

We Find Health in Your Diversity.

iRweb:データ解析ガイド

免疫系レパートリー増幅から次世代シーケンシング

この製品は研究目的のみに使用できます。臨床診断目的には使用できません。



We Find Health in Your Diversity.

iRepertoire® is a registered trademark of iRepertoire, Inc. The iR logo is a trademark of iRepertoire, Inc. Illumina®, HiSeq®,and MiSeq®, are registered trademarks of Illumina, Inc. HiSeq2000™ and GAllx™ are trademarks of Illumina, Inc. 454®, 454 Sequencing®, GS FLX Titanium®, and GS Junior® are registered trademarks of Roche Diagnostics GmbH. Ion Torrent® is a registered trademark of Life Technologies Corporation, Inc.

iRepertoire, Inc. does not assume any liability, whether direct or indirect, arising out of the application or use of any products, component parts, or software described herein or from any information contained in this guide. Furthermore, sale of iRepertoire, Inc. products does not constitute a license to any patent, trademark, copyright, or common-law rights of iRepertoire or the similar rights of others. iRepertoire, Inc. reserves the right to make any changes to any processes, products, or parts thereof, described herein without notice. While every effort has been made to make this manual as complete and accurate as possible as of the publication date, iRepertoire assumes no responsibility that the goods described herein will be fit for any particular purpose for which you may be buying these goods.

目 次

1945天		'
解析:	2D Map表示·····	2
解析:	3D Map表示·····	5
解析:	CDR3リストとCDR3代数・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	7
解析:	D50·····	9
解析:	ツリー・マップ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	10
分布解析・・・・・・・・・・・・11		
	V-使用例······	11
	V-トリミング分布例・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	12
	CDR3長分布例······	13
	N-付加分布例······	14
4 4		4-



次世代シーケンサーにより、各ライブラリーから膨大な量の詳細なTCR及びBCRのシーケンス情報が得られますが、そこから意味のある情報を抽出する必要があります。そのデータ解析を容易にするために、弊社は自動のソフトウェア・パイプラインを構築しました。このパイプラインには、ライブラリー作成の際の増幅やシーケンシング工程で発生するエラーを取り除くために、厳密なフィルターを組み込んでいます。

データがフィルター処理された後、数種類の解析が実行されます。

推奨するブラウザ

最良の結果を表示させるために、Mozilla FirefoxかGoogle Chromeのウエッブ・ブラウザを用いて下さい。

ログ・イン

https://irweb.irepertoire.com/nir/のサイトから、取得したアカウントを用いてログインして下さい。もし、まずソフトウェアの内容を確認してみたい場合は、以下のデモ用アカウントを用いてログインすることが出来ます。

Username : demo Password : 12345

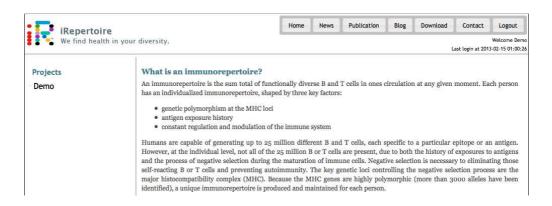


図1: デモ用アカウントでログインした後の最初の表示

データにアクセスするために、左側のパネルから、"Demo "または固有の"サンプル名"を選択します。デモ用アカウントでは、Demo 1, Demo 2, Demo 3,の3種のパネルがあります。この3種の内の1つをクリックします。仮に、Demo 1を選択した場合、IGH (Immunoglobulin Heavy Chain), IGK (Immunoglobulin Kappa chain), IGL (Immunoglobulin Lambda Chain), TRA (TCR-alpha), TRB (TCR-beta), TRD (TCR-delta), 及び TRG-(TCR gamma)の概要にアクセス出来ます。もし、IGHを選択した場合、新たなページが表示され、左側のパネル内の"Show 2D Map,"、"Show 3D Map," "List CDR3,"、"CDR3 algebra,"、"Compute D50,"、"Tree Map."の項目を利用できます。また、"V usage,"、"J usage,"、"V trimming,"、"J trimming,"、"CDR3 length,"、"N-addition."を含む数々の分布解析が可能です。これらのオプションの標準化分布も確認できます。また、これらの解析項目では、個々のデータ・セットの統計概要が確認できます。

解析: 2D Map表示

解析: 2D Map表示

2D Map表示: Heat Map

図2は、大腸がん患者と健常者のT helper集団からの、2次元ヒート・マップの例です。生殖系列V-遺伝子アレルの相対頻度(IMGTデータベースでのアライメント)は、生殖系列J-遺伝子に対してプロットされます。従って、マップの色によって、どのV-J組み合わせが高頻度か低頻度かを即座に判断できます。マップは対話式です。一度、特定のボックスをクリックすると、図3の様に、特定のV-J組み合わせを含んでいるライブラリーの代表する配列が表示されます。個々のV-J組み合わせのライブラリーには全配列が含まれるため、この最終リストには多くの配列が表示されることになります。このリストでは、多量の詳細なシーケンス情報が得られます。その情報としては、翻訳されたプロテイン配列、得られたDNA配列、IMGTデータベースとのアライメント、生殖系列アリル配列との相違、CDR1、CDR2、CDR3の同定の情報が含まれます。更に、CDR3の配列がFASTAのようなヘッダーに表示されます。

CDR1からCDR3の同定は、用いたシーケンス手法に依存します。現在は、Illumina HiSeq 100と150-ペアエンド・リード(PER)、Illumina MiSeq (100,150及び250-PER)そしてRoche 454の3種のシーケンス手法になります。Illumina HiSeqとMiSeq (100-PERと150-PER)では CDR3周辺の約150ベースペアの配列データが得られます。MiSeq 250-PERとRoche 454ではCDR3周辺の約450ベースの配列データが得られ、このことは、高頻度変異パターンをもつ CDR1からCDR3の十分な情報が得られるために、BCRのシーケンスにより適しています。図3の実例に示すように、MiSeq 250-PER によるCDR1、CDR2、CDR3に関する情報が表示されます。全てのシーケンス装置でユニークなCDR3sの同定が可能です。

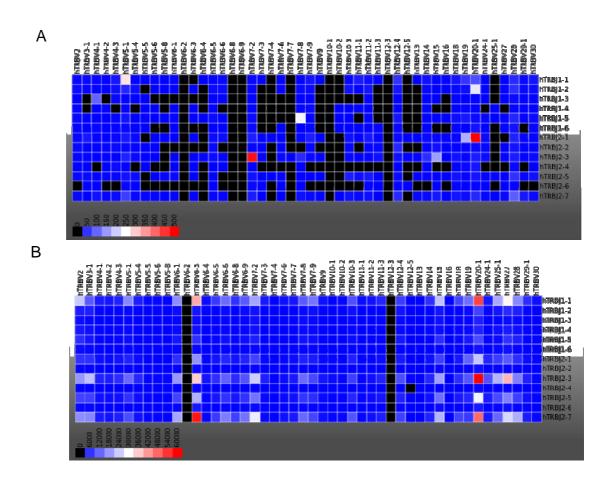


図 2: 大腸がん患者(A)と健常者(B)のT-helper集団のヒート・マップ

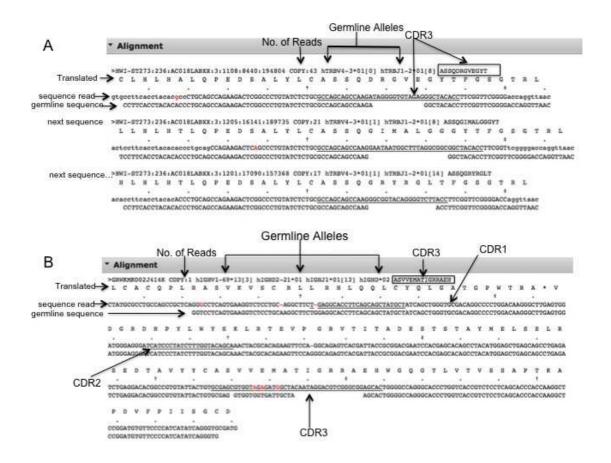


図 3: ヒート・マップ内の、ある四角ボックスをクリックした時の部分的なアライメントの表示。 CDR3周辺の約150 ベース・ペアをカバーしているIllumina HiSeqでのTCRシーケンスでの結果(A)。 CDR1、CDR2、CDR3及びC-領域の初めをカバーするIllumina MiSeq (250-PER)またはRoche 454でのBCRシーケンスの結果(B)。赤でハイライトされたヌクレオチドが生殖系列アリルと異なります。更に、CDR1 - 3と関連している核酸配列には下線が引かれています。10 ヌクレオチド毎に "・"がそのヌクレオチドの上に印され、50ヌクレオチド毎に"†"のマークが印され、100ヌクレオチド毎にそのヌクレオチドの上に"‡"が印されます。

解析: 3D Map表示

ヒート・マップの様なプロットされた情報に加えて、3次元プロットでのV-J頻度の表示が可能です。構成はヒート・マップと似ていますが、頻度は、図4の様にZ軸を基準に、選択した配列に対して赤のカウントと共に棒グラフとしてプロットされます。Z軸の基準値を超えたリード数を持つV-J組み合わせは、その棒グラフ上に赤でリード数が表示されます。3Dマップ、あるいは逆に、ヒート・マップに戻ってJアリルを持つ一つの特定のVアリルのみを確認するためには、特定のV-アリル・カラム又はJ-アリル・カラムを選択します。非常に似ている3Dマップが、図5に示されているように、J-アリルに対する選択されたV-アリルの頻度を示すマップが得られます。

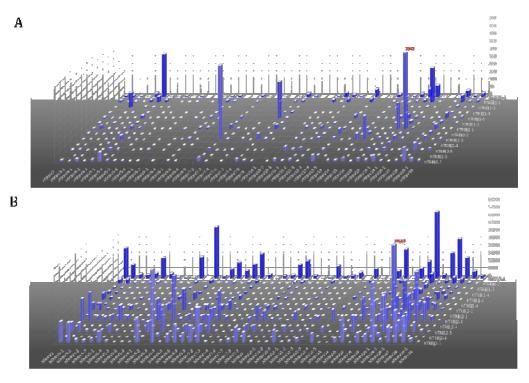


図 4: 大腸がん患者(A)と健常者(B)のT-helper群の3次元マップ

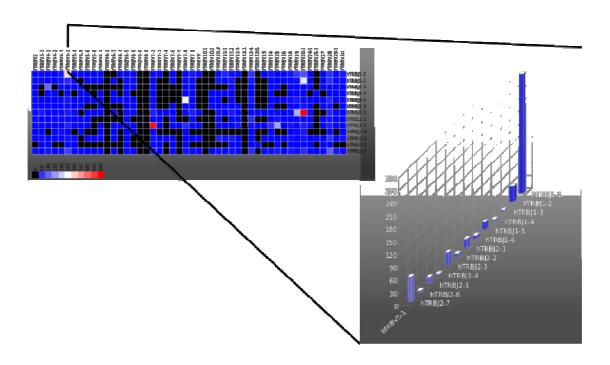


図 5: ある特定のV-アリルがヒート・マップ上で選択されると、J-アリルに対するV-アリルのみの3次元マップが表示されます。

解析: CDR3とCDR3代数

解析: CDR3リストとCDR3代数

CDR3 リスト

CDR3領域は、TCR又はBCRのその領域と高い関連性を示す抗原特異性に多くの研究者が特に 興味を持っています。そこで、選択したライブラリーに対して、図6の様に分類できるリストとして CDR3のリストが確認できます。ある特定のCDR3が選択されると、詳細な配列リストが表示され、ヒート・マップでのリストと似ており、特定のCDR3を含む代表的な配列が示されます。

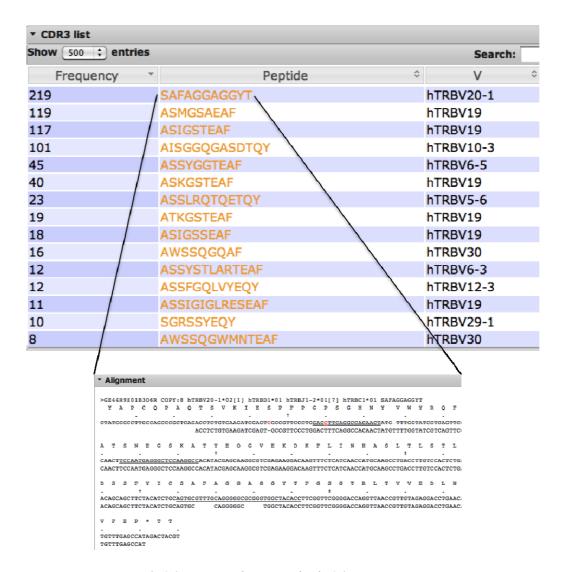


図 6: CDR3配列の分類されたリスト。ある配列が選択されると、配列アライメント・リスト(図 2のアライメントに類似)が、その特定のCDR3の配列のみが表示されます。

CDR3 代数

このソフトウェアの、非常に便利な特徴の一つがCDR3代数です。この機能により、共有のCDR3sを同定するために、ある一つのデータセットと他のデータセットのCDR3配列を比較することが出来ます。CDR3代数を選択すると、図7の様に、選択ボックスが現れます。右側にスクロールする必要がある場合の為に、選択ボックスは一覧になっています。左側のカラム内のボックスから、データセットをクリックして選択すると、現在のデータセットと比較されます。データはCDR3の頻度でフィルターされ、あらかじめセットされた頻度での共有CDR3配列のみが表示されます。図7にその出力例を示しています。共有のCDR3配列を含む、ダウンロード可能なCSVファイルも作成されます。更に、右側の表からデータセットを選択することにより、そのデータセットからCDR3を削除することも出来ます。例えば、患者では共有されているが、健常者では見つからないCDR3のリストが必要な場合に有用です。

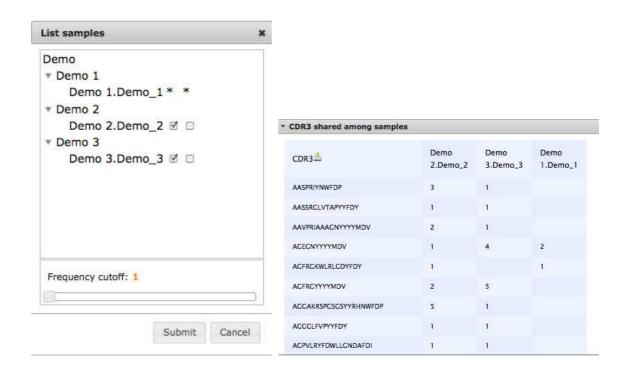


図 7: CD3代数選択ボックスと出力。左側のパネルは、比較するデータセットの選択ボックスを示しており、右側のパネルは3種のデータセットの共有CDR3を比較した時の出力例です。

解析: D50

各ライブラリーの相対的多様性を表示し比較するために、弊社は独自の解析法、D50、を開発しました。D50は、1つの数値の設定でライブラリーの多様性を決定し、T細胞又はB細胞の多様性の程度を定量的に測定します。D50は、サンプル中でのカウントされた全CDR3に対して、累計50%を占めている主要でユニークなT細胞又はB細胞クローンの割合になります。より多くの多様性を示すライブラリーでは、その精密な値は50になるでしょう。低い多様性を示す値は、減少した多様性に関連しています。D50値に加えて、図8に示す様な、計算値のグラフ表示が得られます。

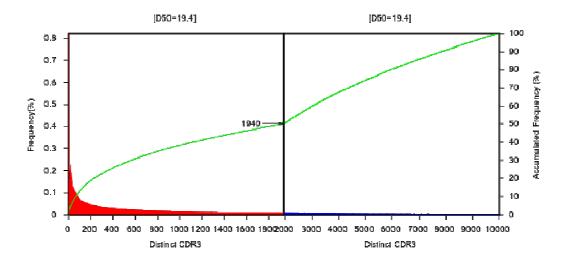


図 8:D50のグラフ表示例

解析: ツリー・マップ

階層マップは多様性を示す、別の表記法です。階層マップ中の、各円形長方形は、それぞれのユニークなV-J-CDR3を表しており、図9の例の様に、各大きさは相対的頻度を示しています。プロットの全領域は、V-使用量に従って、サブ領域に分割され、更に、J-使用量やCDR3の頻度によって分割されます。形状の不均一は、内在的な免疫系レパートリーの固有のバイアスを反映しています。弊社は、これらのマップは判断が非常に難しいために、通常は科学的資料としては用いませんが、科学的レパートリーの芸術的な表現として用いられます。

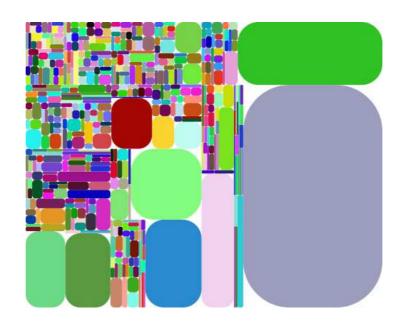


図 9: 大腸がん患者からのT-ヘルパー集団の階層マップ例

分布解析

ソフトウエアは、更にV-使用率(図10)、J-使用率、V-トリミング(図11)、J-トリミング、CDR3長(図12)、及び、N-付加(図13)を含む数種類の分布解析を提供します。これらの解析では、正規化分布も得られます。規則的分布と正規化分布の違いは、どの様にデータをカウントしたかによります。規則的分布は、読み取りカウントデータから直接確認された数を基本としております。正規化分布は、各明確なCDR3を1としてカウントし(V, J, N-付加, CDR3長,等)、固有のCDR3が幾つ確認されたかは関係ありません。

V-使用例

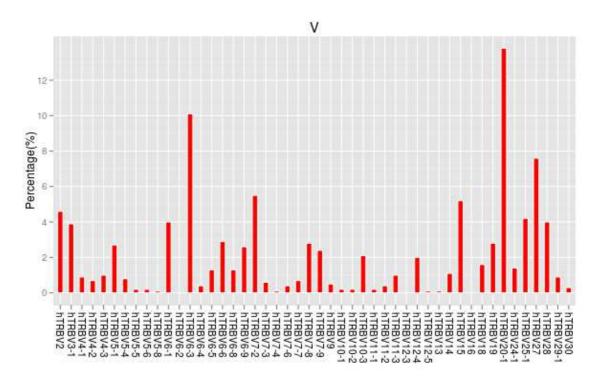


図 10: V-使用率分布。上図は、生殖系列V-アレルを含んでいるリードのパーセンテージで、どのV-アレルがよく使われているかいないかを判定できます。

V - トリミング分布例

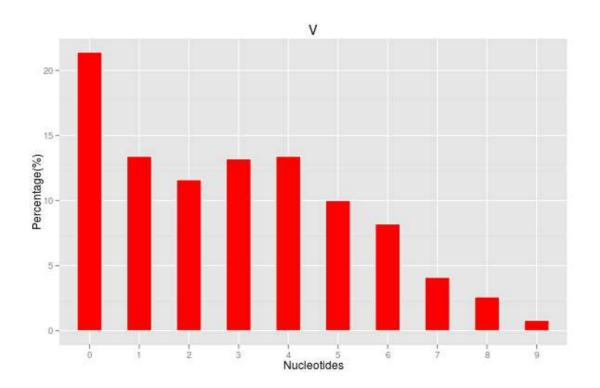


図 11: V-トリミング分布。上図は、V遺伝子中のトリミングされたヌクレオチドのパーセンテージを示しています。例えば、約22%のV-遺伝子は配列中のヌクレオチドのトリミングが無く、一方、約12.5%が1個のトリミングされたヌクレオチドが含まれます。

CDR3長分布例

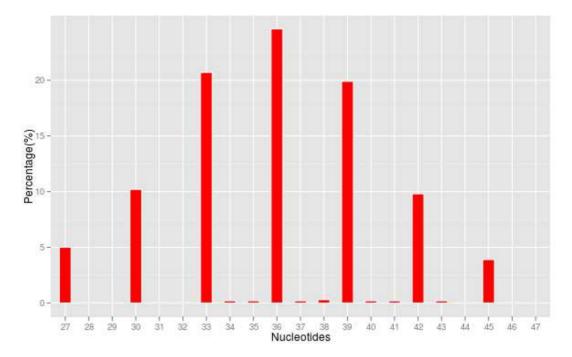


図 12: CDR3長分布。上図は、CDR3領域を構成するヌクレオチドの分布例。例えば、CDR3 配列の約25%が36のヌクレオチドから構成されています。

N - 付加分布例

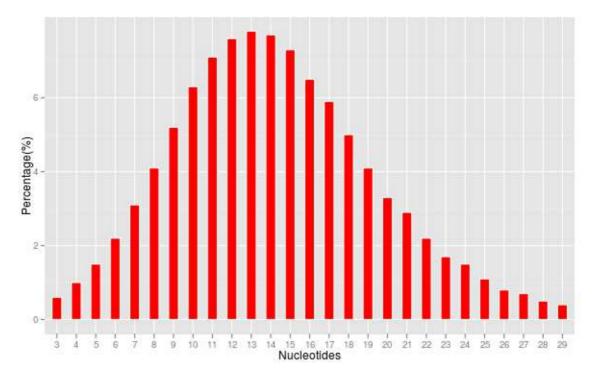


図 13: N-付加分布。上図はN-付加の過程で付加されたヌクレオチドの分布例です。







1. Xu JL, Davis MM: Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 2000, 13(1):37-45.

We Find Health in Your Diversity.

WWW. IREPERTOIRE. COM

