝子まで存在し、Cluster2 では中間の遺伝子が殆どを占め、Cluster3 では小さい遺伝子が殆どを占めるという結果となった。ただし、この結果は各クラスター

の細胞数が影響するため、統計学的検定手法を用いて再度検討する必要がある。

以上の結果からも、遺伝子発現数は生育ステージに大きく影響し、そのバラつきは生育ステージが進むにつれ、収束していくことが考えられた。

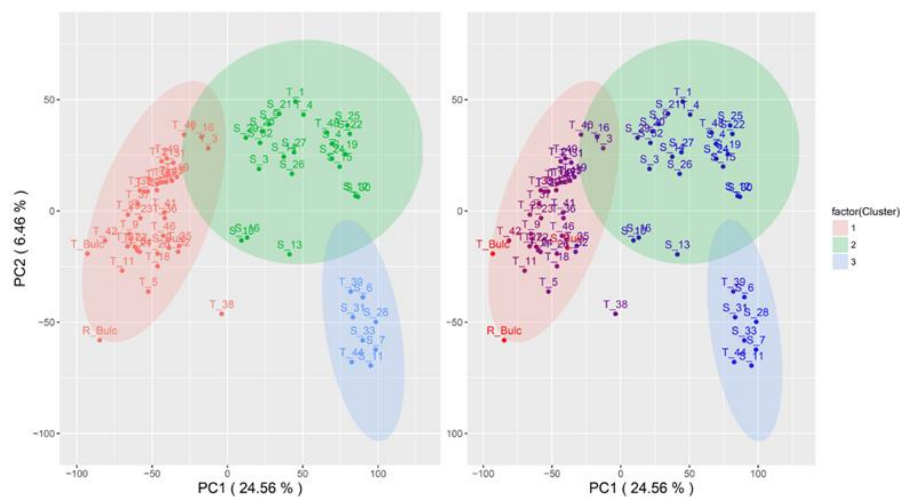


図 38. 同調培養株の主成分分析結果と遺伝子発現数の関係性

遺伝子発現数が最も多い細胞を赤，最も少ない細胞を青で表した。

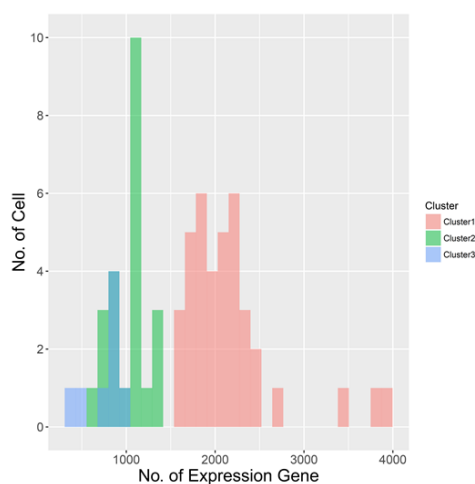


図 39. 同調培養株の遺伝子発現数

3000-4000 遺伝子間に存在する 3 細胞はリング，トロホゾイド，シズントそれぞれのバルク RNA-Seq.

表 28. 各データセットのステージ別の平均遺伝子発現数(RPKM \geq 5)

| Dataset | PRJEB2015 | | | GSE23865 | | |
|-------------------|-----------|-----------|----------|-----------|-----------|-----------|
| | 32 | 40 | 48 | 35 | 40 | 5 |
| 3D7_N_0h (1) | NA | 2678 (31) | NA | NA | 2714 (27) | 2439 (4) |
| 3D7_N_0h (2) | 2398 (3) | 1218 (1) | NA | 2099 (3) | 1899 (2) | NA |
| 3D7_N_0h (3) | 2058 (9) | 1955 (14) | 969 (7) | 2069 (9) | 2061 (15) | 994 (10) |
| 3D7_CQ_6h | 1143 (6) | 1071 (31) | 430 (1) | 1084 (4) | 1156 (26) | 718 (9) |
| 3D7_CQ_12h | 1962 (4) | 1378 (18) | 751 (2) | 1641 (10) | 1365 (16) | 752 (2) |
| 3D7_CQ_24h | 1475 (5) | 1063 (16) | 621 (4) | 1475 (5) | 1092 (13) | 737 (8) |
| 7G8_N_0h | 2033 (6) | 1894 (31) | 850 (2) | 1819 (1) | 1977 (28) | 1475 (11) |
| 7G8_CQ_6h | 1712 (6) | 1096 (29) | 662 (7) | 1575 (4) | 1253 (24) | 737 (14) |
| 7G8_CQ_12h | 1319 (3) | 1114 (10) | 662 (2) | 1516 (5) | 1205 (8) | 731 (7) |
| 7G8_CQ_24h | 1294 (2) | 1178 (21) | 768 (2) | 1304 (3) | 1297 (16) | 904 (8) |
| K1_N_0h | 1605 (11) | 1405 (21) | 1179 (1) | 1521 (10) | 1496 (22) | 1136 (3) |
| K1_CQ_6h | 1392 (5) | 1146 (30) | 702 (7) | 1156 (5) | 1250 (24) | 804 (15) |
| K1_CQ_12h | 2029 (6) | 1509 (19) | 821 (4) | 2083 (3) | 1604 (20) | 966 (6) |
| K1_CQ_24h | 1155 (1) | 1441 (14) | 800 (6) | 1519 (2) | 1466 (11) | 983 (9) |
| Dd2_N_0h | 1309 (2) | 1173 (19) | 861 (11) | 1288 (4) | 1225 (16) | 876 (17) |
| Dd2_CQ_6h | 1891 (2) | 1401 (18) | 801 (1) | 2183 (1) | 1530 (14) | 1033 (6) |
| Dd2_CQ_12h | 2022 (1) | 1154 (21) | 826 (5) | 1734 (2) | 1250 (16) | 904 (11) |
| Dd2_CQ_24h | 1386 (2) | 1351 (14) | 687 (1) | 1360 (8) | 1411 (10) | 780 (2) |
| HB3_N_0h | 1755 (1) | 1444 (12) | 1344 (2) | 1674 (2) | 1438 (9) | 1349 (6) |
| HB3_CQ_6h | 1239 (20) | 1189 (16) | NA | 1158 (23) | 1216 (20) | 713 (2) |
| HB3_CQ_12h | 945 (3) | 1013 (12) | 610 (8) | 1091 (2) | 1021 (13) | 633 (9) |
| HB3_CQ_24h | 872 (3) | 1068 (12) | 510 (2) | 1288 (2) | 1030 (15) | 446 (3) |
| Average | 1571 | 1360 | 781 | 1554 | 1453 | 958 |
| Total Cell Number | 101 | 410 | 75 | 108 | 365 | 162 |

()は細胞数.

表 29. GSE23865 の各 hpi の遺伝子発現数 (RPKM \geq 5)

| Data | hpi 5 | hpi 10 | hpi 15 | hpi 20 | hpi 25 | hpi 30 | hpi 35 | hpi 40 |
|---------------|-------|--------|--------|--------|--------|--------|--------|--------|
| RPKM \geq 5 | 3142 | 3113 | 2984 | 3185 | 3854 | 4212 | 4356 | 4741 |

表 30. PRJEB2015 の各 hpi の遺伝子発現数 (RPKM \geq 5)

| Data | hpi 0 | hpi 8 | hpi 16 | hpi 24 | hpi 32 | hpi 40 | hpi 48 |
|---------------|-------|-------|--------|--------|--------|--------|--------|
| RPKM \geq 5 | 4283 | 4218 | 2734 | 4321 | 4435 | 4264 | 4714 |

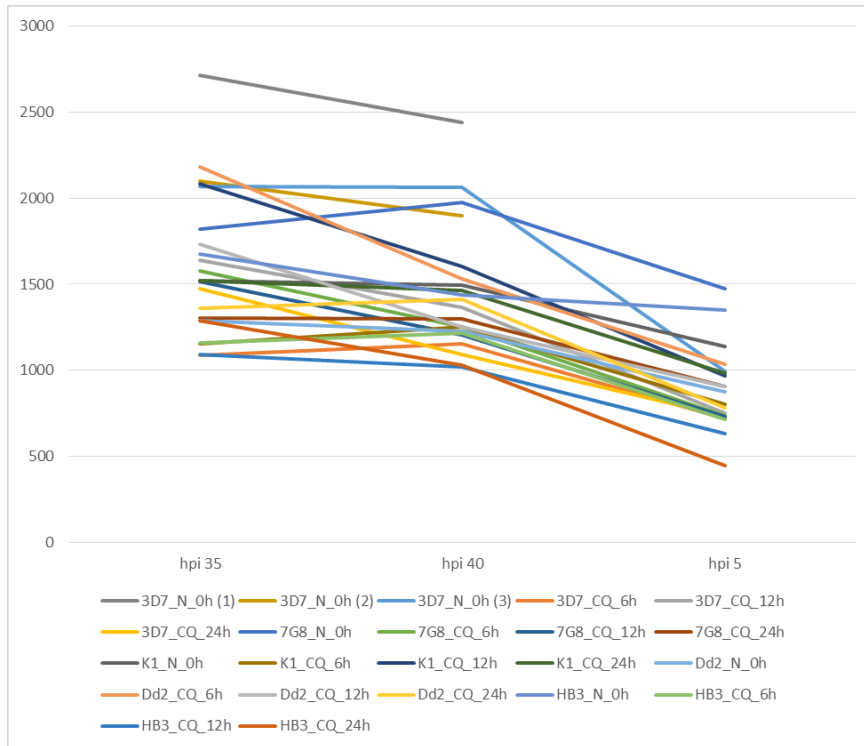


図 40. 各データセットのステージ別の平均遺伝子発現数

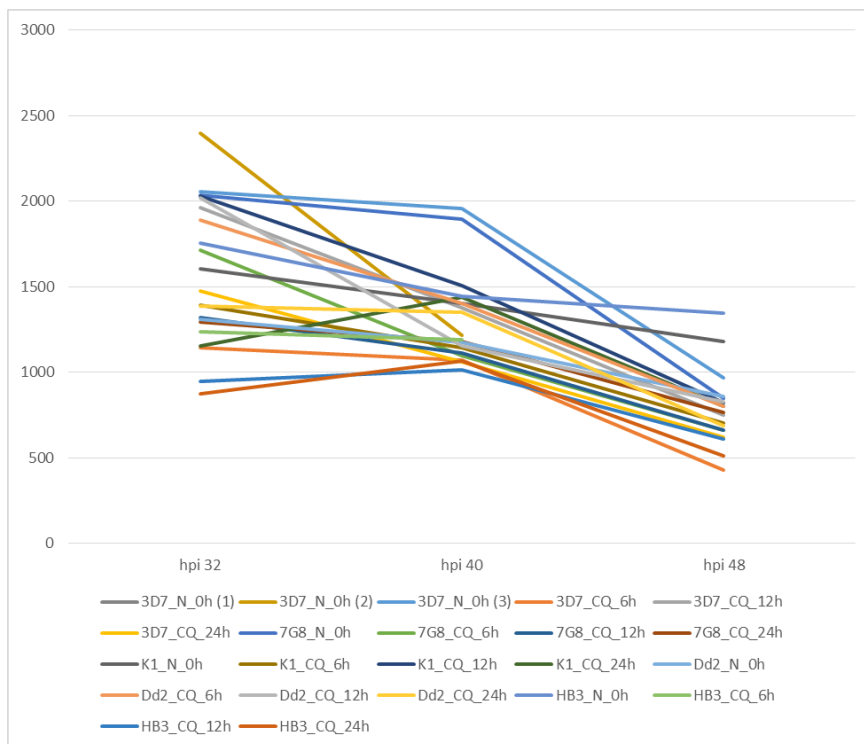


図 41. 各データセットのステージ別の平均遺伝子発現数

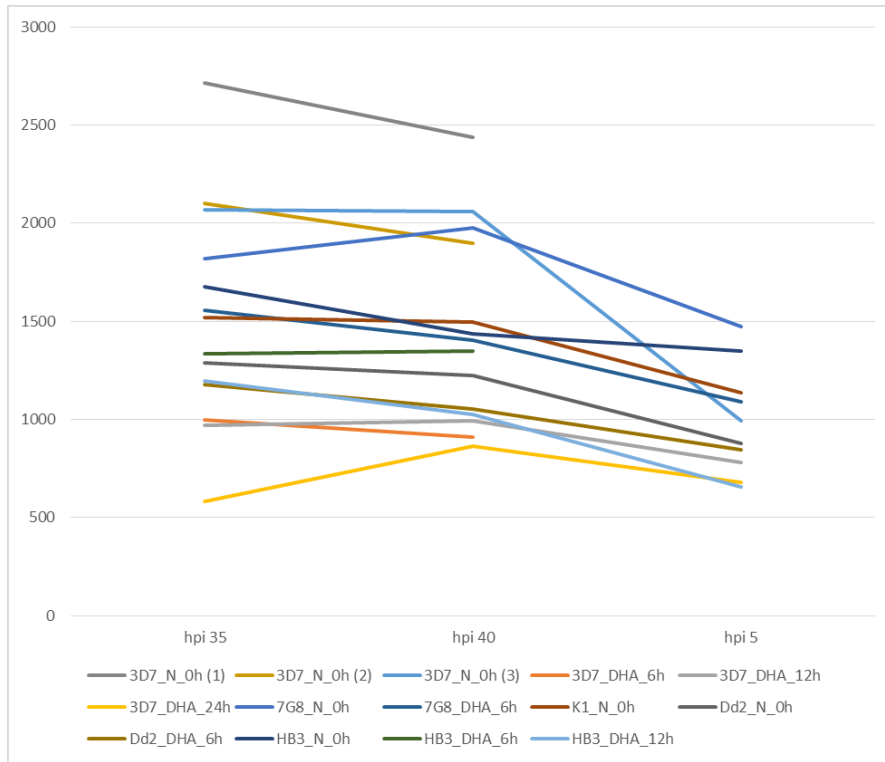


図 42. 各データセットのステージ別の平均遺伝子発現数

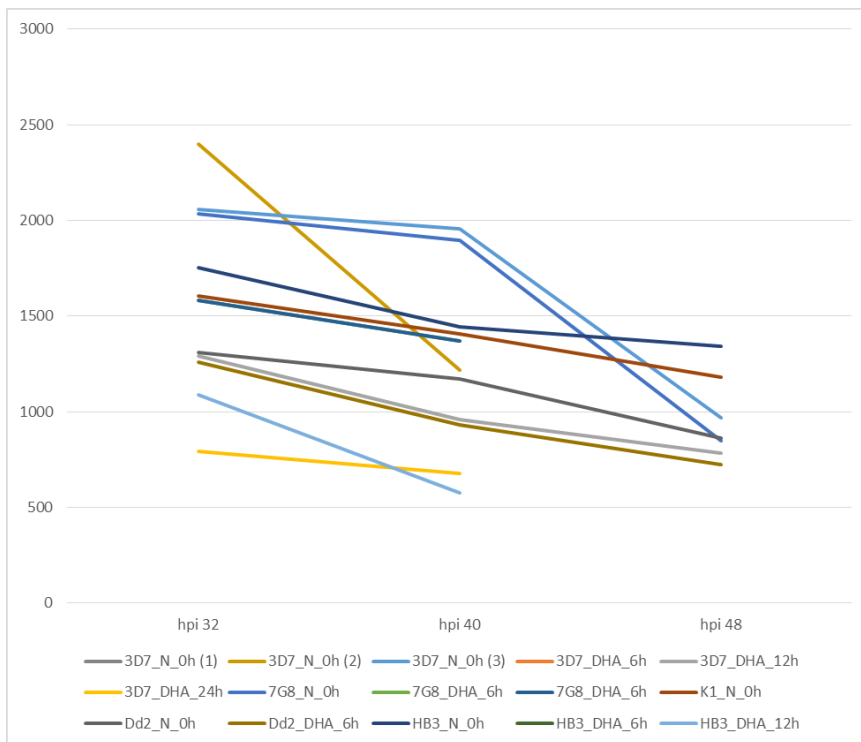


図 43. 各データセットのステージ別の平均遺伝子発現数

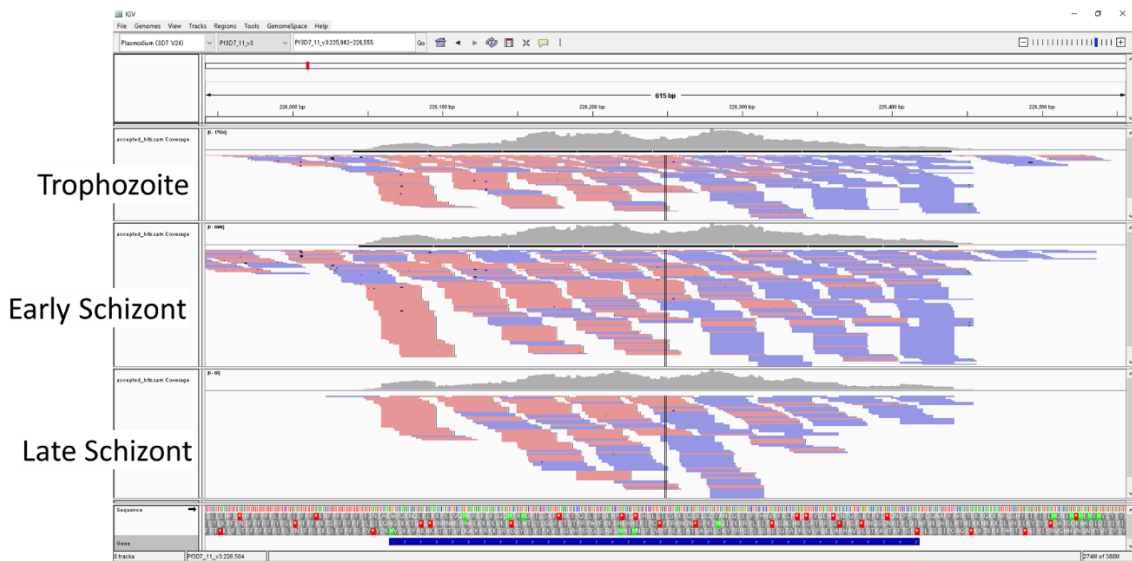


図 44. histone H2B (PF3D7_1105100)近傍にマッピングされたシーケンスタグ

上段：トロホゾイド株，中段：前期シズント株，下段：後期シズント株。

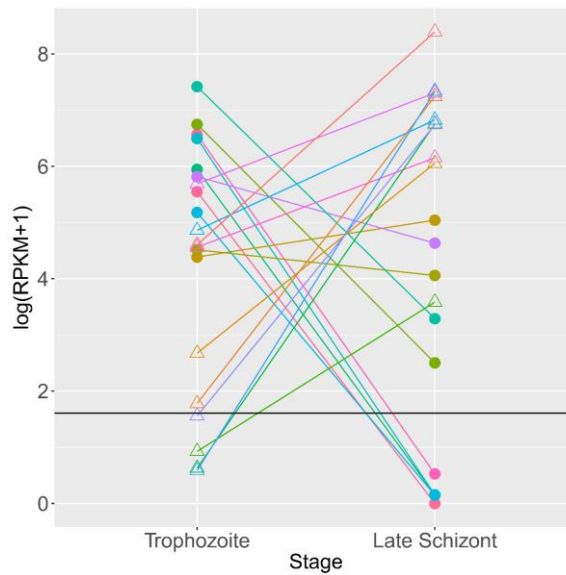


図 45. ステージ経過による発現変動

トロホゾイドとシズントの各マーカー遺伝子 10 ずつ(計 20 遺伝子)において，同調培養株 66 細胞で生育ステージ毎の RPKM 平均値を算出して変動を観測した。

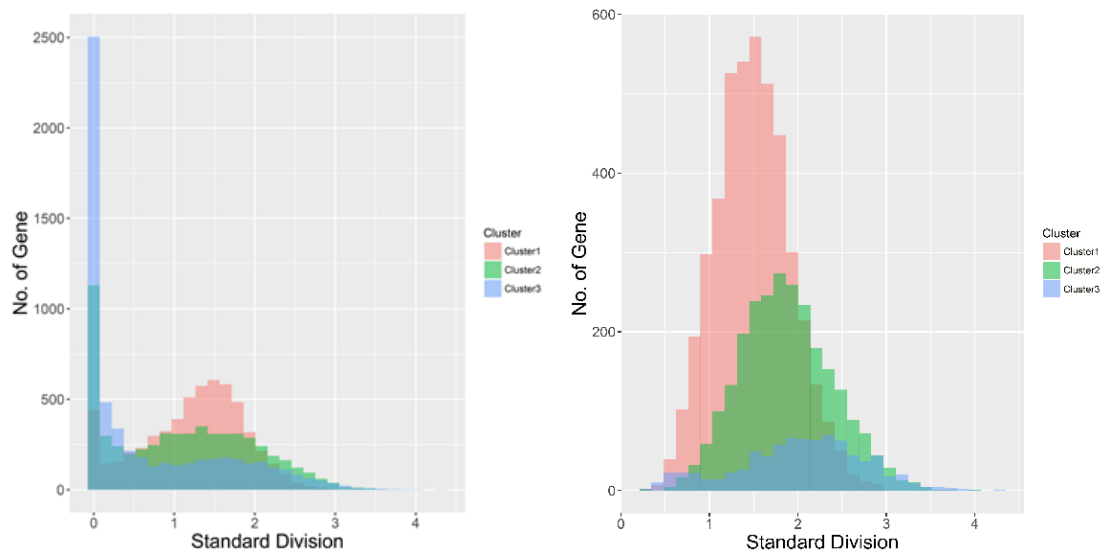


図 46. 各クラスターの分散のヒストグラム

(左)全遺伝子の分散, (右) 2 細胞以上で発現している遺伝子の分散.

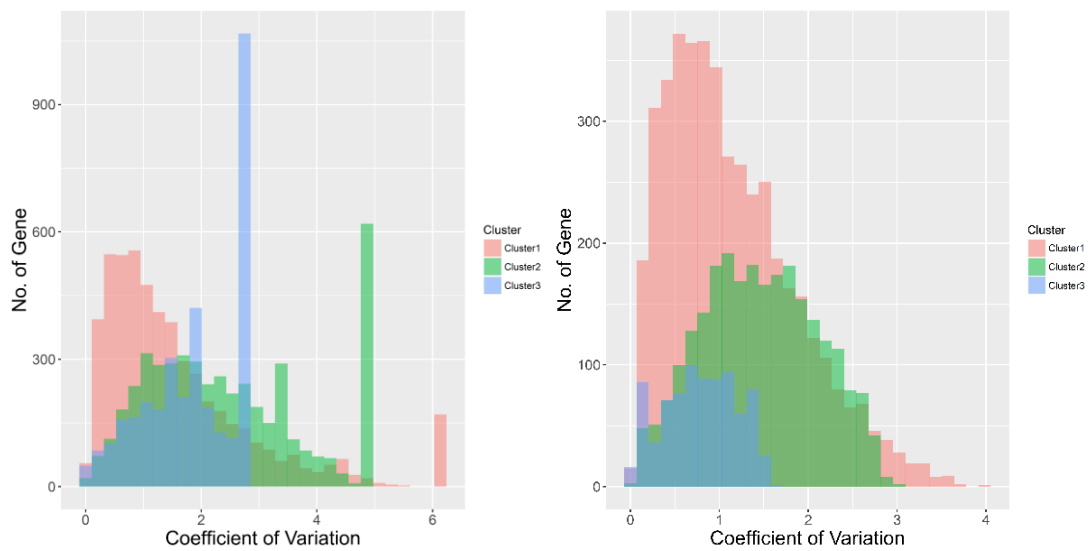


図 47. 各クラスターの変動係数のヒストグラム

(左)全遺伝子の変動係数, (右)2 細胞以上で発現している遺伝子の変動係数.

3.4.2. 遺伝子発現変動

薬剤無添加サンプル、または薬剤添加後時間が同一であるデータセットを 5 種の株でデータをまとめて主成分分析を行った。

薬剤無添加サンプルをまとめて主成分分析を行った結果、各細胞株の生育ステージの影響を大きく受けて分布したと考えられた。同様に、クロロキン、ジヒドロアルテミシニン添加後サンプルについても生育ステージの影響を大きく受けていると考えられた(図 48-52)。

その生育ステージの影響を受けながらもクロロキン添加後 6 時間の主成分分析結果(図 49)から、この条件下においては各細胞のトランスクリプトームに一定の方向性を持った変化を示していることが考えられた。

一方で、クロロキン添加後 12 時間(図 50)、24 時間の主成分分析結果(図 51)からは、クロロキン添加後 6 時間と同様である一定の方向性を持った分布は示されなかった。このことから本研究の実験条件において、薬剤添加後反応を確認するためには、薬剤無添加データとクロロキン添加後 6 時間データの比較が有効であると考えられた。

また、薬剤添加によって特定のトランスクリプトームに変動、または特定のトランスクリプトームを保持する集団のみが薬剤刺激に耐え、その後再び多様性を獲得する可能性が考えられた。

生育ステージの影響を大きく受けることから、生育ステージを分類した上で薬剤添加条件が同一であるデータセットでまとめて主成分分析を行った(図 53-57)。

クロロキン感受性株である 3D7 株、HB3 株とクロロキン耐性株である 7G8 株、Dd2 株、K1 株との間でトランスクリプトームに差が生じるかを観測したが、主

成分分析結果からは確証が得られなかった。各主成分分析結果を確認したが感受性株と耐性株の 2 群にデータが分かれることはなかった(図 54-56)。このことから、耐性の有無でクロロキン添加時のトランスクリプトームに差が存在する可能性は少なく、存在していたとしても大きい変動はないと考えられた。

続いて、生育ステージを分類した上で、株が同一であるデータセットでまとめて主成分分析を行った。薬剤添加前と添加後のトランスクリプトーム変動を比較するためである。

その結果、クロロキン添加によって 3D7、7G8、Dd2、K1 株である程度のトランスクリプトーム変動が存在することが考えられる結果が得られた(図 58,60-62)。この結果は、クロロキンのターゲットとされるトロホゾイド以外の初期シズント、後期シズントステージでも確認された(Slater 1993)。

同様にジヒドロアルテミシニン添加によっても、ある程度のトランスクリプトーム変動が存在することが考えられる結果を得ることができた(図 63-66)。PC1、PC2 の因子負荷量から、薬剤添加後株でどのような変化が起きているかを検討することができると考えた。今後、それぞれの結果について因子負荷量を含めて考察して、薬剤添加時の各データセットでどのような現象が起きているかについて理解を深めていきたい。特に、アルテミシニンは PfPI3K 阻害剤であることを踏まえて解析を進めていくべきであると考えた(Mbengue *et.al.* 2015) (図 67)。ジヒドロアルテミシニンがトランスクリプトーム変動を引き起こし、その影響によって生育ステージ分類が正しく行えていない可能性も存在する(Shaw *et.al.* 2015)。

主成分分析を用いたこの結果は、シーケンスのデプスによる影響がある可能性が考えられるため、各データのタグ数からその影響を調べる必要がある。

得られた結果の信頼性を検証した上で、データの解釈を行い、更なる解析へ

と繋げていきたい。

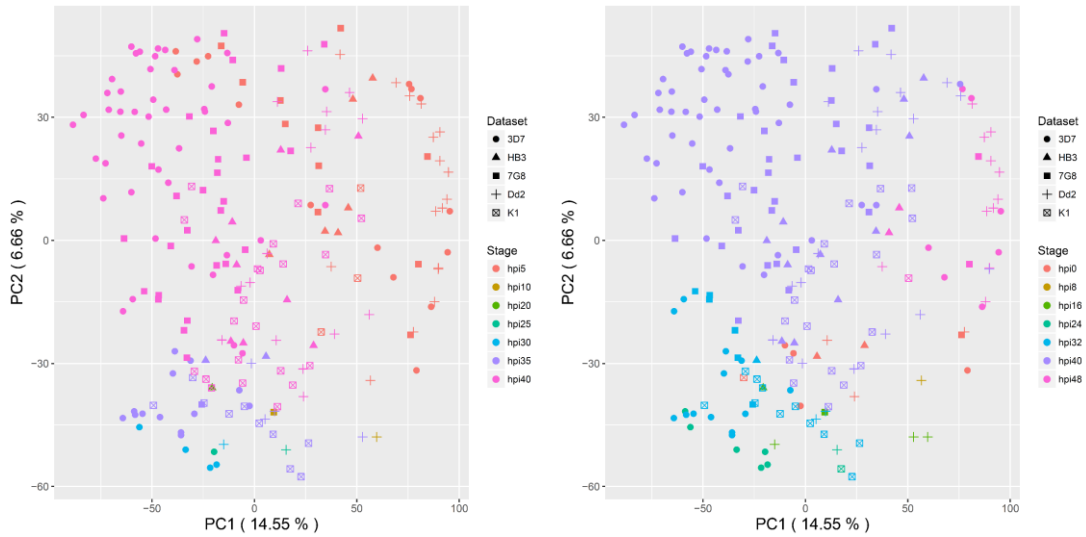


図 48. 主成分分析と生育ステージ (コントロール)

(左)GSE23865 参照, (右) PRJEB2015 参照.

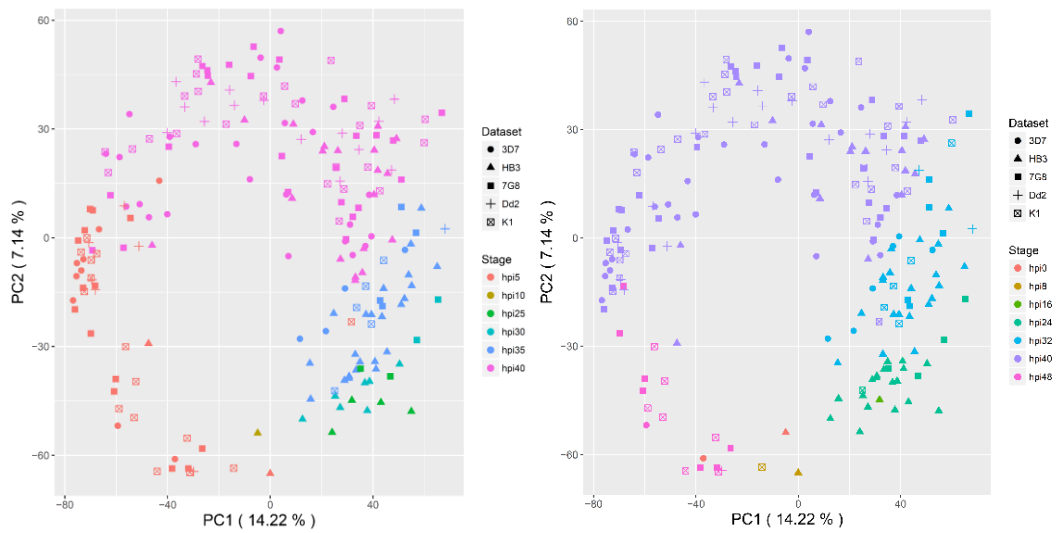


図 49. 主成分分析と生育ステージ (クロロキン添加後 6 時間)

(左)GSE23865 参照, (右) PRJEB2015 参照.

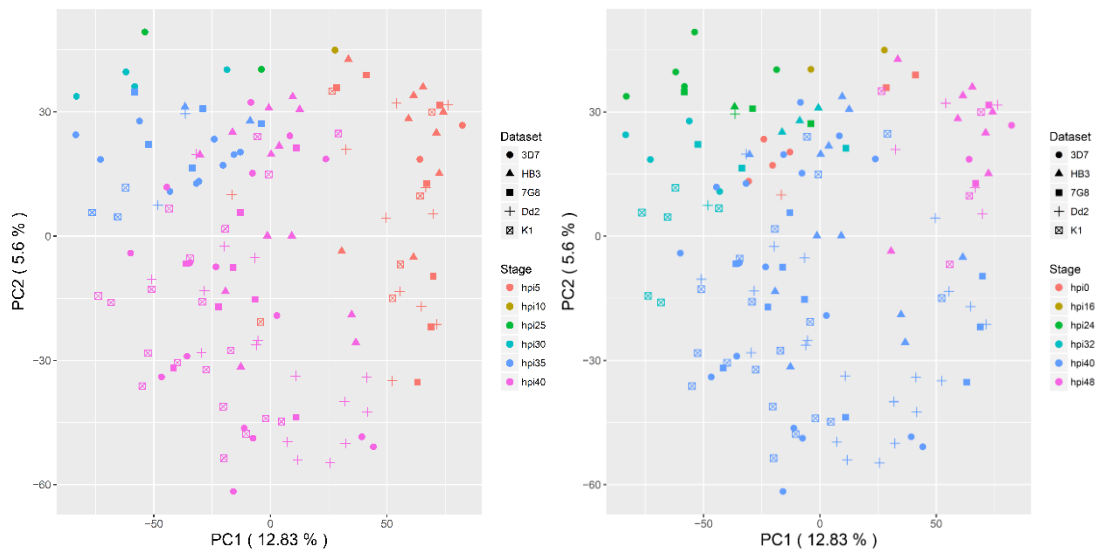


図 50. 主成分分析と生育ステージ (クロロキン添加後 12 時間)

(左) GSE23865 参照, (右) PRJEB2015 参照.

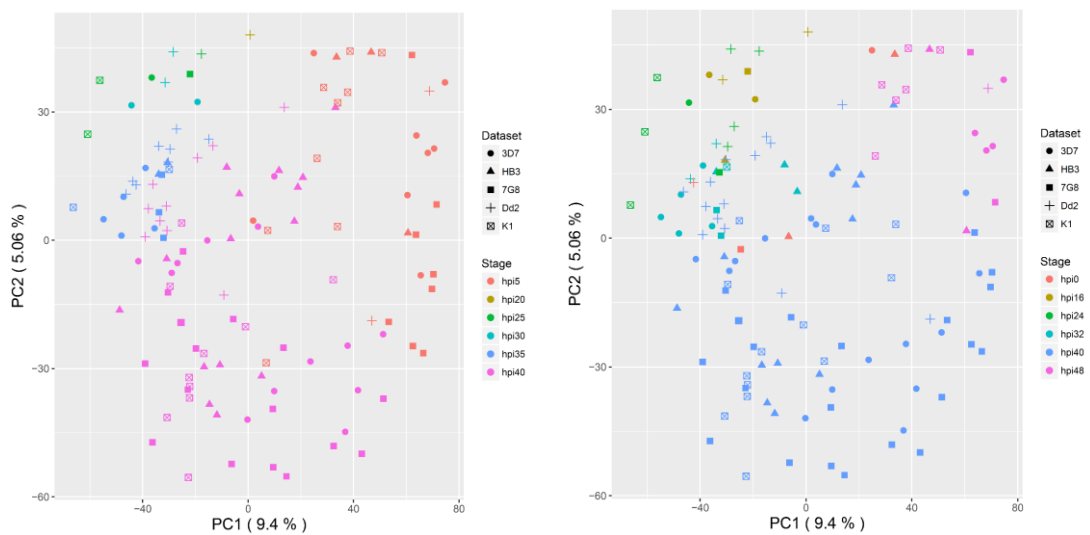


図 51. 主成分分析と生育ステージ (クロロキン添加後 24 時間)

(左) GSE23865 参照, (右) PRJEB2015 参照.

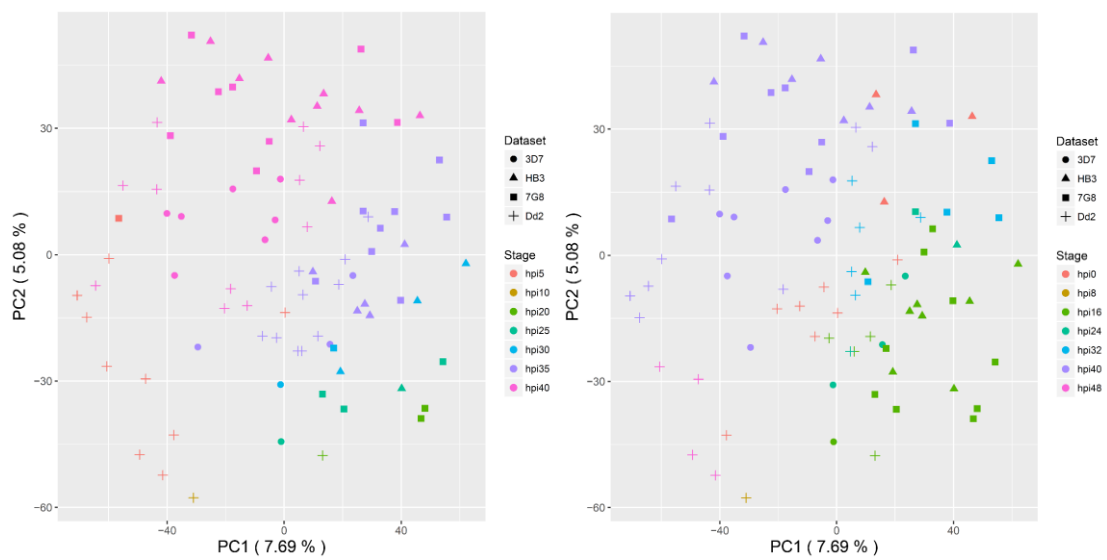


図 52. 主成分分析と生育ステージ (ジヒドロアルテミシニン添加後 6 時間)

(左) GSE23865 参照, (右) PRJEB2015 参照.

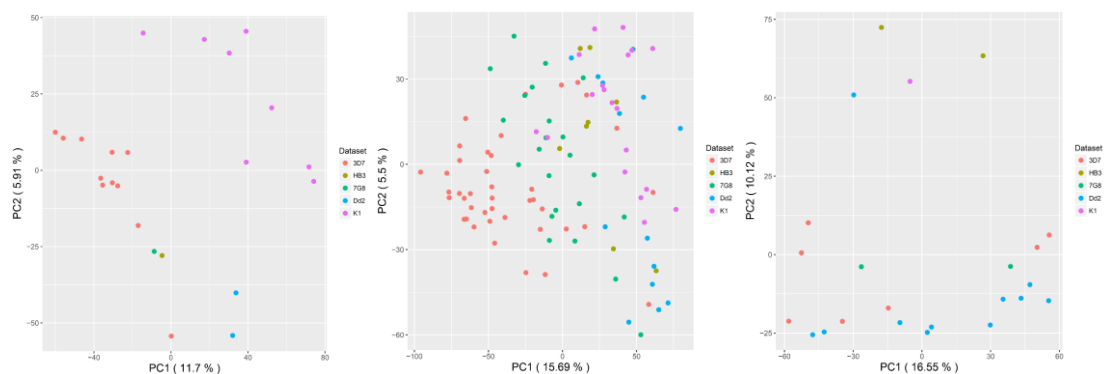


図 53. 主成分分析と生育ステージ (コントロール)

(左) 後期トロホゾイド株, (中) 前期シズント株, (右) 後期シズント株.

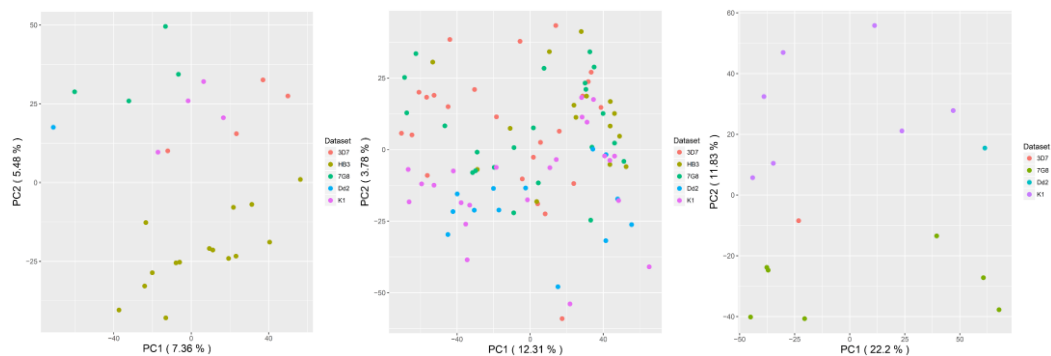


図 54. 主成分分析と生育ステージ (クロロキン添加後 6 時間)

(左) 後期トロホゾイド株, (中) 前期シズント株, (右) 後期シズント株.

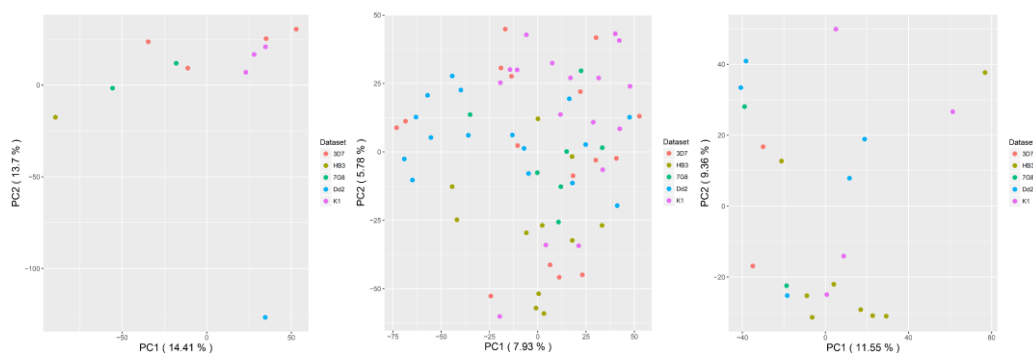


図 55. 主成分分析と生育ステージ (クロロキン添加後 12 時間)

(左) 後期トロホゾイド株, (中) 前期シズント株, (右) 後期シズント株.

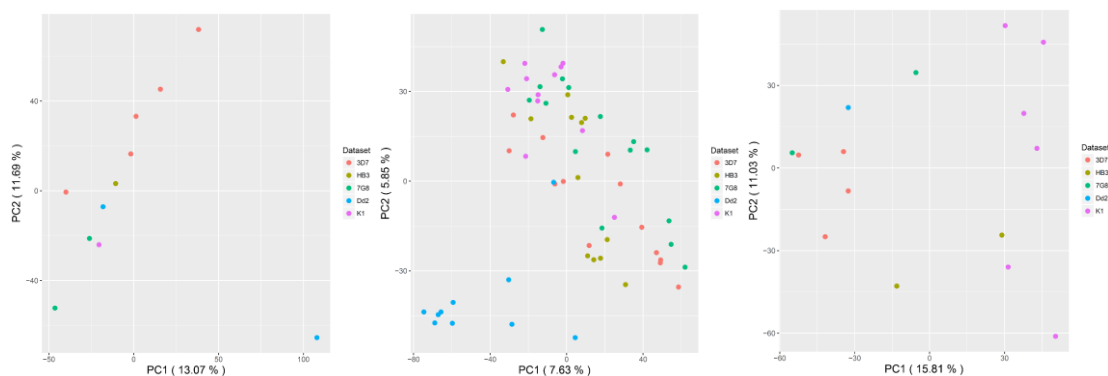


図 56. 主成分分析と生育ステージ (クロロキン添加後 24 時間)

(左) 後期トロホゾイド株, (中) 前期シゾント株, (右) 後期シゾント株.

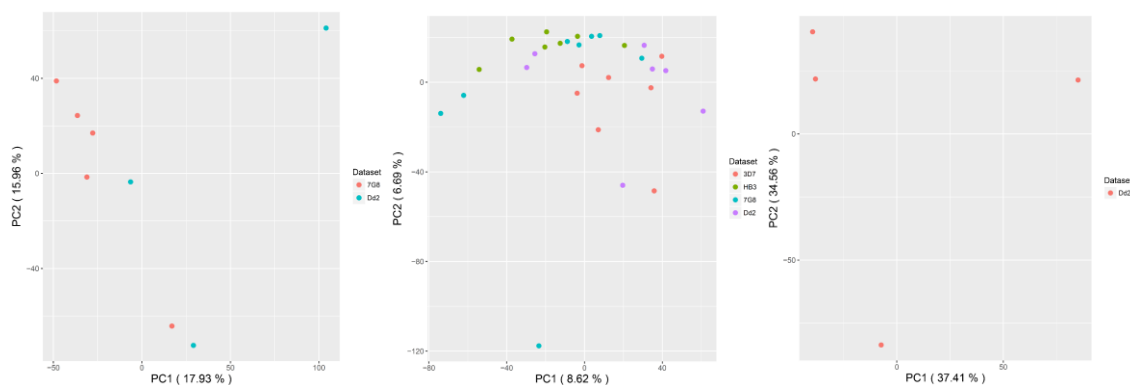


図 57. 主成分分析と生育ステージ (ジヒドロアルテミシニン添加後 6 時間)

(左) 後期トロホゾイド株, (中)前期シゾント株, (右) 後期シゾント株.

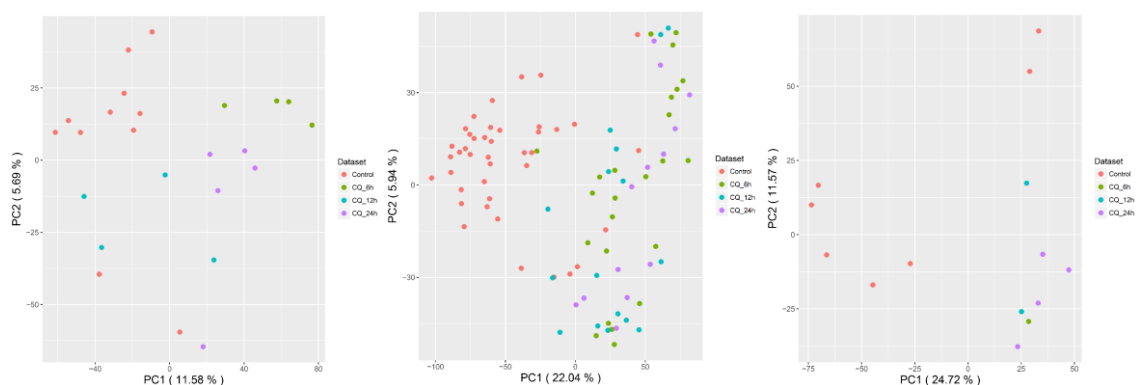


図 58. 3D7 株、薬剤無添加株とクロロキン添加株の主成分分析結果

(左) 後期トロホゾイド株, (中)前期シゾント株, (右) 後期シゾント株.

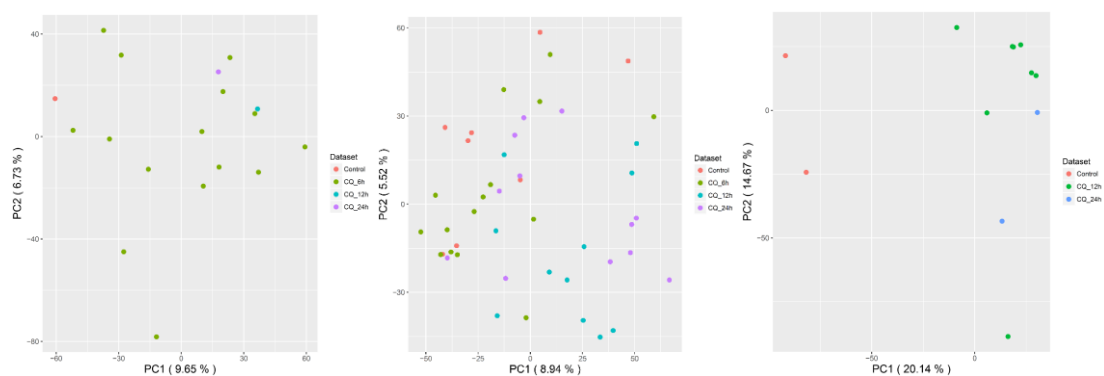


図 59. HB3 株、薬剤無添加株とクロロキン添加株の主成分分析結果
(左) 後期トロホゾイド株, (中)前期シズント株, (右) 後期シズント株.

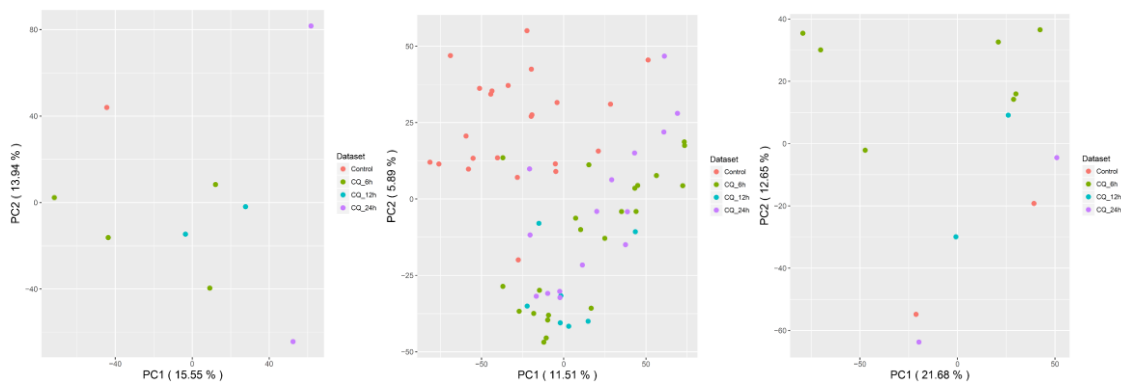


図 60. 7G8 株、薬剤無添加株とクロロキン添加株の主成分分析結果
(左) 後期トロホゾイド株, (中)前期シズント株, (右) 後期シズント株.

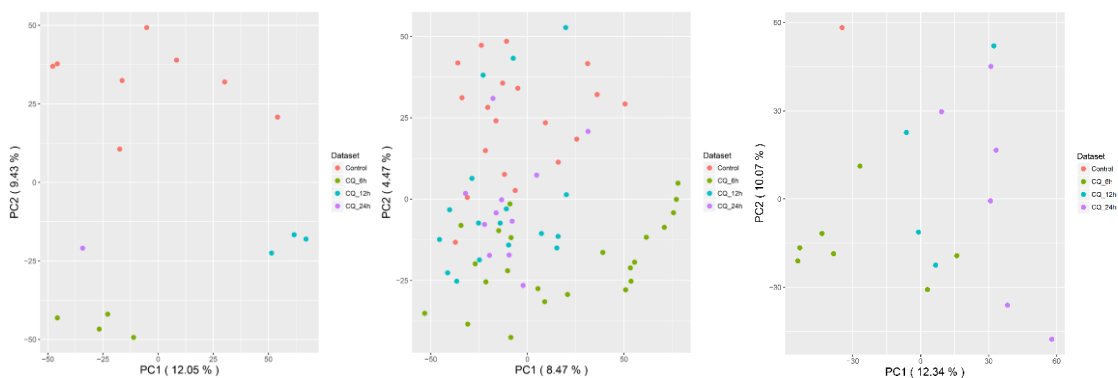


図 61. K1 株、薬剤無添加株とクロロキン添加株の主成分分析結果
(左) 後期トロホゾイド株, (中)前期シズント株, (右) 後期シズント株.

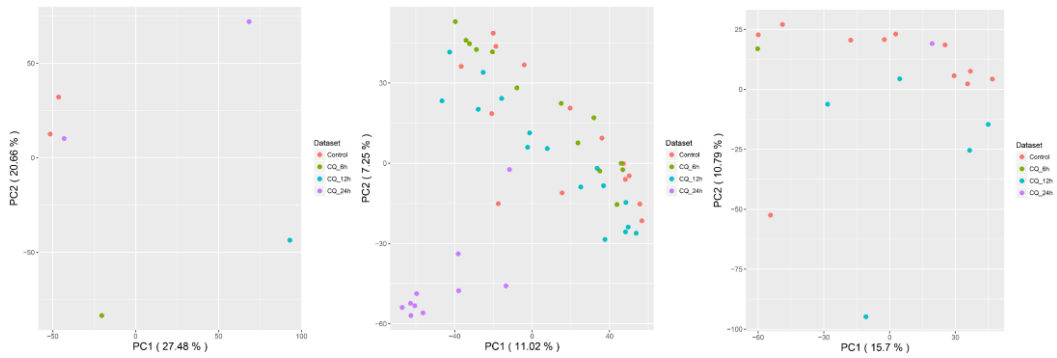


図 62. Dd2 株、薬剤無添加株とクロロキン添加株の主成分分析結果
 (左) 後期トロホゾイド株, (中)前期シizont株, (右) 後期シizont株.

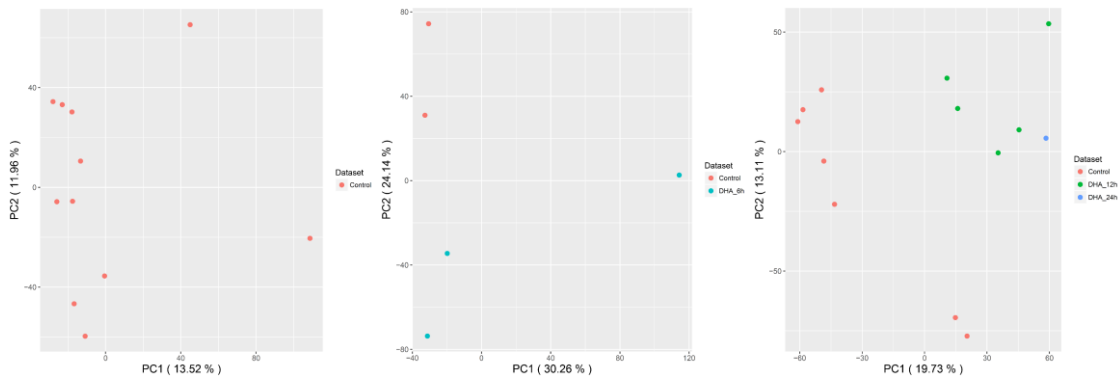


図 63. 3D7 株、薬剤無添加株とジヒドロアルテミシニン添加株の主成分分析結果
 (左) 後期トロホゾイド株, (中)前期シizont株, (右) 後期シizont株.

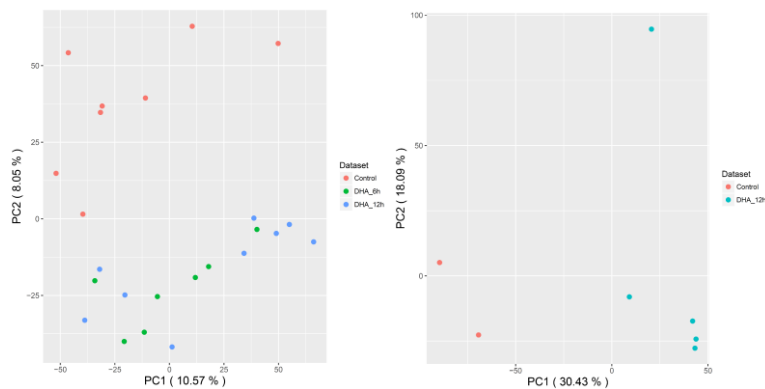


図 64. HB3 株、薬剤無添加株とジヒドロアルテミシニン添加株の主成分分析結果
 (左) 前期シizont株, (右) 後期シizont株.

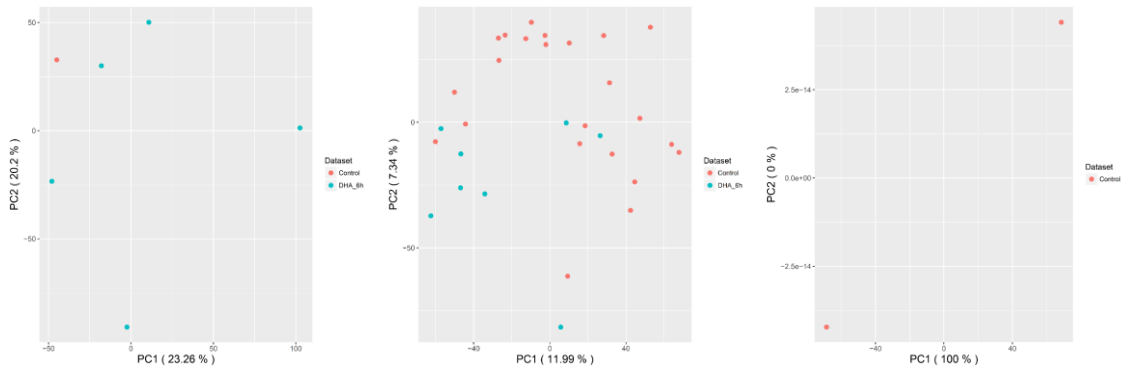


図 65. 7G8 株、薬剤無添加株とジヒドロアルテミシニン添加株の主成分分析結果
 (左) 後期トロホゾイド株, (中)前期シズント株, (右) 後期シズント株.

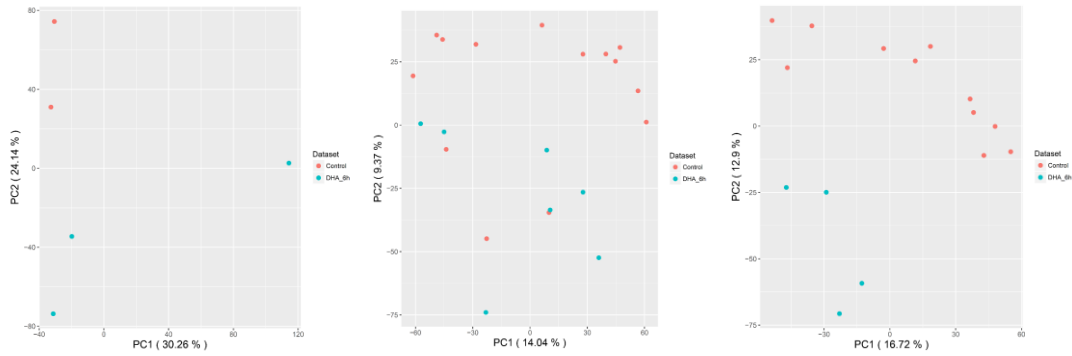


図 66. Dd2 株、薬剤無添加株とジヒドロアルテミシニン添加株の主成分分析結果
 (左) 後期トロホゾイド株, (中)前期シズント株, (右) 後期シズント株.

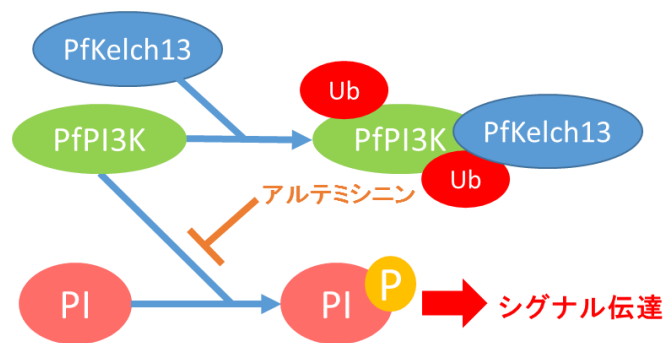


図 67. アルテミシニンの作用メカニズム

PfPI3K による PI のリン酸化を阻害してシグナル伝達に障害を引き起こす

(Mbengue *et.al.* 2015).

3.4.3. 薬剤耐性関連遺伝子の発現量観測

生育ステージを推測した上で、1赤血球単位で各株の薬剤耐性関連遺伝子の発現量をストリップチャートに表して観測した。その結果、各遺伝子はその細胞の生育ステージに必ずしも依存しているとは限らないと考えられる結果を示した。

作製したストリップチャートの一部、3D7、7G8株のCRTとMDR1の発現量について下記に載せた(図68-75)。各ボックスプロットは中央値を線で結んで作成した。

3D7、7G8株、薬剤無添加株のCRTとMDR1において、その発現量はRPKM 0~1000と様々な発現量の細胞が存在することが明らかとなった。3D7株は、3つのduplicateを含めているためバッチ差の影響が懸念されたが、単一のバッチである7G8株やその他の株からも同様の結果が表されていることからこの可能性は少ないと考えた。

ボックスプロットからも薬剤添加を受け、遺伝子発現量は平均と中央値共に変化している。しかし、発現量の平均と中央値が減少している場合であっても、その集団中にはRPKM = 100~1000といった過剰発現している細胞が存在していることが本研究で明らかとなった(図68-75)。

CRT、MDR1を観察した結果から、各細胞の各遺伝子は多様に発現して、この多様性によって生じた薬剤耐性関連遺伝子の高発現の影響によって薬剤添加後に一部の細胞が生き残っている可能性が考えられた。

またCRTやMDR1等の薬剤耐性関連遺伝子の過剰発現を示している細胞がどのようなトランスクリプトームを持つかは大変興味深い。今後は、過剰発現している細胞のキャラクタライゼーションを施すことが重要であると考えた。

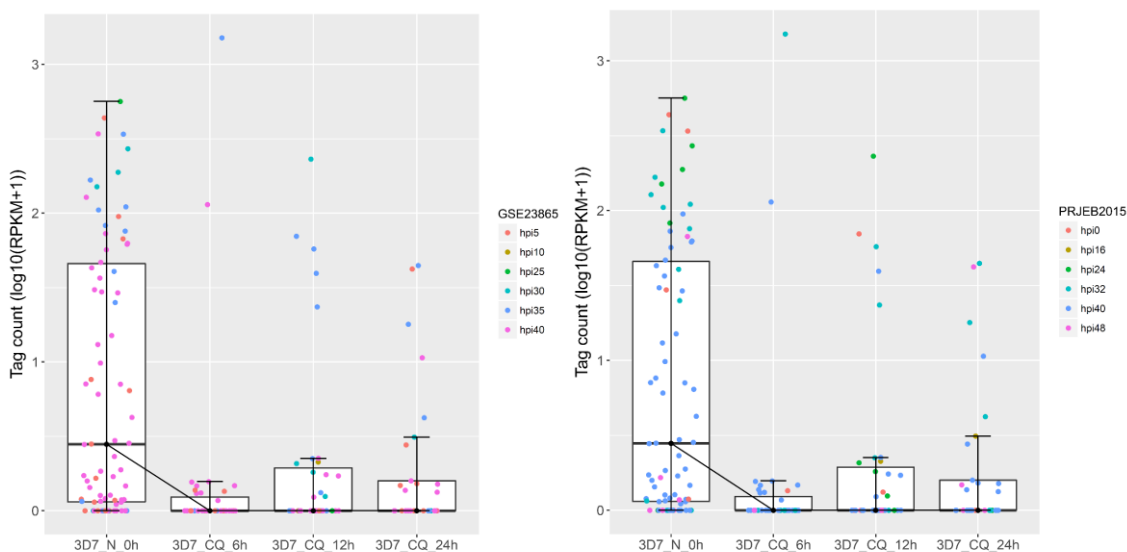


図 68. 3D7 株, 無添加, クロロキン添加株における CRT の遺伝子発現量
 (左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け,
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け.

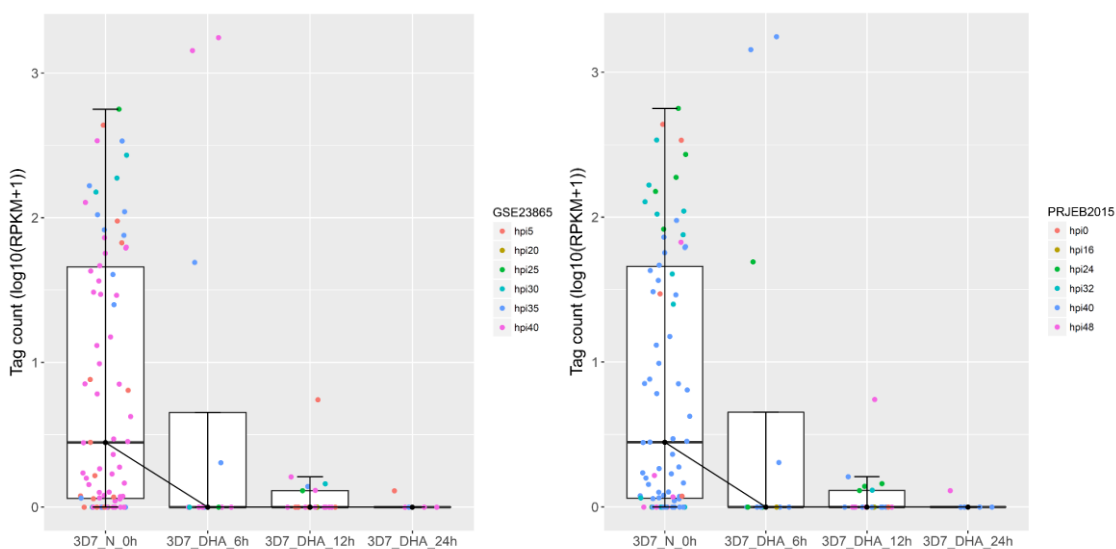


図 69. 3D7 株, 無添加, ジヒドロアルテミシニン添加株における CRT の遺伝子
 発現量

(左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け,
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け.

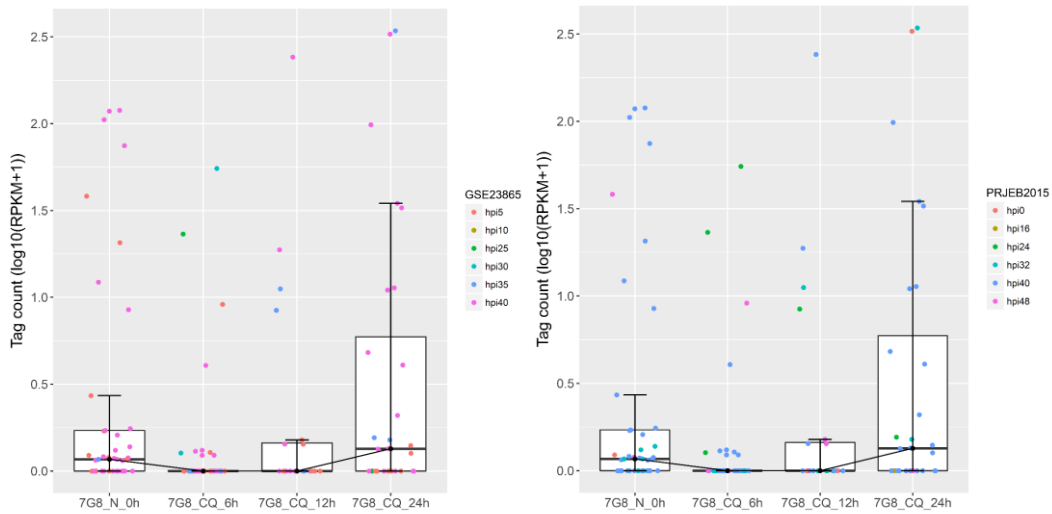


図 70. 7G8 株, 無添加, クロロキン添加株における CRT の遺伝子発現量
 (左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け,
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け.

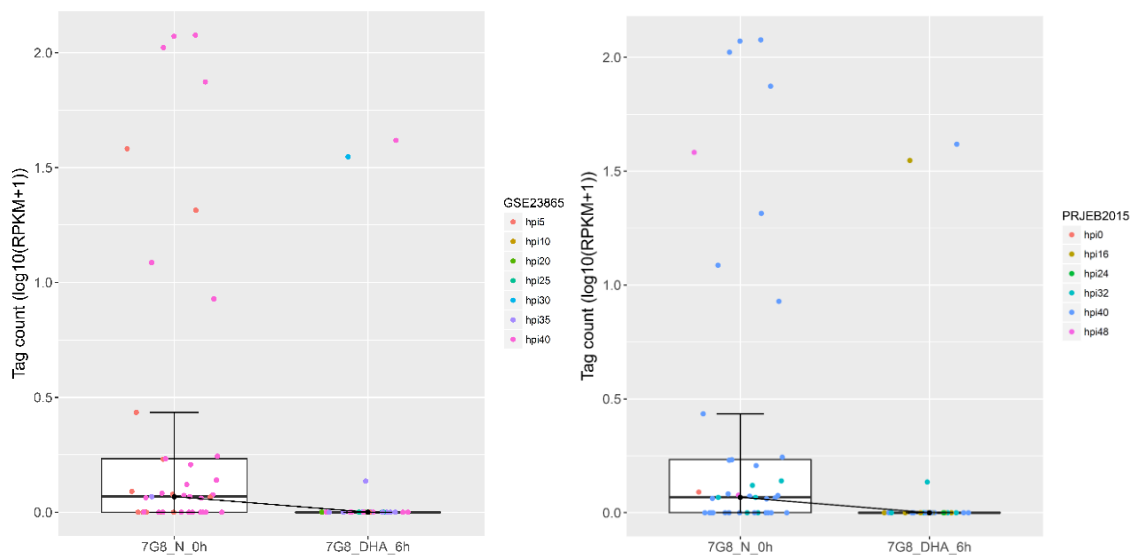


図 71. 7G8 株, 無添加, ジヒドロアルテミシニン添加株における CRT の遺伝子
 発現量

(左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け,
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け.

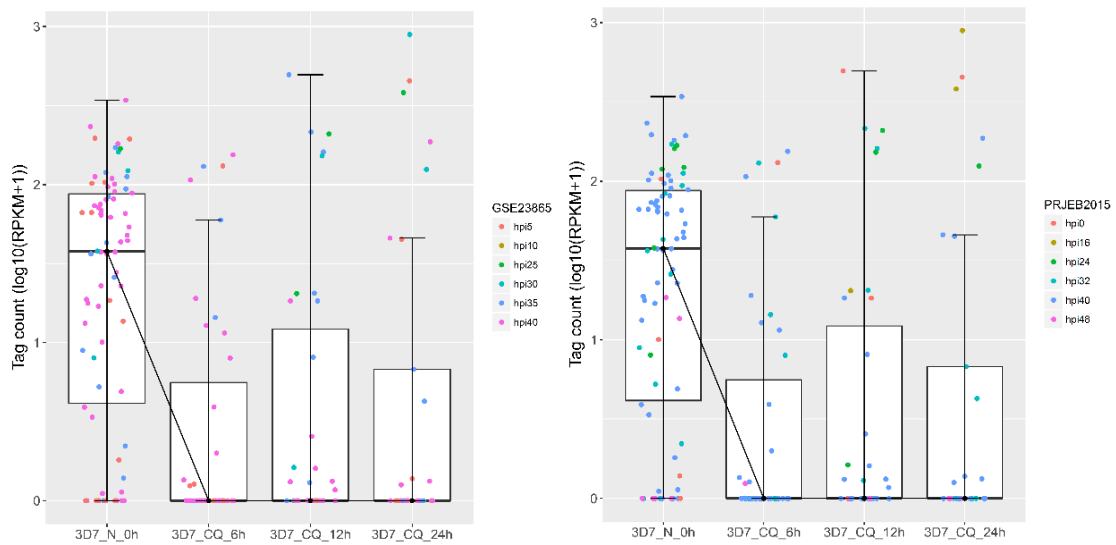


図 72. 3D7 株, 無添加, クロロキン添加株における MDR1 の遺伝子発現量
 (左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け,
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け.

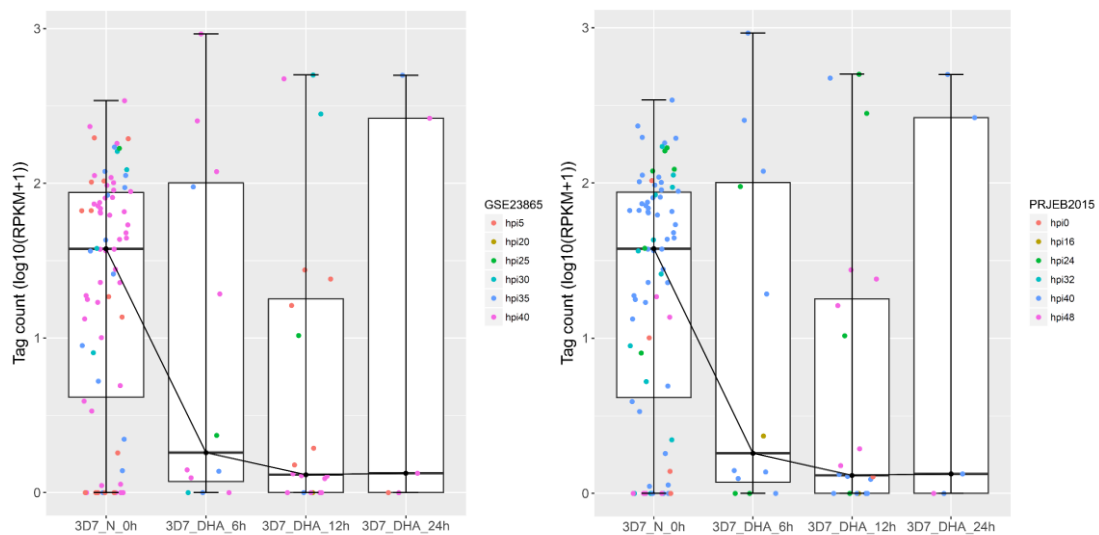


図 73. 3D7 株, 無添加, ジヒドロアルテミシニン添加株における MDR1 の遺伝子発現量

(左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け,
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け.

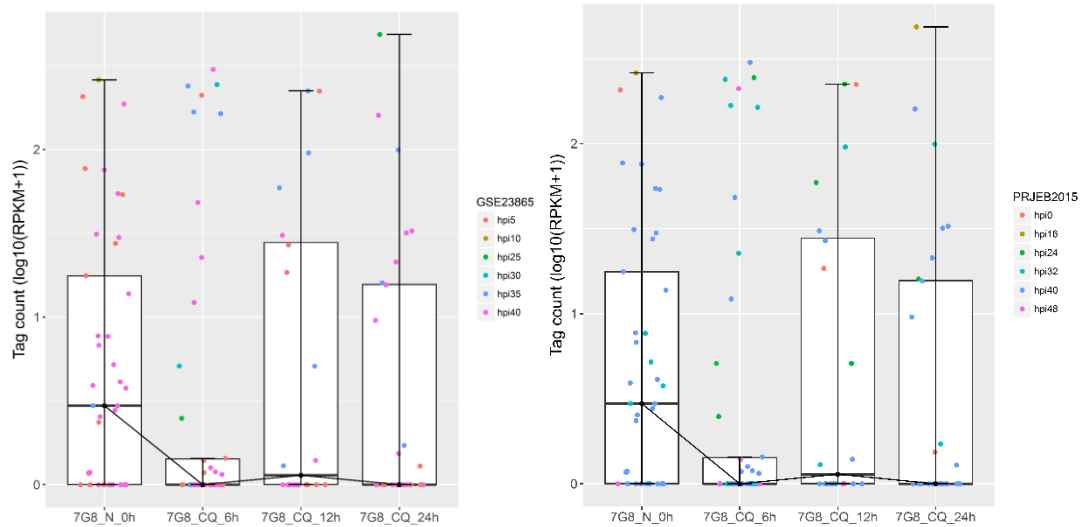


図 74. 7G8 株，無添加，クロロキン添加株における MDR1 の遺伝子発現量
 (左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け，
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け．

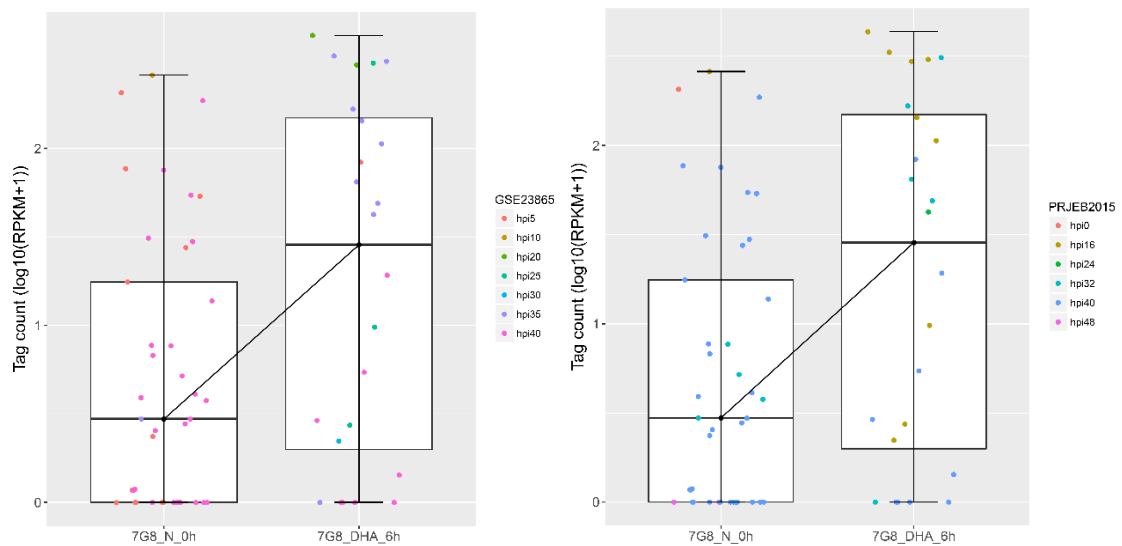


図 75. 7G8 株，無添加，ジヒドロアルテミシニン添加株における MDR1 の遺伝子発現量

(左) GSE23865(hpi5~40)と RNA-Seq データの相関の最高値で色分け，
 (右) PRJEB2015(hpi0~48)と RNA-Seq データの相関の最高値で色分け．

3.4.4. GO 解析

各 1 赤血球 RNA-Seq データの中で、後期トロホゾイド、前期シズント、後期シズントの 3 つに該当する細胞を取り上げ、ステージ毎に平均 RPKM を算出することで *in silico* でバルク(バーチャルバルク)を作成し、RPKM を算出した。生育ステージの分類は、各 1 赤血球 RNA-Seq データと PRJEB2015 のデータの相関係数を算出し、相関係数の最高値を示す hpi から判断した(後期トロホゾイドが hpi32、前期シズントが hpi40、後期シズントが hpi48 を最高値とすると定義した)。バーチャルバルク RPKM を用いて薬剤添加前と添加後(クロロキンまたはジヒドロアルテミシニン添加後 6 時間)を比較し、Up-regulate または Down-regulate された遺伝子について観測した(図 76-80)。

その結果、薬剤無添加データとクロロキン添加データを比較した場合と、薬剤無添加データとジヒドロアルテミシニン添加データを比較した場合の両方で、Up-regulate 以上に Down-regulate された遺伝子の方が多く見つかるという結果となった。特に、ジヒドロアルテミシニン添加で Down-regulate された遺伝子数は、一部のデータセットで 5777 遺伝子中、2000 遺伝子以上が該当するという結果になった(図 79,80)。そのため、今後は発現変動が発生した遺伝子の定義を調整した上で、再度解析を行うべきであると考えた。

続いて、Up-regulate または Down-regulate された遺伝子それぞれで GO 解析を行い、遺伝子発現変動がある遺伝子がどのような傾向を持っているかを検証した。各培養株(3D7、HB3、7G8、Dd2、K1)の生育ステージ毎に GO 解析を行った。

まず初めに、各 GO 解析結果の中で、p 値 0.05 以下の GO term として多くの結果に表れた GO term に注目した。

その結果、Rhoptry (GO:0020008)、microneme (GO:0020009)といったメロゾイドの破裂・感染に関連した構造物に関する遺伝子が多くの結果で Down-regulate されていることが明らかとなった(表 31-34)。Jacobs らの研究から、クロロキン添加によってマラリア原虫の形態に変化が生じることが示唆されている(Jacobs *et.al.* 1988)。そのため、Rhoptry、microneme といった構造物に関わる遺伝子発現にクロロキン添加が影響したのではないかと考えられた。

この結果から、薬剤添加後反応がトランスクリプトーム変動を起こしている可能性が示唆された。結果を踏まえて、薬剤添加前と薬剤添加後 12 時間、24 時間では、どのような結果が得られるかを検証し、更なる考察を深めていきたい。

本研究では、1 赤血球 RNA-Seq データを取得している。そのため、GO 解析を 1 赤血球 RNA-Seq データセット中の各遺伝子の変動係数を規準に、どのような遺伝子が濃縮されているかを検証することも可能である。また、バルク RPKM 対 1 赤血球 RPKM 比較して、どのように濃縮されるかを検証することで、各 1 赤血球 RNA-Seq データのキャラクタライズを行うことも可能であると考えられる。

RPKM の変動だけではなく、様々な視点でデータを比較し、解析してどのような生命現象が起きているかを探索していくことが重要であると考えた。

また、得られた GO Term に属する各遺伝子に注目して、ネットワーク解析等の特定の遺伝子に焦点を当てた解析手法を用いて各 1 赤血球 RNA-Seq データのキャラクタライズを行うことも可能である。

熱帯熱マラリア原虫集団中に存在するトランスクリプトームの中に存在する多様性のバリエーションの全体像を把握することの端緒となることが期待できると考えた。

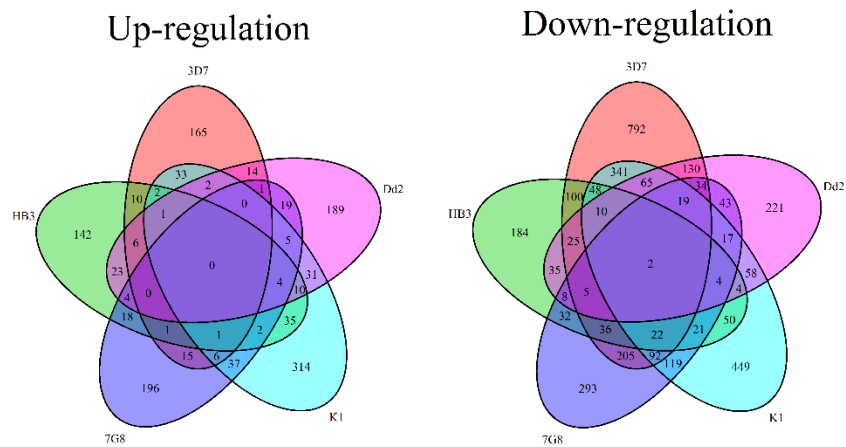


図 76. 後期トロホゾイド株にてクロロキン添加後6時間で変動した遺伝子のベン図

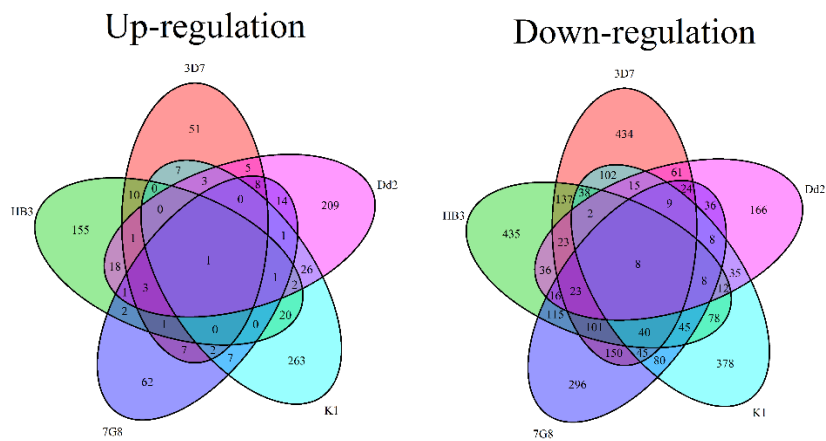


図 77. 前期シゾン株にてクロロキン添加後6時間で変動した遺伝子のベン図

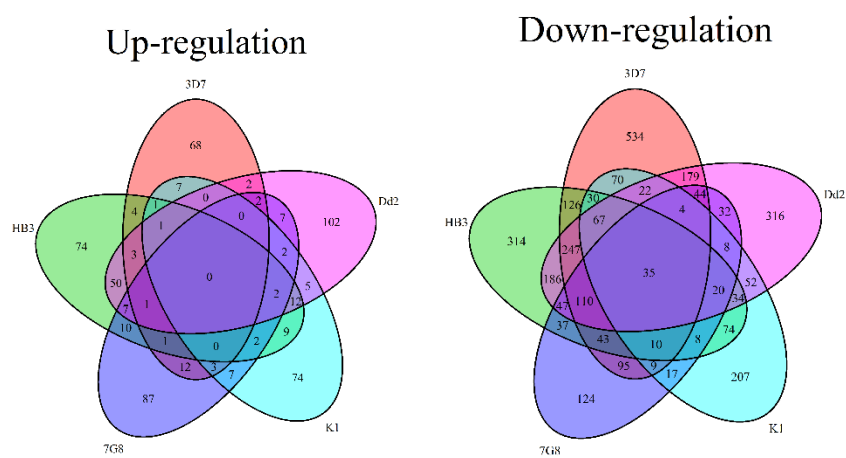


図 78. 後期シゾン株にてクロロキン添加後6時間で変動した遺伝子のベン図

Up-regulation

Down-regulation

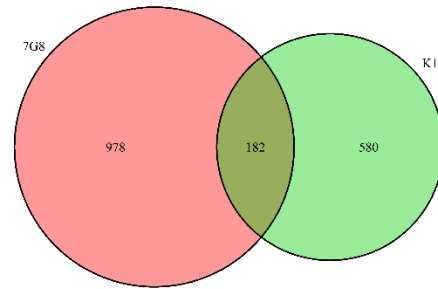
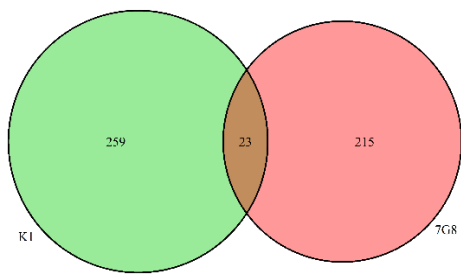


図 79. トロホゾイド株にてジヒドロアルテミシニン添加後 6 時間で変動した遺伝子のベン図

Up-regulation

Down-regulation

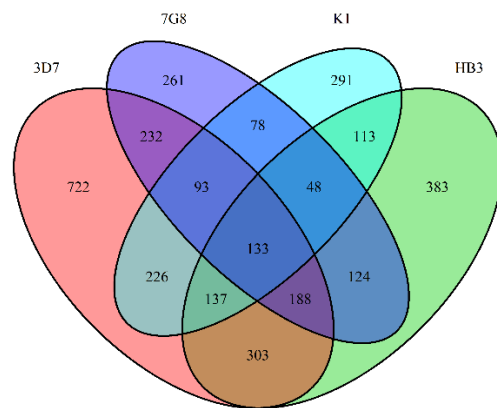
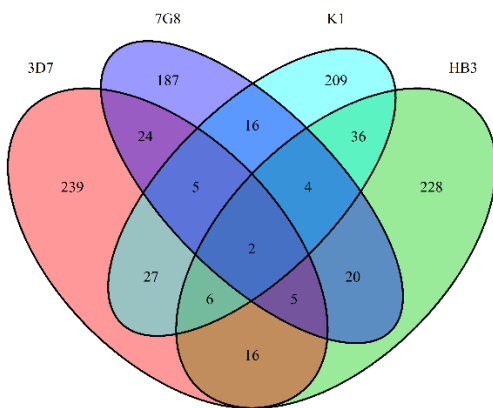


図 80. 前期シズント株にてジヒドロアルテミシニン添加後 6 時間で変動した遺伝子のベン図

表 31. GO:0020009(microneme)でクロロキン添加後 6 時間に Down-regulate
が確認された実験系

| Sample | Stage | P.value | x | m | n | k |
|--------|----------------|----------|----|----|------|------|
| 3D7 | Late Schizont | 0.012057 | 8 | 13 | 5764 | 1625 |
| HB3 | Early Schizont | 0.005706 | 7 | 13 | 5764 | 1117 |
| 7G8 | Early Schizont | 4.16E-06 | 10 | 13 | 5764 | 1004 |
| 7G8 | Late Schizont | 0.000192 | 7 | 13 | 5764 | 643 |
| K1 | Early Schizont | 0.039893 | 5 | 13 | 5764 | 903 |

表 32. GO:0020009(microneme)でジヒドロアルテミシニン添加後 6 時間に
Down-regulate が確認された実験系

| Sample | Stage | P.value | x | m | n | k |
|--------|------------------|----------|----|----|------|------|
| 7G8 | Late Trophozoite | 0.030432 | 6 | 13 | 5764 | 1160 |
| 7G8 | Early Schizont | 1.04E-06 | 11 | 13 | 5764 | 1157 |

表 33. GO:0020008(rhoptry)でクロロキン添加後 6 時間に Down-regulate が
確認された実験系

| Sample | Stage | P.value | x | m | n | k |
|--------|------------------|----------|----|----|------|------|
| 3D7 | Late Schizont | 0.000539 | 14 | 22 | 5755 | 1625 |
| HB3 | Early Schizont | 0.004693 | 10 | 22 | 5755 | 1117 |
| 7G8 | Late Trophozoite | 4.83E-05 | 12 | 22 | 5755 | 952 |

表 34. GO:0020008(rhoptry)でジヒドロアルテミシニン添加後 6 時間に
Down-regulate が確認された実験系

| Sample | Stage | P.value | x | m | n | k |
|--------|------------------|----------|----|----|------|------|
| K1 | Late Schizont | 0.048808 | 8 | 22 | 5755 | 1126 |
| HB3 | Early Schizont | 0.009055 | 11 | 22 | 5755 | 1429 |
| 7G8 | Late Trophozoite | 0.000353 | 12 | 22 | 5755 | 1160 |
| 7G8 | Early Schizont | 0.020109 | 9 | 22 | 5755 | 1157 |

第4章 総括

本研究は、熱帯熱マalaria原虫集団中に存在することが考えられるトランスクリプトーム多様性が、原虫自身の薬剤耐性獲得のリザーバーとして寄与している可能性を検証する目的で行った。DNA配列の変異による薬剤耐性獲得の前段階に、トランスクリプトーム多様性が薬剤耐性へ寄与している可能性について検証を試みた。

熱帯熱マalaria原虫感染赤血球(赤内型マalaria原虫)の1赤血球RNA-Seqを行い、そのデータ取得に初めて成功した。マalaria原虫を培養する段階からC1 Single-cell Auto prep systemを用いたシングルセル化とRNA-Seqを行うまでの様々な問題点を解消しながら、実験手法を構築した。

一連の実験を行い、27の実験系から計897細胞の1赤血球RNA-Seqデータを取得した。RNA-Seqデータは薬剤無添加サンプルと抗マalaria薬(クロロキンまたはジヒドロアルテミシニン)添加サンプルから3つの時間点(添加後6、12、24時間)で取得した。

得られた1赤血球RNA-Seqデータの信頼性と再現性を検証し、本研究の解析に用いる上では十分と考えられる結果を得ることができた。

続いて、熱帯熱マalaria原虫の生育ステージを推定するための手法を考案した。赤内型マalaria原虫は、その生活環の中でリング、トロホゾイド、シズントの大きく分けて3つの生育ステージに分けられる。マalaria原虫は生育ステージによってトランスクリプトームが大きく異なるため、この生育ステージを取得した1赤血球RNA-Seqデータ毎に分類する必要がある。

上記で記した27の実験系から取得したデータの解析を行う上で必要となるパイプライン構築を行う目的から、生育ステージを実験的に一定の精度で均一化

させた同調培養株で 2 つの実験系(トロホゾイド同調培養株、シズント同調培養株)から計 66 細胞の 1 赤血球 RNA-Seq データを取得した。

生育ステージが明らかである同調培養株から取得した 1 赤血球 RNA-Seq データを用いて、各取得データの生育ステージを推定する手法を考案し、実際に利用できるかを検証した。

主成分分析、マーカー遺伝子の発現から生育ステージを同定する手法を考案し、実際に解析に用いた結果、一定の成果を上げることができた。しかし、これらの手法には解析を行う上での様々な欠点が明らかになった。

ここで明らかとなった欠点を補うために、公共データセットを援用する手法を考案し、一定の精度で利用できることを確認した。また各種解析を行う場面で、柔軟に利用できるといった利点も持っていることから、こちらの手法を主として利用することとした。

生育ステージを分類した上で、次に各データセット中のトランスクリプトーム解析に取り組んだ。発現変動の指標の一つになると考え、1 赤血球単位における遺伝子発現数及び 1 赤血球データセットにおける各遺伝子の分散(及び変動係数)を算出した。薬剤添加時のトランスクリプトーム変動を観測するために、考案した解析方法であったが、トロホゾイドから後期シズントにかけて遺伝子数が減少する可能性を示唆するといった結果を得ることができた。

1 赤血球 RNA-Seq により、従来のバルク RNA-Seq 以上の解像度でデータを取得できたことで、このような可能性を検討することができた。この可能性は、研究の本筋からそれる内容であるが、更なる解析を進めて、熱帯熱マラリア原虫という生物へのより深い理解へ繋がる可能性があり興味深い内容であった。

続いて、トランスクリプトーム多様性が薬剤耐性へ寄与している可能性について検証をするために、薬剤無添加サンプルと抗マラリア薬(クロロキンまたは

ジヒドロアルテミシニン)添加サンプルの1赤血球単位のトランスクリプトーム変動の観測を試みた。主成分分析や RPKM から、薬剤や抗マラリア薬の有無の影響を検証した。その結果、一定の遺伝子発現変動に加えて、細胞間に存在するトランスクリプトーム多様性を示唆する結果を得ることができた。

薬剤耐性遺伝子に注目して発現多様性を観測した結果、一定の発現のバラつきを観測することに成功した。特に、薬剤耐性遺伝子の過剰発現を起こしている株の存在を確認した。このような特徴を持った株の詳細な解析が本研究の目的である、薬剤耐性とトランスクリプトーム多様性の関連性についてアプローチできることを期待している。また、トランスクリプトーム多様性が一体何に起因するかは検証していきたい。

トランスクリプトーム多様性が生じる原因の一つに、転写因子の関連性が考えられる。マラリア原虫は転写因子について明らかになっている部分が少なく、転写因子の数が限られていることが予測されており、通常の転写制御とは異なるメカニズムを持つことが予想されている(Ay *et.al.* 2014)。転写制御メカニズムの違いの影響により、転写制御の安定性が小さいと仮定すると、それに起因してトランスクリプトーム多様性が生まれると考えられる。もし、そのような現象が生じているとしたら大変興味深い。

本研究で得られた1赤血球 RNA-Seq データ並びに、解析パイプラインと解析結果から新たな知見が生まれることを期待して、更なる解析を行っていきたい。

引用文献

Andrew F.G. Slater. (2002) Chloroquine: Mechanism of drug action and resistance in *plasmodium falciparum*. *Pharmacology & Therapeutics* **57**, 203-235.

Anne-Catrin Uhlemann, Angus Cameron, Ursula Eckstein-Ludwig, Jorge Fischbarg, Pavel Iserovich, Felipe A Zuniga, Malcolm East, Anthony Lee, Leo Brady, Richard K Haynes & Sanjeev Krishna. (2012) A single amino acid residue can determine the sensitivity of SERCAs to artemisinin. *Nature Structural & Molecular Biology* **19**, 264.

Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ Jr, Treatman C, Wang H. (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research* **37**,D539-543.

Alassane Mbengue, Souvik Bhattacharjee, Trupti Pandharkar, Haining Liu, Guillermina Estiu, Robert V. Stahelin, Shahir S. Rizk, Dieudonne L. Njimoh, Yana Ryan, Kesinee Chotivanich, Chea Nguon, Mehdi Ghorbal, Jose-Juan Lopez-Rubio, Michael Pfrender, Scott Emrich, Narla Mohandas, Arjen M. Dondorp, Olaf Wiest & Kasturi Haldar. (2015) A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature* **520**, 683-687.

Alex K. Shalek, Rahul Satija, Joe Shuga, John J. Trombetta, Dave Gennert, Diana Lu, Peilin Chen, Rona S. Gertner, Jellert T. Gaublomme, Nir Yosef, Schraga Schwartz, Brian Fowler, Suzanne Weaver, Jing Wang, Xiaohui Wang, Ruihua Ding, Raktima Raychowdhury, Nir Friedman, Nir Hacohen, Hongkun Park, Andrew P. May and Aviv Regev. (2014) Single cell RNA Seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363-369.

Ben Langmead & Steven L Salzberg. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359.

Benoit Witkowski, Chanaki Amaratunga, Nimol Khim, Sokunthea Sreng, Pheaktra Chim, Saorin Kim, Pharath Lim, Sivanna Mao, Chantha Sopha, Bamey Sam, Jennifer M Anderson, Socheat Duong, Char Meng Chuor, Walter R J Taylor, Seila Suon, Odile Mercereau-Puijalon, Rick M Fairhurst, Didier Menard. (2013) Novel phenotypic assays for the detection of artemisinin-resistant *Plasmodium falciparum* malaria in Cambodia in-vitro and ex-vivo drug-response studies. *The Lancet Infectious Diseases* **57**, 914-923.

Chris Lambros and Jerome P.vanderberg. (1979) Synchronization of *Plasmodium falciparum* Erythrocytic Stages in Culture. *The Journal of Parasitology* **65**, 418-420.

Christopher V. Plowe, Joseph F. Cortese, Abdoulaye Djimde, Okey C. Nwanyanwu, William M. Watkins, Peter A. Winstanley, Jose G. Estrada Franco, Rene E. Mollinedo, Juan Carlos Avila, Jose Luis Cespedes, Darrick Carter and Ogobara K. Doumbo. (1997) Mutations in *Plasmodium falciparum* Dihydrofolate Reductase and Dihydropteroate Synthase and Epidemiologic Patterns of Pyrimethamine-Sulfadoxine Use and Resistance. *The Journal of Infectious Diseases* **176**, 1590-1596.

Daniel Ramsköld, Shujun Luo, Yu-Chieh Wang, Robin Li, Qiaolin Deng, Omid R Faridani, Gregory A Daniels, Irina Khrebtukova, Jeanne F Loring, Louise C Laurent, Gary P Schroth & Rickard Sandberg. (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology* **30**,777-782.

Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley and Steven L Salzberg. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36.

Elya Dekel, Anna Rivkin, Meta Heidenreich, Yotam Nadava, Yifat Ofir-Birin, Ziv Porat, Neta Regev-Rudzki. (2017) Identification and classification of the malaria parasite blood developmental stages, using imaging flow cytometry. *Methods* **112**, 157-166.

Ferhat Ay, Evelien M. Bunnik, Nelle Varoquaux, Sebastiaan M. Bol, Jacques Prudhomme, Jean-Philippe Vert, William Stafford Noble and Karine G. Le Roch. (2014) Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Research* **24**, 974–988.

Frédéric Ariey, Benoit Witkowski, Chanaki Amaratunga, Johann Beghain, Anne-Claire Langlois, Nimol Khim, Saorin Kim, Valentine Duru, Christiane Bouchier, Laurence Ma, Pharath Lim, Rithea Leang, Socheat Duong, Sokunthea Sreng, Seila Suon, Char Meng Chuor, Denis Mey Bout, Sandie Ménard, William O. Rogers, Blaise Genton, Thierry Fandeur, Olivo Miotto, Pascal Ringwald, Jacques Le Bras, Antoine Berry, Jean-Christophe Barale, Rick M. Fairhurst, Françoise Benoit-Vical, Odile Mercereau-Puijalon & Didier Ménard. (2014) A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* **505**, 50-55.

Gretta H. Jacobs, Ayo M. J. Oduola, Dennis E. Kyle, Wilbur K. Milhous, Samuel K. Martin And Masamichi Aikawa. (1988) Ultrastructural study of the effects of chloroquine and verapamil on *Plasmodium falciparum*. *The American Journal of Tropical Medicine and Hygiene* **39**, 15-20.

Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **15**, 2078-9.

Hiasindh Ashmi Antony, Subhash Chandra Parija. (2016) Antimalarial drug resistance: An overview. *Tropical Parasitology* **6**, 30-41.

Hiroko Asahi, Tamotsu Kanazawa, Nakami Hirayama, Yousei Kajihara.

(2005) Investigating serum factors promoting erythrocytic growth of *Plasmodium falciparum*. *Experimental Parasitology* **109**, 7-15.

Hiroko Asahi, Shinji Izumiyama, Mohammed Tolba, Bethel kwanza Bentum.

(2011) *Plasmodium falciparum*: Differing effects of non-esterified fatty acids and phospholipids on intraerythrocytic growth in serum-free medium. *Experimental Parasitology* **127**, 708-13.

James T Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz & Jill P Mesirov. (2011) Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26.

Jennifer M Reynolds, Kamal El Bissati, Jens Brandenburg, Arthur Günzl and Choukri Ben Mamoun. (2007) Antimalarial activity of the anticancer and proteasome inhibitor bortezomib and its analog ZL3B. *BMC Clinical Pharmacology* **7** :13.

Karl Pearson. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* **2**, 559-572.

Laurens van der Maaten and Geoffrey Hinton. (2008) Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579-2605.

Lydia Mata-Cantero, Maria J Lafuente, Laura Sanz and Manuel S Rodriguez.

(2014) Magnetic isolation of *Plasmodium falciparum* schizonts iRBCs to generate a high parasitaemia and synchronized in vitro culture. *Malaria Journal* **13**, 112.

Malcolm J. Gardner, Neil Hall, Eula Fung, Owen White, Matthew Berriman, Richard W. Hyman, Jane M. Carlton, Arnab Pain, Karen E. Nelson, Sharen Bowman, Ian T. Paulsen, Keith James, Jonathan A. Eisen, Kim Rutherford, Steven L. Salzberg, Alister Craig, Sue Kyes, Man-Suen Chan, Vishvanath Nene, Shamira J. Shallom, Bernard Suh, Jeremy Peterson, Sam Angiuoli, Mihaela Pertea, Jonathan Allen, Jeremy Selengut, Daniel Haft, Michael W. Mather, Akhil B. Vaidya, David M. A. Martin, Alan H. Fairlamb, Martin J. Fraunholz, David S. Roos, Stuart A. Ralph, Geoffrey I. McFadden, Leda M. Cummings, G. Mani Subramanian, Chris Mungall, J. Craig Venter, Daniel J. Carucci, Stephen L. Hoffman, Chris Newbold, Ronald W. Davis, Claire M. Fraser and Bart Barrell. (2002) Genome sequence of the human malaria parasite. *Plasmodium falciparum*. *Nature* **419**, 498-511.

Michael B. Reed, Kevin J. Saliba, Sonia R. Caruana, Kiaran Kirk & Alan F. Cowman. (2000) Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature* **403**, 906-909.

Nicholas Fisher, Roslaini Abd Majid, Thomas Antoine, Mohammed Al-Helal, Ashley J. Warman, David J. Johnson, Alexandre S. Lawrenson, Hilary Ranson, Paul M. O'Neill, Stephen A. Ward and Giancarlo A. Biagini. (2012) Cytochrome b Mutation Y268S Conferring Atovaquone Resistance Phenotype in Malaria Parasite Results in Reduced Parasite bc₁ Catalytic Turnover and Protein Expression. *The Journal of Biological Chemistry* **287**, 9731-9741.

Maria Cartolano, Bruno Huettel, Benjamin Hartwig, Richard Reinhardt, Korbinian Schneeberger. (2016) cDNA Library Enrichment of Full Length Transcripts for SMRT Long Read Sequencing. *PLoS ONE* **11**, e0157779.

Nuno Vale, Luísa Aguiar & Paula Gomes. (2014) Antimicrobial peptides: a new class of antimalarial drugs? *frontiers in Pharmacology* **5**, 275.

Philip J. Shaw, Sastra Chaotheing, Pavita Kaewprommal, Jittima Piriyaongsa, Chayaphat Wongsombat, Nattida Suwannakitti, Pongpisid Koonyosying, Chairat Uthaipibull, Yongyuth Yuthavong and Sumalee Kamchonwongpaisan. (2015) Plasmodium parasites mount an arrest response to dihydroartemisinin, as revealed by whole transcriptome shotgun sequencing (RNA-seq) and microarray study. *BMC Genomics* **16**:830.

Ping Wang, Darren R. Brooks, Paul F.G. Sims, John E. Hyde. (1995) A mutation-specific PCR system to detect sequence variation in the dihydropteroate synthetase gene of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **71**, 115-125.

Richard Bartfai, Wieteke A. M. Hoeijmakers, Adriana M. Salcedo-Amaya, Arne H. Smits, Eva Janssen-Megens, Anita Kaan, Moritz Treeck, Tim-Wolf Gilberger, Kees-Jan Francoijs, Hendrik G. Stunnenberg. (2010) H2A.Z Demarcates Intergenic Regions of the Plasmodium falciparum Epigenome That Are Dynamically Marked by H3K9ac and H3K4me3. *PLOS Pathogens* **6**, e1001223.

RunYe, Dongwei Hu, Yilong Zhang, Yufu Huang, Xiaodong Sun, JianWang, XuediChen, HongningZhou, DongmeiZhang, Mathirut Mungthin & Weiqing Pan. (2016) Distinctive origin of artemisininresistant *Plasmodium falciparum* on the China-Myanmar border. *Scientific Reports* **6**, 20100.

Sabina Dahlström, M. Isabel Veiga, Andreas Mårtensson, Anders Björkman and J. Pedro Gil. (2009) Polymorphism in PfMRP1 (Plasmodium falciparum Multidrug Resistance Protein 1) Amino Acid 1466 Associated with Resistance to Sulfadoxine-Pyrimethamine Treatment. *Antimicrobial Agents and Chemotherapy* **53**, 2553-2556.

Stephanie G. Valderramos, David A. Fidocke. (2006) Transporters involved in resistance to antimalarial drugs. *Trends in Pharmacological Sciences* **27**, 594-601.

The World Health Organization. World Malaria Report 2015.

<http://www.who.int/malaria/publications/world-malaria-report-2015/report/en/>.

Thomas D Otto, Daniel Wilinski, Sammy Assefa, Thomas M Keane, Louis R Sarry, Ulrike Böhme, Jacob Lemieux, Bart Barrell, Arnab Pain, Matthew Berriman, Chris Newbold and Manuel Llinás. (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Molecular Microbiology* **76**, 12-24.

Tolga Eichhorn, Dominic Winter, Berthold Büchele, Natalie Dirdjaja, Martin Frank, Wolf-Dieter Lehmann, Rolf Mertens, R. Luise Krauth-Siegel, Thomas Simmet, Joachim Granzin, Thomas Efferth. (2013) Molecular interaction of artemisinin with translationally controlled tumor protein (TCTP) of *Plasmodium falciparum*. *Biochemical Pharmacology* **85**, 38-45.

Zbynek Bozdech, Jingchun Zhu, Marcin P Joachimiak, Fred E Cohen, Brian Pulliam and Joseph L DeRisi. (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biology* **4**, R9.

Zbynek Bozdech, Manuel Llina, Brian Lee Pulliam, Edith D. Wong, Jingchun Zhu, Joseph L. DeRisi. (2003) The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*. *PLoS BIOLOGY* 1, E5.

謝辞

本研究は、2015年4月から2017年3月にかけて東京大学大学院新領域創成科学研究科メディカル情報生命専攻 生命システム観測分野（鈴木穰研究室）において行ったものである。

本研究の遂行に当たり、東京大学大学院メディカル情報生命専攻 鈴木穰教授には、懇切丁寧なご指導ご鞭撻を賜り心から深く感謝申し上げます。

お忙しい中、本論文審査の副査を引き受けてくださった、東京大学大学院メディカル情報生命専攻 津田宏治教授、程久美子准教授に御礼申し上げます。

また、研究を進める上で有益なご助言をいただいたゲノムシステム医療科学分野 菅野純夫教授、工学院大学先進工学部生命化学科 水島一菅野純子教授に衷心より感謝の意を表します。

実験サンプルの提供と助言をしていただきました、杏林大学医学部感染症学講座 朝日博子博士に深く感謝いたします。

的確な助言と激励をくださったダイナコム株式会社 若栗浩幸博士、関真秀特任助教、Lucky Ronald Runtuwene 特任研究員に深く感謝いたします。

今村聖実様、阿部佳澄様、昆布恵美様、石川由似様、島津幸恵様、Jakub Makalowski 様、小林悦子様には、シークエンシング及び数々の実験を行っていただきました。鳥谷恵子様、ダイナコム株式会社 堀内映実様、久世裕太様には、シークエンスデータの解析をサポートしていただきました。荒内貴子様、清水佳津子様には研究活動に関わる様々な手続きを行っていただきました。

本研究を実施するにあたり、実験や解析など様々な面で多くの方々に協力していただいたため、成果を上げることが初めて出来ました。生命システム観測分野の技術支援員、出向研究員、秘書の皆さまに深く感謝いたします。

研究活動や研究生活の中で多くの時間を共に過ごし、大変お世話になりました先輩と後輩、OB・OGに深く感謝いたします。

本研究の実験で用いた赤血球を提供していただきました、日本赤十字社に感謝いたします。

最後に、研究活動を支えていただいた両親と兄に心より謝意を表します。