

平成 21 年 5 月 12 日現在

研究種目：基盤研究(C)

研究期間：2007～2008

課題番号：19590667

研究課題名（和文） 個人識別実務に即した SNP タイピングの迅速化および効率化の検討

研究課題名（英文） Promotion of rapid and high-throughput SNP typing for human identification in Forensic practices

研究代表者

橋谷田 真樹 (HASHIYADA MASAKI)

東北大学・大学院医学系研究科・技術専門員

研究者番号：40374938

研究成果の概要：

次の single nucleotide polymorphism(SNP)データベースを構築し、その個人識別有用性を確認した。常染色体上 SNP150 座位(100 名)、Y 染色体上 SNP24 座位(243 名)、およびミトコンドリア DNA 上の SNP24 座位(215 名)である。さらに、通常は約 45 分かかかる Fast TaqMan 法の反応時間を検討し、DNA の抽出からタイピング結果を得るまで、わずか 30 分しか要さない反応系を確立した。また、Microsoft Excel によるスプレッドシートを新たに作製することで、一度に最大 96SNP の解析を可能にし、識別精度の向上が図られた。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	2,400,000	720,000	3,120,000
2008 年度	1,200,000	360,000	1,560,000
年度			
年度			
年度			
総計	3,600,000	1,080,000	4,680,000

研究分野： 医歯薬学

科研費の分科・細目： 社会医学・法医学

キーワード： DNA多型医学, SNP

1. 研究開始当初の背景

高度情報化社会の現代では、本人であることを証明・認証する技術の確立が非常に重要な課題である。人が自分自身を証明するためには、自らが唯一持つ生物学的情報（生体情報）を提示すればよく、それを自動的に行うのがバイオメトリクス認証技術である。われわれは、生涯不変で自らを絶対的に示すことができる生体情報としてデオキシリボ核酸(DNA)に着眼し、研究を重ねてきた。しかしながら DNA 情報を生体認証として利用するためには、検査時間と倫理的な側面という 2

つの大きな解決すべき問題が残されている。

当分野ではこれまでの生体認証への応用研究において、DNA 識別情報をハッシュ関数処理することにより、個人情報保護対策も十分可能となることを確認した。さらに、解析時間の短縮のために、近年分子生物学の領域で注目されている single nucleotide polymorphism(SNP)マーカーに着目した。SNP とは、個人間において、ある場所の塩基配列が一塩基だけ違うことを意味し、疾患の原因や薬剤応答性における個人差の要因とされているものである。そのため主に臨床医

学を中心に研究が進められている。個人識別の観点から SNP を見ると、識別精度は劣るものの、多数部位の高速解析が可能である。技術革新により検出の高速化・経済化が見込まれる SNP 情報の利用は大いに期待でき、そのシステムが確立されれば、バイオメトリクス認証技術としての応用だけではなく、迅速かつ効率的な個人識別システムを構成することが可能と考える。しかし現時点では SNP は 1 座位における識別精度が劣るため、精度の向上には多くの SNP 座位を検査する必要があるが、われわれ日本人のデータベースは未だ不十分である。そこでデータベースを整備し、かつ迅速かつ効率的な SNP 解析システムの開発をめざし、本研究の申請に至ったわけである。

ちなみに SNP による個人識別データベースの構築およびそのバイオメトリクス認証技術への応用については国内だけではなく、世界に先駆けて当分野が手がけた点も強調しておく。

2. 研究の目的

個人識別を目的とした SNP のデータベースは世界的に見ても確立されておらず、日本人試料によるデータベースの構築が第一の目的である。当分野にはこれまで 100 人の DNA 試料による常染色体上 SNP96 座位のデータが収集されているが、人数・座位数ともに不十分であり、これらを充実させる必要がある。その際、常染色体上のみならず、男性特異的な Y 染色体上に存在する SNP(Y-SNP)、また、母系遺伝を示すミトコンドリア DNA(mtDNA) 上に存在する SNP(mtDNA-SNP)も対象とする。

第二の目的は解析時間短縮の検討である。SNP 解析には多種多様な検出方法が報告されているが、本研究では解析時間が最も短い Fast TaqMan®法を採用し、解析時間のプロトコルに変更を加え、検査時間の短縮を試みる。最終的には DNA 試料の採取から SNP データを得るまで、30 分しか要さない検査系を確立する。

第三の目的は識別精度向上の検討である。SNP は通常、二対立遺伝子の多型であり、1 ヶ所の SNP 解析ではヒトを 3 タイプにししか識別することができない。そのため多数の SNP 座位を検査し、識別精度を上げる必要がある。Fast TaqMan®法では、解析機器のハード的な問題から、一度に検査できるのは SNP24 座位が限界であり、識別精度が十分とは言えないという問題が残されていた。そこで、新たなスプレッドシートを作製し、機器に付属している解析ソフトと併用することで、一度に最大 SNP96 座位の解析を試みる。

3. 研究の方法

いずれの SNP 解析においてもゲノム DNA 約 5ng をテンプレートとし、Fast TaqMan®法により解析を行った。解析機器には ABI PRISM 7500 FAST Real-Time PCR System, および StepOne Plus Real-Time PCR System (Applied Biosystems)を用いた。

なお、本研究は東北大学大学院医学系研究科倫理委員会の承認を得ている (受付番号 2005-62)。

(1) SNP データベースの構築

・常染色体上の SNP

100 名の DNA を試料とし、150 座位の SNP 解析を行った。SNP 座位は以下のデータベースより、遺伝子領域以外もしくはイントロン領域から、互いに物理的に距離が離れていることを条件に選択した。

- ・ JSNP database (http://snp.ims.u-tokyo.ac.jp/index_ja.html)
- ・ TaqMan SNP Genotyping Assays database (<https://products.appliedbiosystems.com/>)

・ミトコンドリア DNA 上の SNP

細胞内小器官であるミトコンドリアは、核内の DNA とは全く別に独自の DNA を持つ。また、それらは母系遺伝の形質を持ち、かつ、一つの細胞内に数百から数千個と多数存在することから、非常に古い骨試料などからも多型情報を得ることができる。そのため家系をさかのぼり、その配列の違いを基に母系を確認し、人類学的な研究や大まかな人種の推定を行うことが可能となっている。この配列の違いも SNP と利用して捉えることができることから、日本人試料を用いてその解析方法の確立とデータベースの構築を行った。

日本人 404 名の DNA 試料を用いて解析を行い、プローブは MITOMAP (<http://www.mitomap.org/>)にある A human mitochondrial genome database から、31 座位を選択し作製した。これら試料をアフリカ系、ヨーロッパ系、アジア系の 3 タイプに分けた後、さらにアジア系を細分化するためのマーカーで識別した。

・Y 染色体上の SNP

ヒトの性別は X 染色体と Y 染色体という性染色体によって決定されており、XX だと女性、XY であれば男性となる。この Y 染色体は通常父親から男児へそのまま受け継がれるため、男児と父親間の親子鑑定、身元確認等に広く利用されている。Y 染色体には繰り返し配列や SNP(Y-SNP)も多数存在し、これまでの人類学的な研究から、多型に民族特異性があることが判明している。Y 染色体上の SNP により分類されるタイプをハプログル

ープとよび、A~Rの18タイプに分類されている。その中で日本人はハプログループC,D,Oの3系統に大きく分類され、C系統は北アジア、O系統は東ユーラシア全体に広く分布する一方で、D系統は日本人にのみ高頻度で見られ、日本人に極めて特徴的なタイプとされている。そこで我々は日本人におけるY-SNP多型を短時間かつ効率よく解析するため、上述のFast TaqMan®法を用いて検討を行った。

血縁関係のない243人の日本人男性ボランティアから得られたゲノムDNAを試料とし、International Society of Genetic Genealogy (<http://www.isogg.org/tree/Main06.htm>), およびY Chromosome Consortium (YCC) (http://ycc.biosci.arizona.edu/nomenclature_system/frontpage.html)から次のY-SNPマーカーを選択した。まず日本人によく観察されるハプログループC,D,O系統を分類するために、マーカーのM130, M174, M175を、さらに、日本人では稀に観察されることがあるNおよびQ系統を検出するためにM231, M242を選択した。さらに、識別精度の向上を図るために、頻度の高いC,D,O系統の下流から、C系統で3種(M8, M38, M217), D系統で7種(M15, M179, M116.1, M125, P42, M151, P47), そしてO系統で9種(M119, M95, SRY465, M119, 47Z, M122, M159, M7, M134)のマーカーをそれぞれ選択した。これらにより、C系統は4タイプ、D系統は8タイプ、O系統は10タイプに細分化が可能となった。

(2) 解析時間短縮の検討

DNAの抽出は口腔粘膜を擦過した綿棒から行い、迅速化のためにQuickGene-800 (FUJIFILM)を使用した。この機器を使用することで1サンプルであれば5分以内の抽出を可能であり、同時に作業者の熟練度に関係なく、均一な濃度および精製度のDNAが採取できる。

抽出後のSNP解析では、Fast TaqMan®法の反応時間、すなわち反応回数および反応時間を見直すことで短縮を図った。

(3) 識別精度の向上

識別精度の向上を図るためには、単に解析するSNP座位の数を増やせばよい。しかし、Fast TaqMan®法では、1SNP座位につきポジティブ、およびネガティブコントロールを必要とするため96穴のプレートでは一度に24座位の解析が限界となる。そこで、毎回解析の度にコントロールと一緒に反応させるのではなく、過去に採取したデータとの比較で未知試料のタイピングを行えば、コントロール試料を必要としないため、一度に最大96

座位の解析が可能となる。そのためのソフトウェアであるMicrosoft® EXCELを用いたスプレッドシート(日立ソフトウェアエンジニアリング)を作製した。

4. 研究成果

(1) SNPデータベースの構築

① 常染色体上SNP

本研究における新たな常染色体上SNP54座位のデータを表1に示す。頻度計算、および統計処理解析を行った結果、SNP各座位における同値確率は0.375~0.407を示し、54座位の平均は0.381であり、解析を行ったSNP54座位の総合同値確率は 2.19×10^{-23} と算出された。同値確率とは他人でありながらも同じSNPタイプを持つ確率を意味し、SNPの場合、一つの座位では最も識別精度が高くても0.375である。これは単純に1000人検査すれば375人が同じSNP型を持つことを意味する。そこで、いくつもの座位を解析することにより識別精度を向上させるわけである。また、Arlequin software ver.3.1を用いてexact testを行ったところ、ハーディ・ワインベルグの平衡が保たれていた。さらに、連鎖不平衡検定の結果、同じ染色体上に存在するSNP間の相関は見られず、個人識別を目的としたSNPデータベースの充実に図ることができた。

No.	NCBI SNP Reference	Group ID	allele 1	allele 2	同値確率
1	rs648386	D11S4175	A	C	0.375
2	rs2576036	D18S468	A	C	0.375
3	rs2608848	D4S1597	G	T	0.375
4	rs1866039	D2S2264	C	T	0.375
5	rs4857914	D3S3606	C	T	0.375
6	rs2068998	D1S245	C	G	0.375
7	rs2010451	D8S1769	C	T	0.375
8	rs9361855	D6S257	A	C	0.375
9	rs10158294	D1S233	A	G	0.375
10	rs6883939	D5S2031	A	T	0.375
11	rs9359420	D6S460	C	G	0.375
12	rs155202	D12S345	G	A	0.376
13	rs7082530	D10S185	A	G	0.376
14	rs2235099	D7S2496	A	G	0.376
15	rs10764053	D10S548	G	T	0.376
16	rs729174	D13S265	C	G	0.376
17	rs7606043	D2S347	C	G	0.376
18	rs9867647	D3S1614	A	T	0.376
19	rs468158	D5S647	A	G	0.376
20	rs387205	D13S1320	C	G	0.376
21	rs1995042	D11S905	G	A	0.376
22	rs1428073	D5S422	C	T	0.376
23	rs2189579	D4S414	C	T	0.377
24	rs2070759	D12S1708	T	G	0.377
25	rs7921240	D10S1790	A	C	0.377
26	rs705588	D6S1671	A	C	0.377
27	rs6966595	D7S486	A	C	0.377
28	rs252313	D16S3100	G	A	0.378
29	rs2085292	D4S1587	A	G	0.378
30	rs4717222	D7S506	C	A	0.378
31	rs9021	D8S514	A	G	0.378
32	rs6559872	D9S167	G	T	0.378
33	rs1381996	D3S3681	C	T	0.379
34	rs4662260	D2S112	A	G	0.380
35	rs870875	D1S2878	A	C	0.380
36	rs4834211	D4S1615	C	T	0.381
37	rs2802668	D1S252	A	T	0.381
38	rs2026034	D9S1874	C	T	0.381
39	rs897397	D3S1277	G	T	0.382
40	rs1356050	D8S543	C	G	0.382
41	rs11111069	D12S346	C	G	0.386
42	rs1896797	D15S205	A	G	0.386
43	rs944642	D9S1776	C	G	0.386
44	rs3770161	D2S126	C	T	0.387
45	rs958552	D1S2766	A	G	0.387
46	rs9309664	D2S165	A	G	0.387
47	rs1467817	D6S308	C	T	0.388
48	rs1950284	D14S274	C	T	0.389
49	rs1424642	D2S2369	A	C	0.389
50	rs12523326	D5S471	C	T	0.389
51	rs3751972	D17S1857	A	C	0.391
52	rs7593254	D2S364	A	G	0.394
53	rs701760	D4S406	C	G	0.405
54	rs752353	D9S1817	A	G	0.407

表1 常染色体上54座位のデータ

② ミトコンドリア DNA 上 SNP

解析の結果、404 サンプルは 2 つの大きなハプログループ、すなわち M に 265 サンプル、および N に 139 サンプルが分類された。アフリカ系を示すハプログループ L1/L2, および L3 に分類されるものは見られなかった。ハプログループ M,N の下流では、1 サンプル（ハプログループ H）を除いてヨーロッパ系を示唆する I/X/W, K/U, V, J, T に分類されたものは観察されず、ハプログループ M11, M12 以外のアジア系を示すハプログループに分類された。中でもっとも高い出現頻度は、ハプログループ D で 0.396 であり、次に M7 の 0.129, B の 0.106 と続いた。まとめたものを図.1-1~1-3 に示す。

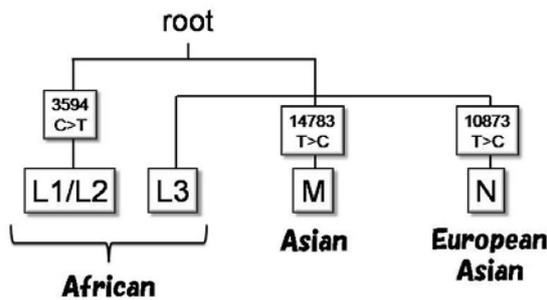


図.1-1 3人種の大まかな分類

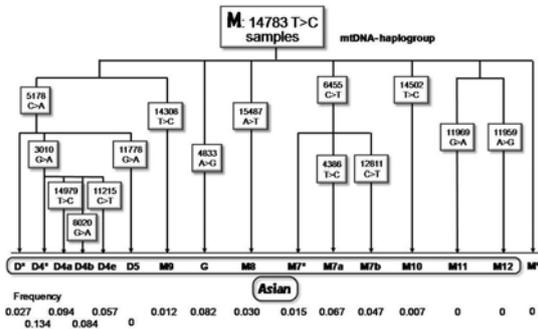


図.1-2 M 系統下流の細分化

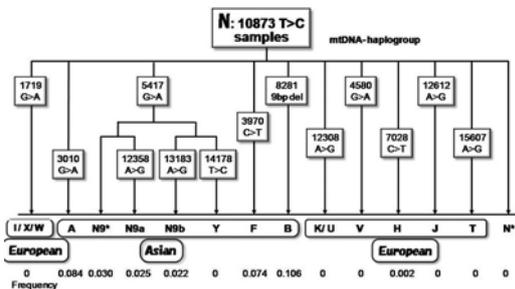


図.1-3 N 系統下流の細分化

以上より、Fast TaqMan® 法による mtDNA-SNPs, 31 座位を識別する解析工程が確立されると同時に、404 名による日本人ミトコンドリアデータベースが構築された。

③ Y 染色体上 SNP

243 サンプルは C 系統 25 名, D 系統 84 名, O 系統 130 名に加えて, N 系統 3 名, Q 系統 1 名に分類され, 出現頻度はそれぞれ C: 0.103, D: 0.346, O: 0.535, N: 0.012, Q: 0.004 であった。アジアにおける他の民族と同じように, O 系統および C 系統が高い頻度で観察されたが, 他の民族と違い日本人には D 系統の頻度が高いのが大きな特徴であった。また, N および Q 系統もごくわずかながら観察された。C, D, O, N, Q 系統による大まかなハプログループの分類に加えて, 日本人に高頻度で観察される C, D および O 系統を細分化の結果を図 4 に示す。C 系統の 25 名は 4 タイプに, また, 日本人で最も高頻度で観察された O 系統 130 名は 8 タイプにそれぞれ細分化された。その結果を図.2-1 ~ 2-3 に示す。

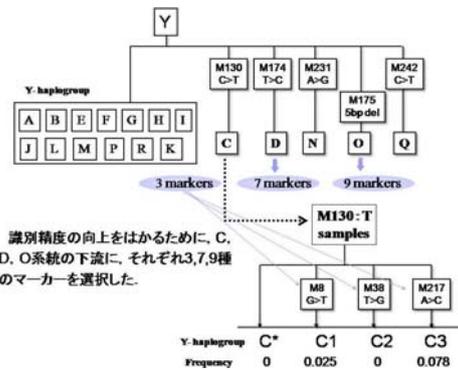


図.2-1 日本人特有のタイプの識別(C,D,O 系統の分類)

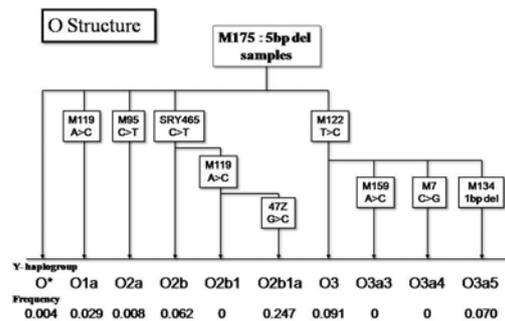


図.2-2 ハプログループ D の細分化

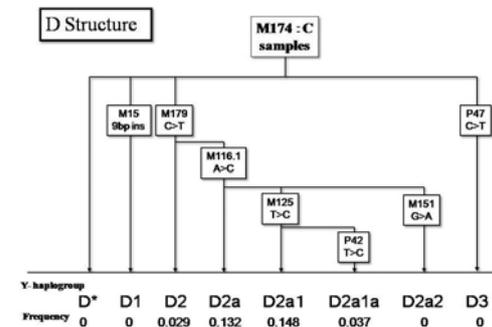


図.2-3 ハプログループ O の細分化

以上より,それぞれの系統において今回選択されたマーカーの有効性が確認された。

(2) 解析時間短縮の検討

TaqMan 法においては,元になるテンプレート DNA を増幅させると同時に SNP タイピングを行うことで時間の短縮が図られている。つまり,基本は PCR 反応であり,反応の内訳はポリメラーゼ酵素による DNA 鎖の複製反応とその繰り返しである。そこで, TaqMan 反応全体の時間を短縮させるためには 1 回あたりの反応時間,もしくは反応回数 of どちらかを減らせばよい。反応回数を 40 サイクルから 35, 30 サイクルと減らしたところ, TaqMan 反応が不十分のため蛍光強度が上がり, SNP の判定が不能となる座位がほとんどであった。一方,複製反応時間については, 62°C・20 秒という反応時間を 15 秒, 10 秒, 5 秒と短縮させて結果を観察したところ,ほとんどの座位で結果に変化は見られなかった。これは反応領域が 150 塩基程度と短いため,プローブのアニーリング,その後続くポリメラーゼによるテンプレート DNA 鎖複製とプローブの破壊による蛍光の発光がスムーズに進んだためと思われる。しかし,やはり反応時間を短くしたことで結果が思わしくなくなるものが 150 座位のうち 10 数座位程度,散見された(図 3)。そのため,識別能力が高く,かつ短い反応時間でも検出が可能な SNP 座位の選択が必要となってくる。

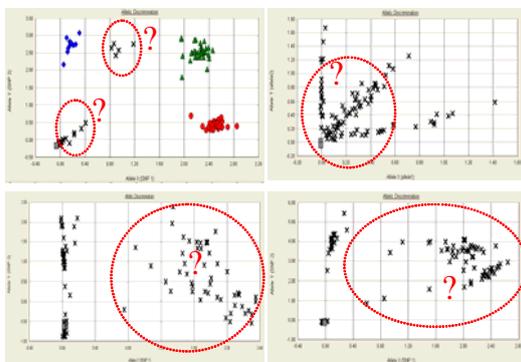


図 3. 条件を変えたことにより蛍光強度がばらついた例

以上の検討から,サイクル数を減らすのではなく,ポリメラーゼによる反応時間を短縮することで,全体の TaqMan 反応時間を 1 時間から 20 分程度に短縮することが可能となった。

(3) 識別精度の向上

一度に多数の SNP を解析するために, Microsoft®EXCEL を用いた新たなスプレッドシートを開発した。以下にスプレッドシートを用いた解析の流れを示す。

- ① ポピュレーションデータを採取した際のデータをそれぞれの SNP 座位ごとに CSV ファイル形式で書き出し,保存しておく。これまでに 120 座位のデータが蓄積されており,これらをテンプレートファイルとする。
- ② あるヒトの SNP を解析し,そのデータを CSV ファイルで保存する。
- ③ われわれが新たに作製したスプレッドシートを立ち上げ,4-1 のデータを読み出し,②のデータと比較・統合し,SNP のタイピングを行う。

図 4 にこれらの概念を示す。それぞれの座位ごとにポジティブ,ネガティブコントロールを用いる代わりに,以前のデータとの比較によるタイピングを行うことで,一度に最大 96 座位の SNP 解析が可能となった。

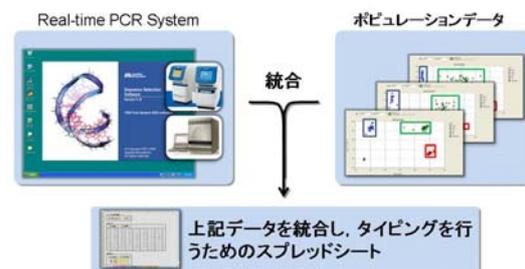


図 4. スプレッドシートを用いた解析の概念図

5. 主な発表論文等

(研究代表者,研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

- ① Masaki Hashiyada, Yukio Itakura, Masato Funayama. Polymorphism of eight X-chromosomal STRs in a Japanese population. Forensic Sci. Int. Genetics Supplement Series 1. 150-152(2008) 査読有
- ② Masaki Hashiyada, Kazuo Umetsu, Isao Yuasa, Akiyoshi Tamura, Aya Matsusue, Koichi Suzuki, Seiichi Kashimura and Masato Funayama. Population genetics of 17 Y-chromosomal STR loci in Japanese. Forensic Sci. Int. Genetics Vol.2:e69-e70(2008) 査読有
- ③ Hashiyada M, Itakura Y, Funayama M. Paradox in the Biometric Personal Authentication System Using STR Polymorphism - Practical Matching Probabilities Evaluate the DNA personal ID System -. Res Pract Forens Med 2007; 50: 235-239. 査読有

[学会発表] (計 10 件)

- ① 橋谷田真樹, 板倉征男, 舟山真人. 新たな

一塩基多型(SNP)情報による個人識別精度の向上. 2009年暗号と情報セキュリティシンポジウム. 大津. 大津プリンスホテル. 2009年1月20~23日.

- ② 橋谷田真樹, 板倉征男, 舟山真人. 日本人におけるSNPsデータベースの充実と識別精度向上の検討. 日本DNA多型学会第17回学術集会. 東京. 私学会館. 2008年11月20, 21日.
- ③ Masaki Hashiyada, Yukio Itakura, Masato Funayama. Haplogroup determination of mitochondrial DNA in Japanese using single nucleotide polymorphisms analysis. The 19th International Symposium on Human Identification. Hollywood, California, USA. October 13-16, 2008.
- ④ Masaki Hashiyada, Yukio Itakura, Shirushi Takahashi, Jun Sakai, Masato Funayama. Development of a spreadsheet for SNPs typing using Microsoft[®] EXCEL. The 7th International Symposium Advances in Legal Medicine. Osaka, JAPAN. September 1-5, 2008.
- ⑤ 橋谷田真樹, 板倉征男, 舟山真人. Fast TaqMan法を用いた日本人におけるY-SNP解析. 日本DNA多型学会第16回学術集会. 2007. 11.15-16. 大阪
- ⑥ Hashiyada M, Itakura Y, Funayama M. Population genetics of eight X-chromosomal STRs in a Japanese population. The 18th International Symposium on Human Identification. Hollywood, California, USA. October 1-4, 2007.
- ⑦ Hashiyada M, Itakura Y, Funayama M. Polymorphism of eight X-chromosomal STRs in a Japanese population. 22nd Congress of the International Society for Forensic Genetics. Copenhagen, Denmark. August 21-25, 2007.
- ⑧ Nakayashiki N, Takamiya M, Shimamoto K, Aoki Y, Hashiyada M. Studies on differentially methylated parental allele in imprinted genes. 22nd Congress of the International Society for Forensic Genetics. Copenhagen, Denmark. August 21-25, 2007.
- ⑨ 境 純, 橋谷田真樹, 高橋識志, 金涌佳雅, 金武 潤, 舟山真人. 複数の人間をSNPで検出するには何座位必要か? 第91次日本法医学会総会. 2007.5.16-18. 秋田.

[図書] (計5件)

- ① 橋谷田真樹, 板倉征男, 舟山真人. Fast TaqMan法を用いた日本人におけるY-SNP解析. DNA多型 16. 東京:東洋書籍, 2008,

p279-281.

- ② 橋谷田真樹, 板倉征男, 長嶋登志男, 境純, 舟山真人. SNPデータベースの構築と解析の高速化. DNA多型 15. 東京:東洋書籍, 2007, p293-295.
- ③ 境 純, 橋谷田真樹, 金武 潤, 舟山真人. 擬父と擬父の祖父のいずれかが生物学的な父親かを問う場合の父権肯定確率計算法. DNA多型 15. 東京:東洋書籍, 2007, p199-200.
- ④ 中屋敷徳, 高宮正隆, 熊谷章子, 青木康博, 橋谷田真樹. DMPA (differentially methylated parental allele)検出法に関する検討 III. DNA 多型 15. 東京:東洋書籍, 2007, p260-263.
- ⑤ 橋谷田真樹, 齊藤 直, 板倉征男, 舟山真人. 一塩基多型(SNP)による個人識別の実験的検証と解析の高速化. 2007年 暗号と情報セキュリティシンポジウム(SCIS2007) 予稿集CD-ROM

6. 研究組織

(1)研究代表者

橋谷田 真樹 (HASHIYADA MASAKI)
東北大学・大学院医学系研究科・技術専門員
研究者番号: 40374938

(2)研究分担者

境 純 (SAKAI JUN)
東北大学・大学院医学系研究科・助教
研究者番号: 20431504

(3)連携研究者