

ジャポニカアレイ[®] ジェノタイピングサービス

ジャポニカアレイ[®] 3つの特長

日本人に特化

短期間で解析

低コスト





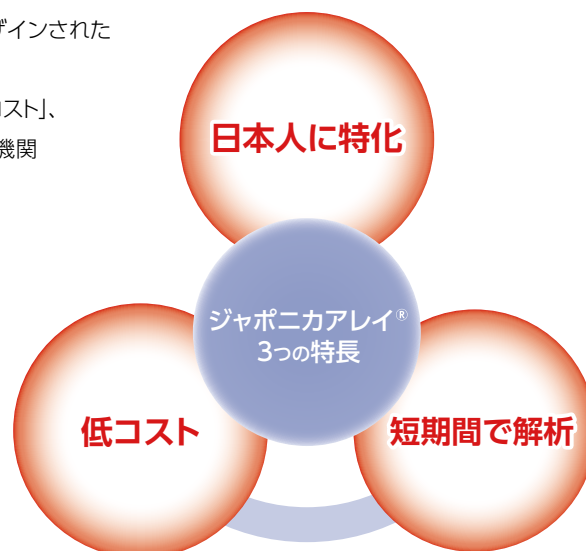
● ジャポニカアレイ[®]*¹ ジェノタイピングサービス*²の概要

ジャポニカアレイ[®]*¹は、国立大学法人東北大学東北メディカル・メガバンク機構 (ToMMo) が構築した「全ゲノムリファレンスパネル (1KJPN パネル)」を基に、COI東北拠点が社会実装した日本人ゲノム解析ツールです。日本人に特徴的な塩基配列を持つ約66万箇所の一塩基多型 (SNP:スニップ) を一枚のチップに搭載しており、短期間で日本人のゲノム情報を解析するツールです。その解析結果から約30億塩基の全ゲノム情報を疑似的に再構成 (インピュテーション*³) できる設計となっています。

ジャポニカアレイ[®]*¹は高品質かつコスト競争力のある米・サーモフィッシュャーサイエンティフィック社製の Axiom[™] プラットフォームを採用し、独自のコンテンツでカスタムデザインされたアレイです。

ジャポニカアレイ[®]*¹ ジェノタイピングサービスは「日本人に特化」、「低コスト」、「短期間で解析」を特長としており、大学、病院、製薬企業などの研究機関向けにご提供いたします。*²

- *¹ ジャポニカアレイ[®]は国立大学法人東北大学の登録商標です。
- *² ジャポニカアレイ[®] ジェノタイピングサービスは研究者向けサービスであり、研究用途での活用を目的としています。本サービスの解析結果は、診療・診断用途ではありません。
- *³ インピュテーションはジャポニカアレイ[®] ジェノタイピングサービスには含まれません。インピュテーションサービスは、ジャポニカアレイ[®] ジェノタイピングサービスとは別にToMMoが実施いたします。また、使用するリファレンスパネルは1KJPN以降の場合があります。詳細につきましては、ToMMoのホームページをご覧ください。
URL:<http://www.megabank.tohoku.ac.jp/tommo/genomepf/imputation>



● ジャポニカアレイ[®]の構成①

ジャポニカアレイ[®]には、インピュテーションの性能を最適化するために、「全ゲノムリファレンスパネル (1KJPN)」の中のSNPのランキング付けの結果、約64万個のタグSNPを搭載されています。さらに、薬剤応答性に関するSNPなど、過去の知見に基づいたSNPを約2万個搭載しており、既知のマーカーを利用した解析も行うことが可能です。

ジャポニカアレイ[®]に搭載されているSNPリストは、下記東北大学 東北メディカル・メガバンク機構インシリコ解析室のウェブサイトで公開中です。URL:<http://nagasakiilab.csml.org/ja/japonica>

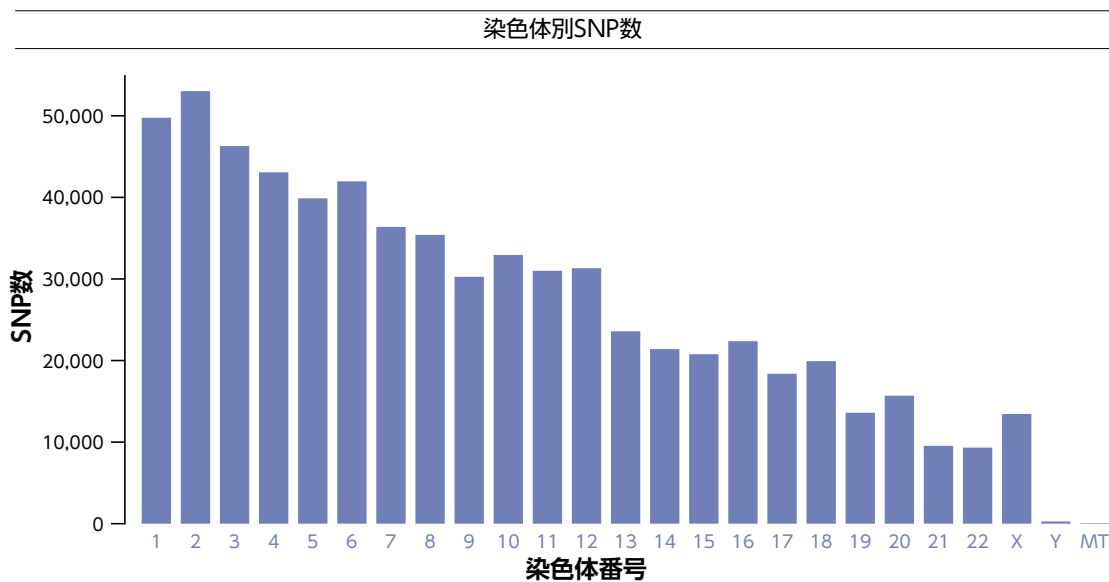
ジャポニカアレイ[®]に搭載されているSNPの内訳

カテゴリ	SNP数	割合
タグSNP	638,269	96.8%
薬剤応答関係 (ADME)	2,028	0.31%
Y染色体	275	0.04%
ミトコンドリア	70	0.01%
NHGRI GWAS Catalog	10,798	1.64%
HLA	3,906	0.59%
その他	3,990	0.61%
合計 (上記は重複を含む)	659,253	—



ジャポニカアレイ[®]の構成②

ジャポニカアレイ[®]は「全ゲノムリファレンスパネル(1KJPN)」のゲノム情報をリーズナブルなコストで引き出せる設計です。常染色体上のタグSNPは各染色体上に均等に分布しており、搭載SNPの平均距離は4.2キロベースと緻密な配置です。日本人集団のゲノムワイド関連解析に最適ですが、がんの染色体コピー数異常やバイオバンク検体の個人識別など多様な目的に使用可能です。



ジャポニカアレイ[®]の構成③

疾病や形質に関連するゲノム多型を効率的に探索するには遺伝子領域(イントロン含む)に多数のプロープを配置するのが効率的です。ジャポニカアレイ[®]は全体の約45%が遺伝子領域+1kbの範囲にあります。多型の多いイントロンと連鎖不平衡にある重要な希少変異はインプテーションによって探索可能です。また、近年その機能が着目されている非コードRNA遺伝子に存在するSNPも多数搭載されています。

ジャポニカアレイ[®]に搭載されている遺伝子領域のSNPの内訳

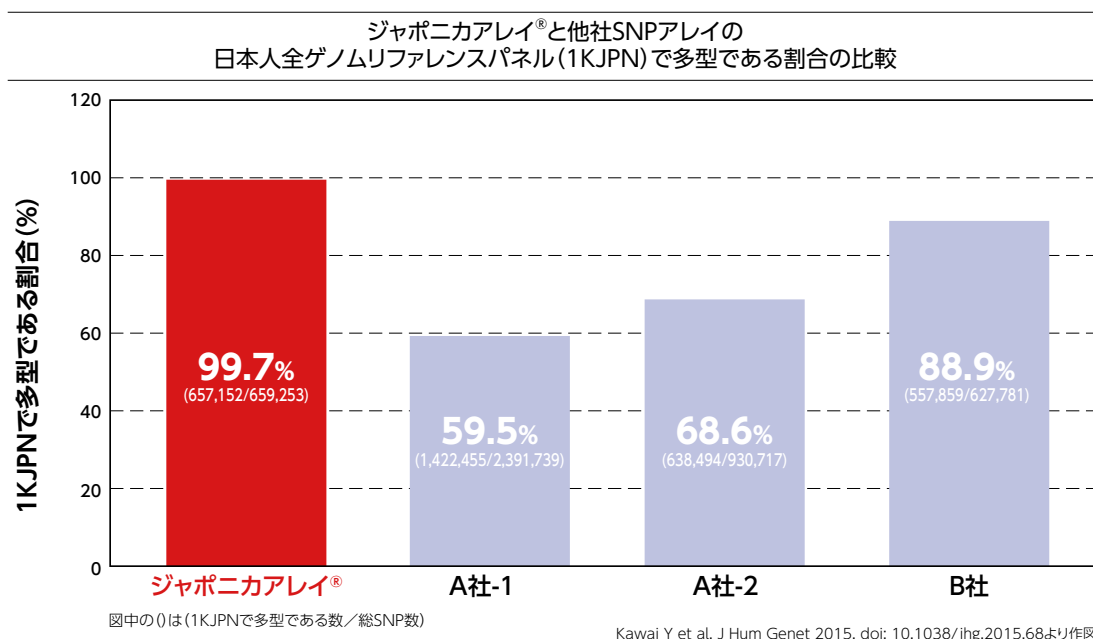
属性情報	定義	搭載SNP
エキソン領域	タンパク質コード領域の多型	12,304
スプライシング	スプライスジャンクションから2塩基対以内の多型	62
非コードRNA	タンパク質をコードしない転写領域上の多型	37,651
5'非翻訳領域	タンパク質コード遺伝子の転写物の5'非翻訳領域	1,041
3'非翻訳領域	タンパク質コード遺伝子の転写物の3'非翻訳領域	6,471
イントロン	イントロン中の多型	239,136
遺伝子上流領域	転写開始点の上流1KB以内の多型	3,978
遺伝子下流領域	転写開始点の下流1KB以内の多型	4,450
遺伝子領域合計		304,093

東芝調べ



● 日本人のSNP解析に適したジャポニカアレイ[®]の構成

ジャポニカアレイ[®]で解析可能な約66万個のSNPのうち99.7%が日本人の全ゲノムリファレンスパネル(1KJPN)の中で多型であり、ジャポニカアレイ[®]は日本人のSNP解析に適しています。



● ジャポニカアレイ[®]のコールレート

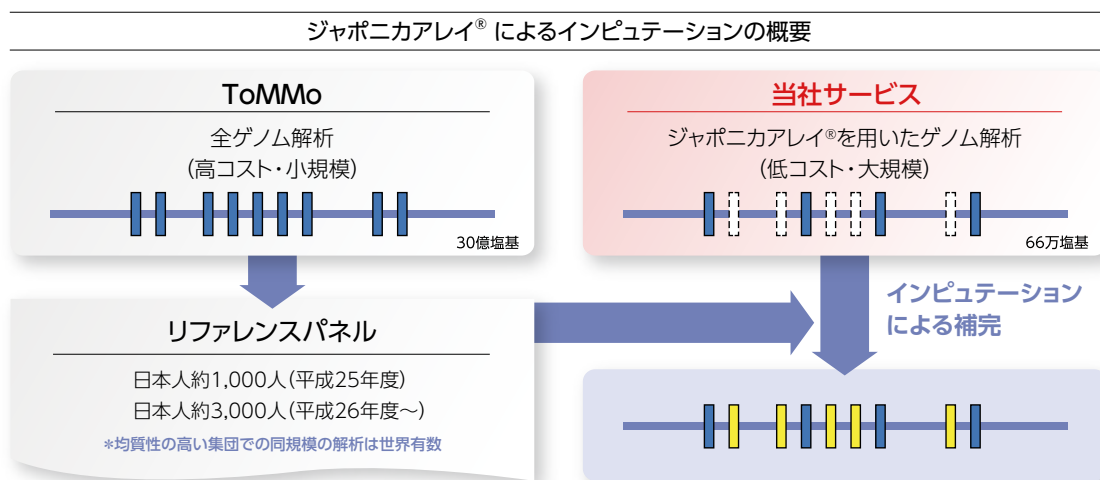
ジャポニカアレイ[®]のコールレートは99.475%^{※4}です。

※4 東芝内でのDNA検体タイピング結果の平均値です。コールレートは検体により異なります。



ジャポニカアレイ[®]によるインピュテーション^{※5}

インピュテーションとはジェノタイプの欠損をもつ解析データに対し、既知の参照多型パネルなどを利用してジェノタイプ推定を行い、より精度の高い解析結果を推測することです。ジャポニカアレイ[®]で得られた解析結果に対し、「全ゲノムリファレンスパネル(1KJPN)」を参照して、全ゲノム解析に近い情報を得ることが可能です。

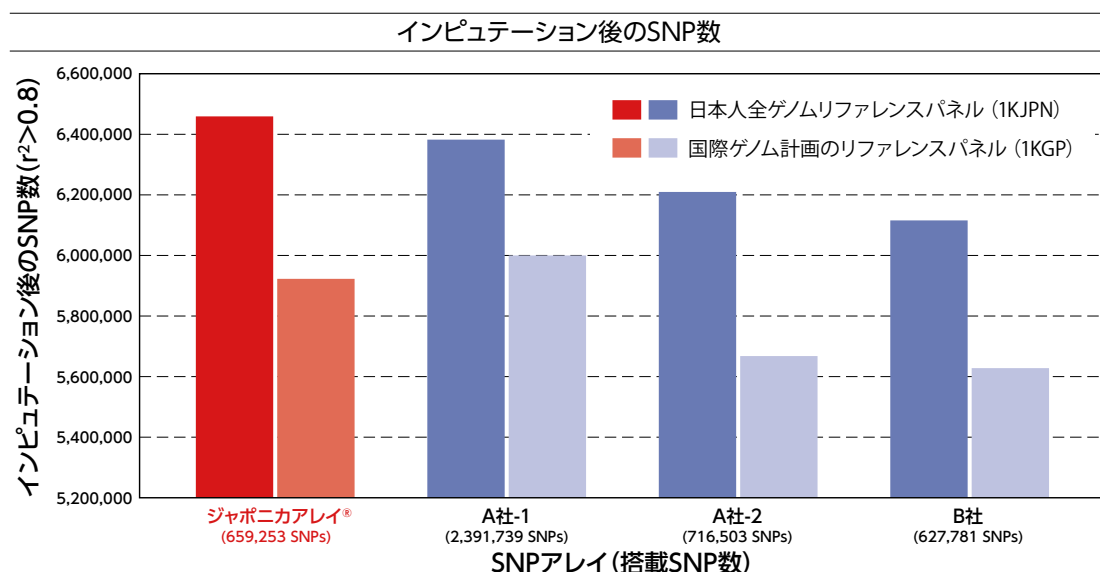


※5 インピュテーションはジャポニカアレイ[®]ジェノタイプングサービスには含まれません。インピュテーションサービスは、ジャポニカアレイ[®]ジェノタイプングサービスとは別にToMMoが実施いたします。また、使用するリファレンスパネルは1KJPN以降の場合があります。詳細につきましては、ToMMoのホームページをご覧ください。
URL:<http://www.megabank.tohoku.ac.jp/tommo/genomepf/imputation>



インピュテーション後のSNP数

ジャポニカアレイ[®]を用いてゲノム解析を行い、その後日本人全ゲノムリファレンスパネル(1KJPN)を参照にインピュテーションを行うことで、アレイに搭載されているSNP数(約66万個)の約10倍の646万個のSNP遺伝子型が得られました。



Medical Science Digest 2015; 41 (5) 226-229より転載(一部改変)



全ゲノムインピュテーションの性能

ジャポニカアレイ®によるインピュテーションの性能について、検体(ToMMo 131検体及びHapMap JPT 89検体)の解析により得られたSNP遺伝子型と日本人全ゲノムリファレンスパネル(1KJPN)中のSNP遺伝子型との相関性を検討しました。いずれの検体を用いた場合にも、高頻度SNP(MAF \geq 5%)については相関係数の二乗(r^2)の平均値が0.97を上回り、相関係数の二乗(r^2)が0.8を超えるSNPの割合は96%を上回りました。低頻度SNP(1% $<$ MAF $<$ 5%)についてはいずれの検体を用いた場合にも相関係数の二乗(r^2)の平均値が0.82を上回り、相関係数の二乗(r^2)が0.8を超えるSNPの割合は70%を上回りました。

全ゲノムインピュテーションの性能

ToMMo 131 検体を使った検証結果

Category	MAF	Mean r^2	Coverage($r^2>0.8$)
高頻度SNP	MAF \geq 5%	0.972	96.9%
低頻度SNP	1% $<$ MAF $<$ 5%	0.822	70.1%

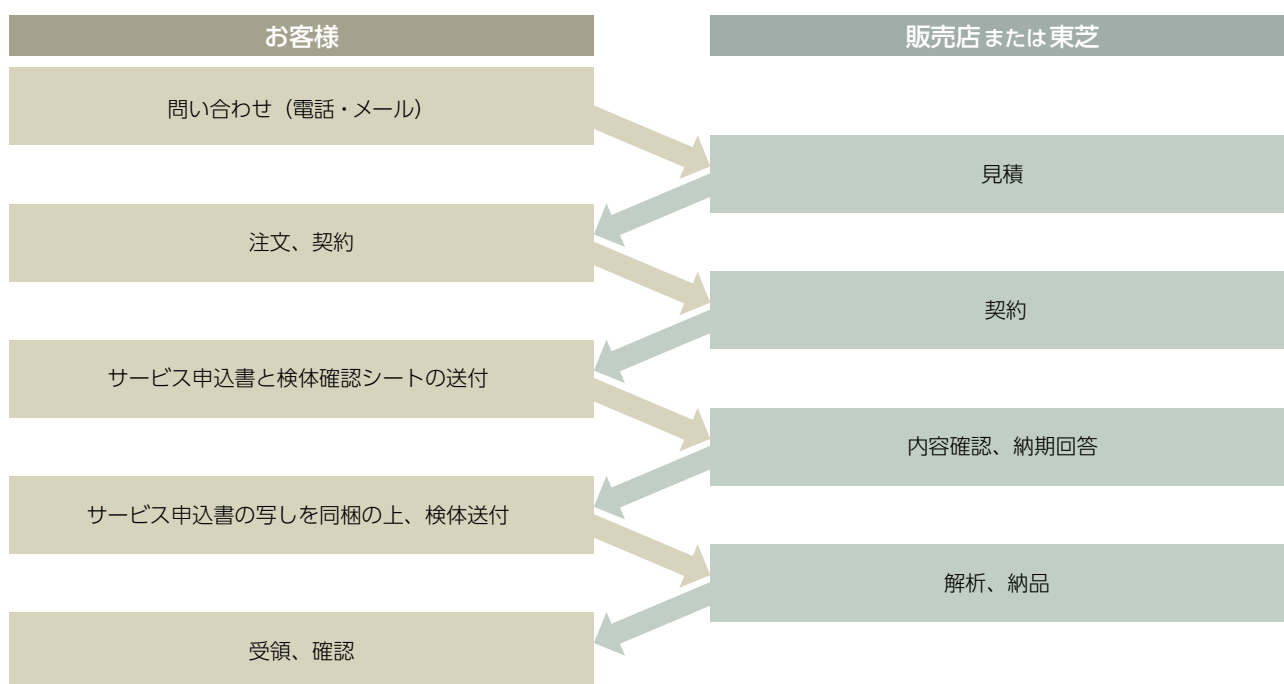
HapMap JPT 89検体を使った検証結果

Category	MAF	Mean r^2	Coverage($r^2>0.8$)
高頻度SNP	MAF \geq 5%	0.976	96.7%
低頻度SNP	1% $<$ MAF $<$ 5%	0.845	71.3%

Kawai Y et al, J Hum Genet 2015, doi: 10.1038/jhg.2015.68



サービスフロー



検体条件

- 96検体をサービスの単位としております。
- 解析する検体は検体提供者からのインフォームドコンセントの取得及び倫理審査委員会の承認を得ていることを条件とさせていただきます。

1. DNA抽出済検体の場合

検体を入れる容器は0.2 mL 96ウェルPCRプレートと致します。

ゲノムDNA濃度	≥50 ng /μL ^{※6, ※7}
溶媒	いずれも、分子生物学グレード。 ○Nuclease (DNase/RNase) -free water (DEPC-treated waterは不可) ○Low TE buffer, pH 8.0 (10 mM Tris-HCl / 0.1 mM EDTA) ○TE Buffer, pH 8.0 (10 mM Tris-HCl / 1 mM EDTA) ^{※6, ※7}
送付容量	≥20 μL (≥1.0 μg) ^{※6, ※7}
A260/A280	>1.6 ^{※7}
鎖長	≥10kbp ^{※7}

2. 全血、パフィーコートの場合

送付容量	≥1.0 mL ^{※6, ※7}
抗凝固剤	ヘパリン、CPD、EDTA ^{※6, ※7}
検体の形態	全血、 パフィーコート

3. 唾液検体の場合

送付容量	2 mL (容器にある採取ラインまでの量) ^{※8}
送付容器	Oragene (OG-500) ^{※8}
唾液採取方法	Orageneが定める使用方法により採取 ^{※8}

※6 検体が条件に満たない場合は、販売店または東芝までお問い合わせください。

※7 東芝が検体受け入れ後に検体品質の確認を行います。その上で上記の品質を満たさない場合は、本サービスの受託を辞退させて頂く場合がございます。その場合でも販売店及び東芝で発生した費用(検体品質確認を含む)はお客様よりお支払頂きます。またこの場合、お客様より送付頂いた血液、唾液またはDNA検体の残余分の返却は別途費用によりお受け致しますが、販売店及び東芝は返却する血液、唾液またはDNA検体の品質の保証は致しかねます。

※8 唾液からのDNAの抽出は検体の状態により異なる為、DNA収量を保証するものではありません。唾液採取は必ずOrageneが定める方法に則って採取してください。

解析

プラットフォーム	Axiom™(サーモフィッシャーサイエンティフィック社)
96検体のデータサイズ	genotyping result: 約2.2GB、SNP summary table: 約70MB
納入の形態	HDDドライブ ^{※9} (容量は検体数による)

※9 HDDドライブは、サービス標準価格には含まれておりません。

納品物

- 解析結果報告書
- お客様の検体管理番号と東芝の検体管理番号の対応表
- 検体の品質確認結果(受入れ時および解析工程中)
- ジェノタイプング解析結果
- その他(アノテーションファイル等)

■納入データ構成

genotyping_result.txt の各列

項目	概要
Probe Set ID	SNP検出に用いられるプローブセットに対するサーモフィッシャーサイエンティフィック社のユニークID
[結果表示用ID]_call_code	Forward StrandにおけるアレルAの塩基とアレルBの塩基の組み合わせで示したジェノタイプ (AT, CG, AG, TC, ..., etc)
[結果表示用ID]_confidence	ジェノタイプコールに対する信頼性を表す値。0に近いほど信頼性が高く、1に近いほど信頼性が低い。0.15以上のジェノタイプはコールされない(No call)
[結果表示用ID]_log_ratio	アレルAとアレルBのシグナル値の対数比
[結果表示用ID]_strength	アレルAとアレルBのシグナル値の対数の平均
Affy SNP ID	SNP検出に用いられるプローブセットに対するサーモフィッシャーサイエンティフィック社のユニークID
Chr_id	染色体番号
Chromosomal Position	染色体上の位置
Chromosome Start	染色体上の開始位置
Chromosome Stop	染色体上の終結位置
Strand	SNPが存在するゲノムDNA鎖の向き
dbSNP RS ID	NCBIのdbSNPデータベースに登録されたSNP登録番号
dbSNP Loctype	NCBIのdbSNPにおけるSNPのLocType
In Hapmap	Hapmapプロジェクトへの登録の有無
Strand Versus dbSNP	dbSNPに登録されているデータと同一鎖か相補鎖かを表示
Probe Count	各プローブセットに含まれるプローブ数
Cytoband	SNPがマップされるサイトバンド領域
Flank	SNP近傍の塩基配列(33塩基)。SNPの位置は[]で表される
Allele A	プローブを設計したゲノムDNA鎖がdbSNPデータベースのRefSNPと同じDNA鎖でない場合は、サーモフィッシャーサイエンティフィック社が公開しているプローブ情報(NetAffx™)に記載されているSNPの塩基配列がこれらのデータベースと異なるケースがある。
Allele B	プローブを設計したゲノムDNA鎖がdbSNPデータベースのRefSNPと同じDNA鎖でない場合は、サーモフィッシャーサイエンティフィック社が公開しているプローブ情報(NetAffx™)に記載されているSNPの塩基配列がこれらのデータベースと異なるケースがある。
Ref Allele	リファレンスアレル
Alt Allele	多型アレル
Associated Gene	SNPの近傍にある遺伝子
Genetic Map	三つのリンケージマップ (deCODE, Marshfield, SLM) から導かれる遺伝的位置
Microsatellite	SNPに対して最も近いマイクロサテライトマーカー (upstream, downstream, overlapping)
Heterozygous Allele Frequencies	Yoruba, Japanese, Han Chinese, CEPH studiesからのヘテロアレルの頻度
Allele Frequencies Count	Yoruba, Japanese, Han Chinese, CEPH studiesからのメジャーアレル数とマイナーアレル数
Allele Frequencies	Yoruba, Japanese, Han Chinese, CEPH studiesからのメジャーアレル頻度とマイナーアレル頻度
Minor Allele	SNPのマイナーアレル
Minor Allele Frequency	SNPのマイナーアレル頻度
OMIM ID	NCBIのOMIMデータベースの登録ID
Biomedical	本サービスでは使用しない
Annotation Notes	本サービスでは使用しない

SNP_summary_table.txt の各列

項目	概要
Probe Set ID	SNP検出に用いられるプローブセットに対するサーモフィッシャーサイエンティフィック社のユニークID
Minor Allele Frequency	No callを含まないすべてのジェノタイプコールにおけるアレルAの頻度とアレルBの頻度のうち、頻度の低いもの
H.W. p-Value	Hardy Weinberg p値
affy_snp_id	SNP検出に用いられるプローブセットに対するサーモフィッシャーサイエンティフィック社のユニークID
CR	全サンプルを通じたSNP判定率
FLD	contrast軸におけるヘテロクラスタと二つのホモクラスタの距離のうち小さい方。何れかのクラスタが空であればNAとなる。
HomFLD	contrast軸におけるホモクラスタ同士の距離。二つのホモクラスタがなければNAとなる。
HetSO	size軸におけるヘテロクラスタ中心のホモクラスタ中心からの距離
HomRO	contrast軸における原点と原点に最も近いホモクラスタ中心の距離
nMinorAllele	マイナーアレルの数
Nclus	クラスタの数
n_AA	アレルAAの数
n_AB	アレルABの数
n_BB	アレルBBの数
n_NC	No Callの検体数
hemizygous	SNPがHemizygousであるとき1、そうでないとき0
HomHet	一つのホモクラスタと一つのヘテロクラスタの二つのクラスタを持つSNPを識別するHomHet尺度を利用するとき1、そうでないとき0
ConversionType	QCメトリックに基いたSNPの分類結果。PolyHighRes, MonoHighRes, Off-Target Variant, CallRateBelowThreshold, NoMinorHom, Other, Hemizygousの何れかに分類される。PolyHighRes, MonoHighRes, NoMinorHom, Hemizygousの利用を推奨
BestProbeset	プローブが同じSNPに割り当てられている際にConversionTypeに基づいて決められる優先順位。値が1であるプローブの利用を推奨。
Call Modified	本サービスではすべてFALSE

改良等のために本サービスの内容の一部を変更することがありますので、ご了承ください。

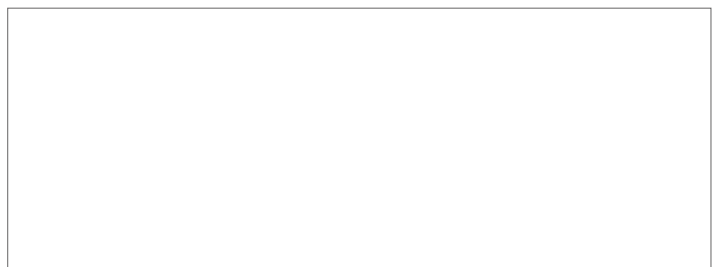
TOSHIBA

株式会社 **東芝** 研究開発本部

本部企画部 ライフサイエンス推進室

〒105-8001 東京都港区芝浦1-1-1 TEL: 03-3457-2984

Mail: HdqLS-PSG@ml.toshiba.co.jp



※この資料の内容は、2018年8月現在のものです。改良等のために本サービスの内容の一部を変更することがありますので、ご了承ください。