CMV primer R	80 μL
HHV6 primer F	40 μL
HHV6 primer R	40 μL
B19 primer F	60 μL
B19 primer R	60 μL
BKV JCV primer F	40 μL
BKV JCV primer R	40 μL
VZV primer F	20 μL
VZV primer R	20 μL
Total	520 μL

＊2　各プライマーは100 pmol/μLに調製したものを用いる.

❷ 内在性コントロール遺伝子（IC）β-グロビンプライマーミックスの作製

プライマー濃度を4 pmol/μLに調製する.

β-グロビン primer F	4 μL
β-グロビン primer R	4 μL
Nuclease free water	92 μL
Total	100 μL

❸ マルチプレックスPCR用マスターミックスの作製

＜1反応分＞

Primer（❶で調製したもの）	0.60 μL
IC primer（❷で調製したもの）	0.40 μL
Buffer	1.50 μL
定性用増幅酵素	0.25 μL
dH_2O	2.25 μL
Total	5.00 μL

2. 反応液の調製

試薬の調製は，あらかじめ4℃で冷却しておいたLightCycler Centrifuge AdaptersにLightCycler Capillaries（20 μL）を立てて行う.

❶ ミネラルオイルを3 μLずつキャピラリーに入れる

❷ マスターミックスを5 μLずつ入れる

❸ テンプレートを0.2 μg添加してピペッティングで混合し，ヌクレアーゼフリー水で全容量10 μLに調製する＊3

＊3　ここでは全容量10 μLにしているが最大で倍の20 μLまで増やせる.

❹ キャピラリーをアダプターごと高速微量遠心機で1,000×g（3,000 rpm）で3秒遠心する＊4

＊4　キャピラリーの蓋が開いているため，ミストの発生によるコンタミネーションの危険がある.
それを避けるため高速で遠心しないこと.

❺ ハイブリプローブミックスを5μLずつ入れる

各FITC標識プローブおよび各LcRed標識プローブを各0.02 pmol/μLに調製する.

＜Aセットプローブの場合＞

HSV-1,2	FITC標識プローブ*5	2μL
HSV-1,2	LcRed640標識プローブ	2μL
VZV	FITC標識プローブ	2μL
VZV	LcRed640標識プローブ	2μL
B19	FITC標識プローブ	2μL
B19	LcRed640標識プローブ	2μL
HHV-6	FITC標識プローブ	2μL
HHV-6	LcRed705標識プローブ	2μL
CMV	FITC標識プローブ	2μL
CMV	LcRed705標識プローブ	2μL
BKV, JCV	FITC標識プローブ	2μL
BKV, JCV	LcRed705標識プローブ	2μL
β-Globin	FITC標識プローブ	2μL
β-Globin	LcRed640標識プローブ	2μL
Nuclease free water		972μL
total		1,000μL

＊5　各プローブは100 pmol/μLに調製したものを用いる.

❻ キャッピングツールを使用してキャピラリーに蓋をし，カローセルにセットする

カローセルごとLightCycler 2.0にセットし，PCR反応を行う.

❼ マルチプレックスPCRの実行

〈リアルタイムPCR条件〉

TOTAL 40分

			温度変化速度 (℃/s)	データ取得タイミング	
熱変性	95℃	2分	20	Single	
↓					
熱変性	95℃	2秒	20	None	40サイクル
アニーリング	58℃	15秒	20	None	
伸長反応	72℃	15秒	1	None	
↓					
冷却	40℃	30秒	20	None	

3. 融解曲線分析，判定

❶ PCR終了後，カローセルごとLightCycler 2.0から取り出し，LC Carousel Centrifuge 2.0で遠心し，PCR反応液とハイブリプローブミックスを混合する

❷ カローセルを逆さにして暗所で1分静置する

❸ 再びLightCycler 2.0にセットし，融解曲線分析を行う

<融解曲線分析条件>

TOTAL 10分

			温度変化速度 (℃/s)	データ取得タイミング	
ハイブリダイズ	40℃	00秒 *6	20	None	]3サイクル
熱変性	95℃	10秒	20	None	]
↓					
熱変性	95℃	00秒	20	None	
ハイブリダイズ	40℃	20秒	4	None	
解離	80℃	00秒	0.2	Continuous	
↓					
冷却	40℃	10秒	20	None	

＊6　0秒は，設定温度に到達させることが目的である．

❹ AnalysisでTm Callingを選択し，Color Compensation＊7をOnにする

SettingでManual Tmを選び，Tm値＊8を手動で確認し，チャンネルとTm値からウイルスの種類（図2A参照）を判定する．

＊7　Color Compensationデータはあらかじめ取得しておかないと蛍光の漏れこみによりF2/F1，F3/F1の両方に同じピークが現れ判定を誤る恐れがある．

＊8　サンプルの塩濃度が高いとTmも塩濃度に依存して変化するのでTm値全体が上がる．ICのピークを見て例えば1℃高ければ他のウイルスのピークも1～2℃高くなる．LightCycler 4.0のソフトで調節できる．

実験例

1．2本のキャピラリーで合計12種類のウイルスを測定（図2）

A，Bセットで12種類のウイルスについて，図2Aに示されているTm値の差異によって明確に区別でき，検出・同定が可能になる．図2Aにウイルス種類別のTm値の目安を示す．このように，あらかじめポジティブコントロールで実際に検出するTm値を確認しておく必要がある．

もしネガティブコントロールが陽性になった場合はコンタミの可能性がある．試薬をすべて変える．変えても検出する場合はPCRのアニーリング温度を最適化する．あるいは新しいプライマーに変更する．

A

セット	チャンネル	Target	Tm(℃)
A	640 (F2)	IC(β-globin)	51
		HSV-1	56
		VZV	61
		B19	64
		HSV-2	69.5
	705 (F3)	HHV-6	53.5
		CMV	60
		BKV	65
		JCV	70.5
B	640 (F2)	IC(β-globin)	51
		EBV	64
	705 (F3)	HHV-7	56
		HHV-8	62.5
		HBV	66

F1：F1TC
F2：LcRed 640
F3：LcRed 705

B

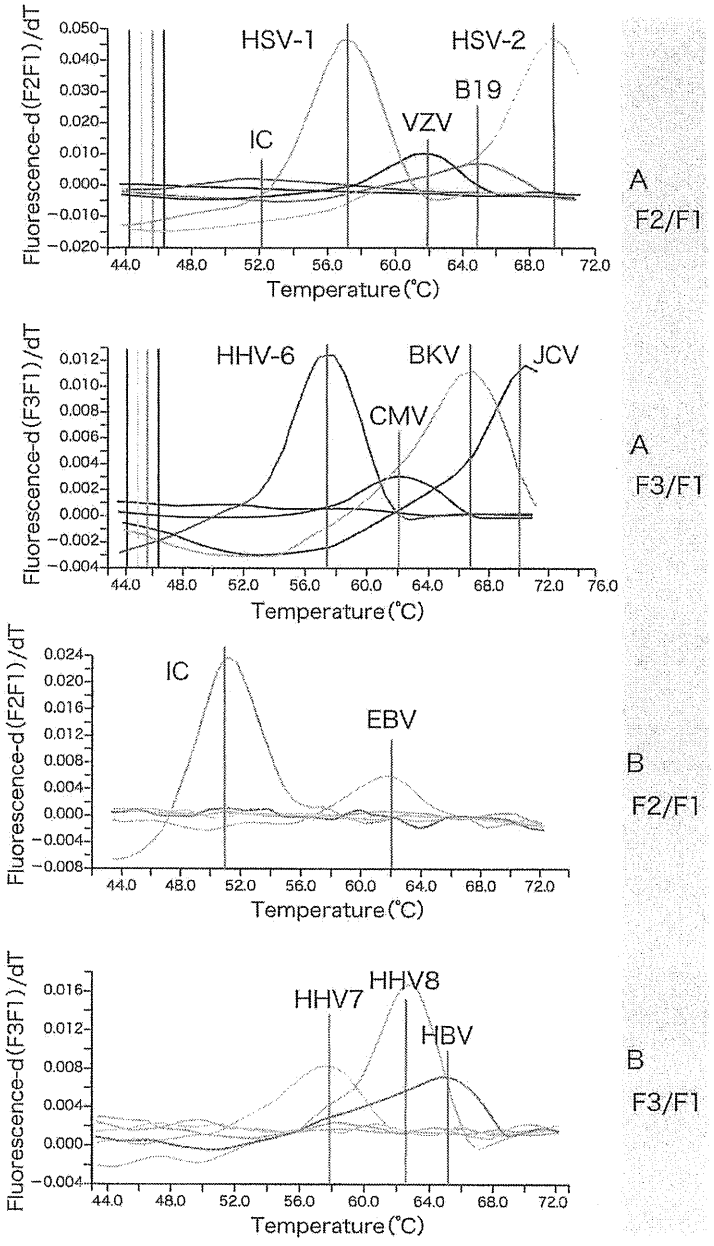

図2　標準DNAを用いたウイルス A，Bセット2本の検出結果

IC遺伝子を含め，ピークが何も検出されない場合は，サンプル抽出がうまくいっていない場合が想定される．

2. 構築されたウイルス検査系を使用した眼疾患検査への利用例

眼科疾患においてブドウ膜炎の原因の多くは自己免疫疾患とウイルス・細菌などによる感染症のいずれかで起こることが知られている．原因によって治療法が全く異なるため，原因を迅速に決定し適切な治療を行うことが患者QOLを確保するうえできわめて重要である．また採取できる眼科検体は微量であり，個別に多くの項目を測定するためには検体を薄める必要があり，検出感度が低下する．したがって，マルチプレックスPCRにより高感度かつ多項目の同時測定

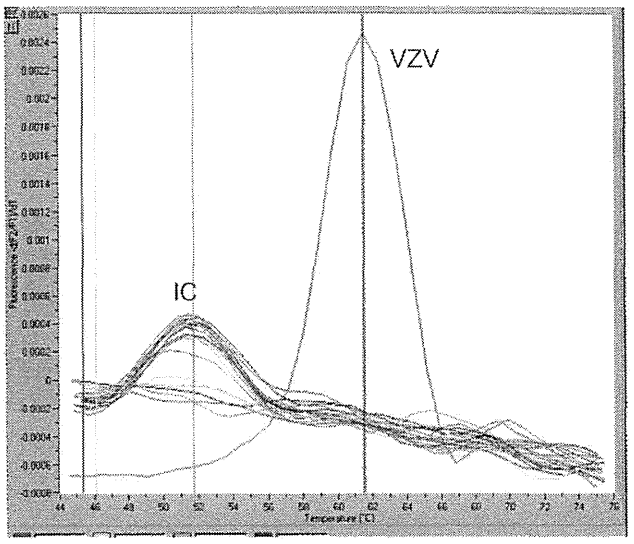

図3　複数の検体の1つからVZVが検出された例

を行うことが望ましい[2]．

図3は複数サンプルを同時に測定した例で，1検体から水痘帯状疱疹ウイルス（varicella-zoster virus：VZV）が検出された．その他のサンプルからはIC（β-Globin）のみが検出され，ウイルス陰性であったことを示している．

固相化試薬とプレートタイプPCR装置を使用した網羅的検査系

プレートタイプPCR装置を使用した網羅的ウイルス検査系では，プローブをPCR反応後に加えることが難しいためマルチプレックスPCRの感度が低下してしまう懸念がある．予備的検討では，1つの反応場で行うPCR反応を3つ程度に抑えればプライマー・プローブの配列をそれほど吟味しなくても良好な結果を得られるとの結果が出ていた．しかし，多項目の検出を行おうとすると多数のウェルを使用する必要があるため，試薬のセットアップに長時間を要する欠点がある．本検査系では，あらかじめプライマー，プローブ，安定化剤等を固相化した試薬を準備することで，短時間で多くの項目を網羅的に検査することが可能となった．また，プローブをはじめから投入するため，半定量的結果を得ることが可能である．

準備

□ 固相化ウイルス測定試薬（固相化ストリップ）

日和見感染症セット（日本テクノサービス社），DNA・RNAウイルス・マイコプラズマ定性試薬セット（日本テクノサービス社）など．プライマー・プローブ，安定材などが固相化されたもの（図4）．50コピーの検出をCt値40以下で行えるように調製されている．

□ サンプルDNA 100 ng

□ リアルタイムPCR装置

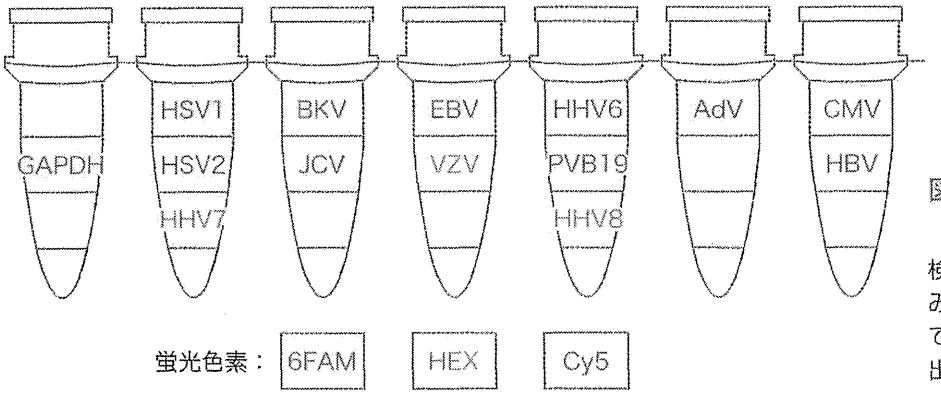

図4　固相化ウイルス測定試薬のイメージ
検出プローブの蛍光色素を組み合わせることで，1ウェルで1～3種類のウイルスを検出できる

LightCycler® 480（ロシュ・ダイアグノスティックス社），CFX96 Touch™ リアルタイムPCR（バイオ・ラッド社），PikoRealリアルタイムPCR（サーモサイエンティフィック社）など．

□ PCR定量用Buffer（#B002，日本テクノサービス社）

□ PCR反応液定量用増幅酵素（#T002，日本テクノサービス社）
詳細については日本テクノサービス株式会社へ直接問合せる．

□ 標準DNA
各検査項目に対応した標準DNA．検査対象のウイルスゲノムなどを取得することが困難である場合が多いため，一般的には，検査する際に増幅させる領域を含むDNA断片をPCRで増幅した産物や，その領域が挿入されたプラスミドを用いる．調製した標準DNAのコピー数は，DNAの濃度と断片の長さから，以下の式で計算する．

$$1\text{コピーの質量（Y g）}=\frac{\text{DNA断片鎖長}\times 660\ (1\text{ bpの平均分子量})}{6.02\times 10^{23}}$$

$$\text{コピー数濃度}=\frac{\text{DNA溶液の濃度（g/}\mu\text{L）}}{1\text{コピーの質量（Y g）}}$$

□ マイクロタイタープレート対応ミキサー

□ 分注機
ウイルスの検査をするにあたって，人為的な間違いやコンタミネーションは結果に甚大な影響を及ぼす．また，一度に多くのサンプルを処理する場合は，その危険性が高くなる．このようなリスクを低減するため当研究室では，小型で安価な自動分注機（Nadeshiko II：#BM-N002，ジーンワールド社）を共同開発・運用している．

プロトコール

1反応分がチューブに固相化された日和見感染症セットの8連ストリップ（8連チューブ）を用いた場合の実験手順を以下に示す．なお，PCRは高感度であるため，コンタミネーションの影響を受けやすい．そのため，反応液を調製する場所は，クリーンに保つことが重要である．当研究室では，検査をする実験室とその他の実験室を別にし，さらにPCR反応液の調製はクリーンベンチ内で行うなど，コンタミネーションのリスクを減らす取り組みを行っている．

❶ 目的に応じたマルチプレックス検出系の選択

❷ 反応液の調製

以下の組成のリアルタイムPCR反応溶液を調製する*1*2．サンプルが多いときには分注機を用いる．

PCR定量用Buffer	9.8 μL
PCR定量用増幅酵素	0.2 μL
サンプルDNA	300 ng
超純水	適量
Total	20 μL

*1 日和見感染症セットは7ウェルで13項目について検査するため，反応液の調製は7～8ウェル分を先に準備して，それを各ウェルに分注する．

*2 8連ストリップには，プライマー・プローブそして安定化材が固相化されている．そのため，反応液を加えたあとよくピペッティング（5～10回程度），あるいはプレートミキサーを使用しよく撹拌し均一化する必要がある．

❸ リアルタイムPCR反応

以下の温度条件でPCR反応を行う．LightCycler® 480，CFX96 リアルタイム PCR，Piko-Realについては，以下の条件で増幅検出できることを確認している．

ポリメラーゼの活性化	95℃	10秒	TOTAL 60分
↓			
熱変性	95℃	10秒	45サイクル
アニーリング	60℃	30秒	

❹ 実験データの解析

リアルタイムPCR反応後，解析装置付属の解析ソフトを用いて増幅曲線を確認し，ウイルスの陰陽判定を行う．

各種装置付属解析ソフトのアルゴリズムは，それぞれ独自のものを使用しているので，それぞれの手順書に従って解析する．ほとんどの解析ソフトが，増幅曲線からCt値を計算するための閾値やバックグラウンドを自動で設定する機能がついているので，これを利用し参考にしながらそれぞれの設定を行うこともできる．

また得られたCt値から，あらかじめ標準DNAを用いて検量線を作成しておくことで，半定量的な情報を得ることができる．

実験例

血液から抽出したDNAと日和見感染症セットを用いて，実際のウイルス検査の結果を図5に示す．本項で紹介したプロトコールを実際の血液サンプルでテストした結果，多くのサンプルから複数のウイルスが検出された．このように，網羅的な検査は，標的を絞った検査では見落とす可能性のあるウイルスを検出することができることが大きなメリットである．

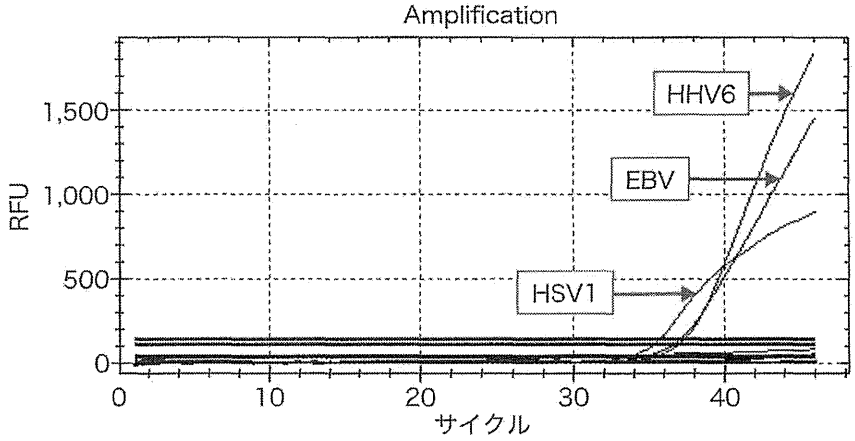

図5　日和見感染症セットを実際の血液サンプルを用いて解析した結果

おわりに

本項では当研究室で主にウイルスの検出を目的として開発したマルチプレックスPCR検出系を解説した．マルチプレックスPCR法による網羅的ウイルス検査は，ウイルス感染症が疑われる疾患の病因特定に有用な情報を与えることができるため，すでに多くの医療施設で利用されている．一方，iPS技術の登場により再生医療に対する注目度が高まっているが，ヒトには多くの病原体が持続感染しているため治療用細胞製剤の原材料には微生物汚染のリスクが避けられない．細胞製剤は滅菌操作をすることが不可能なため，安全に治療を行うためには微生物検査を徹底することが非常に重要であり，マルチプレックスPCRを応用した本検査法は細胞製剤の安全管理法として注目されている．また，本項で記したように，あらかじめ固相化試薬を準備しておけば，さまざまな遺伝子検査を簡便に実施することが可能になる．固相化ストリップの作製技術は日本テクノサービス社と共同開発した成果であり，必要な固相化ストリップの製造を委託することが可能である．本プロトコールに関する技術的な質問は，東京医科歯科大学難治疾患研究所ウイルス治療学 清水則夫（nishivir@tmd.ac.jp）まで．

◆参考文献

1）Ito, K. et al.：Intern. Med., 52：201-211, 2013
2）Sugita, S. et al.：Br. J. Ophthalmol., 92：928-932, 2008
3）『医薬品の品質管理とウイルス安全性』（日本医薬品等ウイルス安全性研究会/編），文光社，2011

IJC
International Journal of Cancer

Match criteria for human cell line authentication: Where do we draw the line?

Amanda Capes-Davis[1], Yvonne A. Reid[2], Margaret C. Kline[3], Douglas R. Storts[4], Ethan Strauss[4], Wilhelm G. Dirks[5], Hans G. Drexler[5], Roderick A.F. MacLeod[5], Gregory Sykes[2], Arihiro Kohara[6], Yukio Nakamura[7], Eugene Elmore[8], Raymond W. Nims[9], Christine Alston-Roberts[2], Rita Barallon[10], Georgyi V. Los[11], Roland M. Nardone[12], Paul J. Price[13], Anton Steuer[14], Jim Thomson[10], John R.W. Masters[15] and Liz Kerrigan[2]

[1] CellBank Australia, Children's Medical Research Institute, Westmead, New South Wales, Australia
[2] American Type Culture Collection (ATCC), Manassas, VA
[3] National Institute of Standards and Technology (NIST), Gaithersburg, MD
[4] Promega Corporation, Madison, WI
[5] Leibniz-Institute, Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Braunschweig, Germany
[6] Japanese Collection of Research Bioresources (JCRB), National Institute of Biomedical Innovation, Osaka, Japan
[7] RIKEN Bioresource Center, Cell Engineering Division, Tsukuba, Japan
[8] Department of Radiation Oncology, University of California, Irvine, CA
[9] RMC Pharmaceutical Solutions, Longmont, CO
[10] LGC, Teddington, Middlesex, United Kingdom
[11] Thermo Fisher Scientific, Rockford, IL
[12] Catholic University of America, Washington, DC
[13] Cell Culture Consultant, Mt. Pleasant, SC
[14] BioReliance Corporation, Rockville, MD
[15] The Prostate Cancer Research Centre, University College London, London, United Kingdom

Continuous human cell lines have been used extensively as models for biomedical research. In working with these cell lines, researchers are often unaware of the risk of cross-contamination and other causes of misidentification. To reduce this risk, there is a pressing need to authenticate cell lines, comparing the sample handled in the laboratory to a previously tested sample. The American Type Culture Collection Standards Development Organization Workgroup ASN-0002 has developed a Standard for human cell line authentication, recommending short tandem repeat (STR) profiling for authentication of human cell lines. However, there are known limitations to the technique when applied to cultured samples, including possible genetic drift with passage. In our study, a dataset of 2,279 STR profiles from four cell banks was used to assess the effectiveness of the match criteria recommended within the Standard. Of these 2,279 STR profiles, 1,157 were grouped into sets of related cell lines—duplicate holdings, legitimately related samples or misidentified cell lines. Eight core STR loci plus amelogenin were used to unequivocally authenticate 98% of these related sets. Two simple match algorithms each clearly discriminated between related and unrelated samples, with separation between related samples at ≥80% match and unrelated samples at <50% match. A small degree of overlap was noted at 50–79% match, mostly from cell lines known to display variable STR profiles. These match criteria are recommended as a simple and effective way to interpret results from STR profiling of human cell lines.

Key words: authentication, human cell lines, cross-contamination, STR profiling, match criteria

Abbreviations: ANSI: American National Standards Institute; ATCC SDO: American Type Culture Collection Standards Development Organization; NCBI: National Center for Biotechnology Information; STR: short tandem repeat

Additional Supporting Information may be found in the online version of this article.

Certain commercial materials, or suppliers, are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Conflicts of interest: The authors note that Douglas Storts and Ethan Strauss are employed by Promega Corporation, a for-profit biotechnology company that develops STR genotyping systems. Rita Barallon and Jim Thomson are employed by LGC; an LGC division, LGC Standards, is a commercial partner of the ATCC. Georgyi Los is employed by Thermo Fisher Scientific. John Masters has been a consultant to Life Technologies.

DOI: 10.1002/ijc.27931
History: Received 15 Jul 2012; Accepted 26 Sep 2012; Online 8 Nov 2012

Correspondence to: Amanda Capes-Davis, CellBank Australia, Children's Medical Research Institute, Locked Bag 23, Wentworthville, NSW 2145, Australia, Fax: +61-2-8865-2801, E-mail: acapdav@gmail.com

What's new?
Short tandem repeat (STR) profiling is the recommended approach for authentication testing of human cell lines. To improve its effectiveness, however, a method by which STR can account for genetic drift arising from the passage of malignant cells is needed. Here, algorithms based on a set of match criteria, eight core STR loci, and amelogenin analysis were found to successfully discriminate between related and unrelated samples. The match criteria used here would bring greater rigor to the interpretation of STR profiling results.

Research into cancer, and many other life-threatening conditions, relies on the culture of continuous cell lines as effective models for the cell type being studied.[1,2] Cell lines cease to be effective models if their behavior changes—for example, due to overpassaging, microbial contamination or misidentification of the cell line.[3,4] Cell line authentication and contaminant testing, and other routine measures to minimize behavioral changes, are an important but often neglected part of good cell culture practice.[2]

A cell line is said to be misidentified when its DNA profile is no longer consistent with the individual donor from whom it was first established. Although misidentification has many causes, including accidental substitution of culture samples, it is often caused by cross-contamination—introduction of another cell line into that culture, resulting in overgrowth by the contaminant. Cross-contamination is a common occurrence within the research community. Ongoing work from the German Collection of Microorganisms and Cell Cultures (DSMZ), evaluating deposited leukemia-lymphoma cell lines ($n = 620$), has shown that 31% of that sample set are contaminated: 18% with mycoplasma, 7% with another cell line and 6% with both.[5,6] Cross-contamination remains a common occurrence even when cell lines are supplied from the originator's laboratory, suggesting that contamination is frequently an early event—often during establishment of the cell line.[5] More than 360 cell lines are known to be cross-contaminated or otherwise misidentified with no known authentic stock, calling into question the validity of any studies involving these cell lines.[6]

Misidentified human cell lines, whether due to cross-contamination or other causes, can be detected through authentication testing. Authentication testing aims to compare the test cell line to other samples (either tissue or cultured cells) from the same donor. Correctly authenticated samples may behave differently in culture (*i.e.*, exhibit phenotypic differences), but testing will show that they come from the same donor and are not cross-contaminated by other commonly used cell lines such as HeLa. A variety of methods exist for authentication testing, but short tandem repeat (STR) profiling has been recommended as the most widely used method currently available for human cell lines.[7,8]

Human STR profiling relies on a polymerase chain reaction (PCR)-based assay examining polymorphic tetranucleotide or pentanucleotide repeats.[9] The varying number of repeats produces amplified DNA fragments of different sizes, which are identified and assigned a numerical value after comparison to a set of size standards. The resulting profile is characteristic of that individual and can be entered into a database, where it is readily compared between laboratories so long as appropriate controls and validation are used.[7,10] STR profiling was initially shown to be effective for international comparison of cell line samples.[7] It was recently recommended by the American Type Culture Collection Standards Development Organization (ATCC SDO) Workgroup ASN-0002 as the best method currently available for human cell line authentication.[8,11] The Workgroup has now published a Standard for authentication of human cell lines by STR profiling, approved by the American National Standards Institute (ANSI), aiming to set out sufficient information for laboratories to perform their own testing or to interpret results obtained from testing laboratories.[12]

As with all authentication methods, STR profiling has advantages and disadvantages. Advantages include extensive data available through population and forensic studies; standardized kits; and implementation of the technique by cell banks worldwide, resulting in publication of STR profiles for the most commonly used cell lines.[9,11] Four of the cell banks have made these STR profiles available through online interactive databases; samples have established provenance, with records of their origin and history kept by the relevant cell bank.[13,14] The combined dataset will be used for a global database currently being developed by the ASN-0002 Workgroup and the National Center for Biotechnology Information (NCBI). Researchers will be able to use the NCBI database to compare the STR results for their cell lines to those for other samples of the same cell line and will be able to add STR profiles from their own testing.[8,11] Some laboratories have undertaken STR profiling of large cell line panels that would fit well with a combined database approach.[15–19] Other groups such as the National Institute of Standards and Technology (NIST) are generating STR profiles specifically for the NCBI database through the Identification of Human Cell Lines Project.[20]

The disadvantages of STR profiling include the inability to fully characterize more than a few species using the available kits. Most primer sets used for cell line authentication amplify only human DNA, although primer sets for other species have been developed and both canine and monkey cell lines have been authenticated using a similar approach.[21,22] Another disadvantage of STR profiling comes from a tendency for some cell lines to undergo genetic drift with continued passage in culture. STR profiles from malignant tissues are known to vary, with loss of heterozygosity and a high incidence of microsatellite instability.[23,24] In culture, genetic drift may be accentuated by suboptimal culture conditions, for example, if cells are

overpassaged or overdiluted when passaged, resulting in selection of variant subclones. Subclone selection through "bottlenecking" can be minimized by the use of low passage cultures, regular feeding and avoidance of overdilution.[25]

Most cell lines show only small amounts of genetic drift.[7,17,26] Such changes generally relate to loss of heterozygosity rather than changes in alleles.[17] However, some cell lines such as CCRF-CEM and Jurkat show considerable variation between subclones, resulting in genetic drift even when good cell culture technique is used.[27,28] CCRF-CEM and Jurkat are widely used, and like HeLa, have often been implicated in cases of cross-contamination.[29] Thus to interpret authentication test results correctly, one must use standardized match criteria to discriminate between "related" (same donor) and "unrelated" (different donor) samples, with some allowance for genetic drift. Any match criteria must also discriminate effectively between cells derived from different donors; if set incorrectly, samples may be said to come from the same donor when that is not the case.

Match criteria generally incorporate an algorithm to compare two STR profiles.[7,16,30–32] One STR profile is defined as the "questioned" profile (the sample being tested), while the other is a "reference" profile, ideally from the same donor. Where another sample from that donor is not available, the questioned STR profile should be compared to other samples from that laboratory and databases online. Comparison to other samples improves the chances of detecting cross-contamination, as most cases of cell line cross-contamination arise from a limited number of commonly used cell lines.[5,6] Various match algorithms have been proposed for comparison of DNA fingerprints, some requiring specialized bioinformatics expertise.[32,33] Although the more complex algorithms can yield a substantial amount of information on authenticity and instability, early validation of STR profiling has shown that simple match algorithms also work well to discriminate between related and unrelated samples.[7,30]

The STR profiles made available by the cell banks generally included eight STR loci plus amelogenin for gender determination.[10] Some cell banks and laboratories now use a larger number, typically 16 loci.[16,31] To make effective recommendations for the human cell line authentication Standard,[12] STR profiles were contributed from four cell banks to give a combined dataset. The duplicates within the dataset were sorted into sets of related (same donor) samples, based on known provenance, and analyzed to determine the minimum number of STR loci when comparing cell line profiles, and the effectiveness of a simple set of match criteria. We also highlighted the commonly used cell lines where match criteria are difficult to apply due to marked genetic drift, and the known misidentified cell lines within the dataset.

Methods

Match algorithms

Three empirical match algorithms were evaluated for STR profile comparison. The sample undergoing authentication is defined as the questioned profile, and a previously authenticated sample used for comparison is defined as the reference profile. Homozygous alleles are counted as one allele.

The first match algorithm is referred to as the "Masters algorithm"[7]:

Percent match = number shared alleles/total number of alleles in the questioned profile.

The second algorithm is a modification of the Masters algorithm:

Percent match = number shared alleles/total number of alleles in the reference profile.

The third match algorithm is referred to here as the "Tanabe algorithm"[30] but is more usually known as the Sørensen similarity index or Sørensen-Dice coefficient[34]:

Percent match = (number shared alleles × 2)/(total number of alleles in the questioned profile + total number of alleles in the reference profile).

Combined dataset of STR profiles

STR profiles for this analysis were contributed by four cell banks: ATCC ($n = 664$), DSMZ ($n = 465$), JCRB ($n = 577$) and RIKEN ($n = 573$), giving a total of 2,279 STR profiles. STR profiles were obtained retrospectively, during authentication testing performed by each cell bank as part of its accession process.

Methods used for DNA extraction varied between the cell banks and over time. Extraction methods included spotting onto FTA cards (Whatman, Maidstone, United Kingdom), using FTA Classic or FTA Elute (formerly IsoCode, Schleicher & Schuell), and use of the High Pure PCR Template Preparation Kit (Roche, Penzberg, Germany). The PowerPlex 1.2 System (Promega, Fitchburg, WI) was used by all four cell banks for STR analysis. Results were made available for eight core STR loci: D5S818, D13S317, D7S820, D16S539, vWA, TH01, TPOX and CSF1PO, with the addition of amelogenin as a ninth locus for gender determination. Electropherograms were analyzed and results were interpreted by each cell bank, in accordance with their own internal STR method validation and Standard Operating Procedures.

Data were made available anonymously. Donors were de-identified, with the exception of the HeLa cell line, where the donor's identity is in the public domain.

Grouping STR profiles into related cell line sets

To identify "related" STR profiles, that is, those coming from a common donor, entries were tagged by manual sorting of the combined dataset. Entries were sorted by:

1. Cell line designation. All STR profiles associated with the same cell line name, or with the same name followed by a different suffix, were tagged as possibly coming from the same donor. For example, HeLa was tagged along with HeLa 229, HeLa AG, HeLa TG and HeLa S3.
2. STR locus. STR profiles that differed at 0–2 loci were tagged as possibly coming from the same donor.

3. Comparison to a separate database of cross-contaminated or misidentified cell lines.[6] Version 6.7 of the database was used for comparison. Cell line designations in common were compared to the published contaminant, and the resulting misidentified cell lines were tagged as coming from the same donor.

The provenance of all tagged entries[14] was assessed, looking for further evidence of relatedness apart from the cell line designation or STR profile. Provenance was determined through the catalog entry for that cell line and a search of the scientific literature using PubMed. For example, HeLa 229, HeLa AG, HeLa TG and HeLa S3 were all documented as HeLa derivatives on the relevant cell bank websites and through published work.[35]

Tagged entries were grouped into related cell line sets. Each set represented all of the STR profiles arising from a single donor. A reference STR profile was selected from each set for further comparison. Normally a reference profile would come from donor tissue or the parental cell line at the lowest available passage. Because no tissue or passage information was available for our study, the reference profile was selected arbitrarily from the entries for the parental cell line.

Provenance information was then used to determine relationships between the reference profile and the other STR profiles in that set. STR profiles were grouped into same cell line (duplicate holdings of that cell line across several cell banks); legitimately related cell lines—either derivatives (daughter cell line) or other cell lines established from the same donor (sister cell line)—and known misidentified or cross-contaminated cell lines. In some cases, provenance could not be used to establish a relationship between similar STR profiles, and in those cases the entry was labeled as "Relationship Unknown."

Developing a validated subset of STR profiles

To look at the effectiveness of match criteria in distinguishing between unrelated cell lines, a validated subset of STR profiles was used. This subset was initiated as the nucleus for the NCBI database and consists of cell lines with known identities, whose STR profiles displayed minimal variation when tested by the cell banks contributing to the dataset.

STR profiles were included in this validated subset if two or more cell banks generated identical STR profiles for that cell line. Data were interpreted by the contributing cell bank. The cell lines in the subset were then compared to the database of cross-contaminated or misidentified cell lines,[6] using version 6.7 of the database for comparison. Any cell line with evidence for misidentification was excluded before the subset was finalized.

Results

A minimum of eight core STR loci is recommended when comparing STR profiles

The STR loci used by the cell banks may differ with respect to the specific loci used and the number of loci used, relative to those described in previous studies.[7,30] To recommend a minimum number of loci, the ASN-0002 Workgroup examined the number of unique STR profile results within the combined dataset. Relationships between cell line samples were not considered at this stage. For successful authentication, it is important to be able to discriminate between unique STR profiles, just as it is to be able to group related samples together.

Discrimination between unique STR profiles was assessed by progressively reducing the number of loci included in the analysis from nine (eight STR loci plus amelogenin) to one. All possible locus combinations were examined (including elimination of amelogenin), giving a total of 512 combinations across the nine loci evaluated. For each locus combination, the number of unique STR profiles resulting was calculated, expressed as a percentage of the combined dataset (n = 2,279). The highest and lowest percentages were recorded for each of the number of loci included in the analysis, giving a range of reductions.

The results are summarized in Figure 1. Nine loci (eight STR loci plus amelogenin) resulted in 77.2% unique STR profile results within the combined dataset. As might be expected, reduction of the number of loci evaluated resulted in loss of discriminating ability. Some loss of discrimination became evident with removal of one locus; unique STR profile results from eight loci ranged from 76.8 to 75.5%. Continued reduction of the number of loci evaluated resulted in a steady loss of discrimination, which was particularly marked when looking at the worst results from all locus combinations.

Because some reduction of discrimination occurred after exclusion of one locus, the recommendation was made that eight core STR loci (plus amelogenin) be used as a minimum when comparing cell line STR profiles. The core loci recommended were those included in this analysis: D5S818, D13S317, D7S820, D16S539, vWA, TH01, TPOX and CSF1PO.

Related cell line samples can be highly variable

To assess the effectiveness of the recommended STR loci for cell line authentication, the combined dataset was examined to identify related STR profiles, that is, those arising from the same donor. The process is described in the Methods section and summarized in Figure 2.

Relatedness was established for more than half of the combined dataset (n = 1,157, 50.8%). A reference profile was chosen for each set of related samples (n = 369). Of the remaining STR profiles, approximately one third were duplicate holdings of the same cell line (352/1,157, 30.4%); another third were legitimately related (330/1,157, 28.5%); and almost one in ten samples were known to be misidentified (91/1,157, 7.9%). Misidentification was documented in previous publications or was confirmed during discussion with the relevant cell bank. Misidentified cell lines were

Number Loci Used	% Unique profiles	
	Best Result	Worst Result
9	77.2	77.2
8	76.8	75.5
7	76.3	74.2
6	75.9	72.3
5	74.6	67.8
4	72.2	46.6
3	61.0	16.6
2	25.4	2.5
1	3.38	0.1

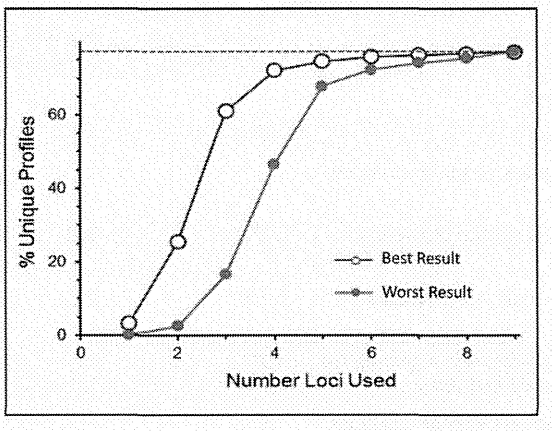

Figure 1. Discrimination between STR profiles based on the number of loci used. The combined dataset (n = 2279) comprised results from eight STR loci (D5S818, D13S317, D7S820, D16S539, vWA, TH01, TPOX and CSF1PO), with the addition of amelogenin for gender determination. To assess discrimination, the number of unique STR profiles existing within the combined dataset with these nine loci was calculated and expressed as a percentage of the total number of STR profiles in the combined dataset. The number of loci was then progressively reduced from nine to one, and the number of unique STR profiles was calculated for each possible locus combination (512 combinations). The "best result" (highest) and "worst result" (lowest) are recorded for each locus number.

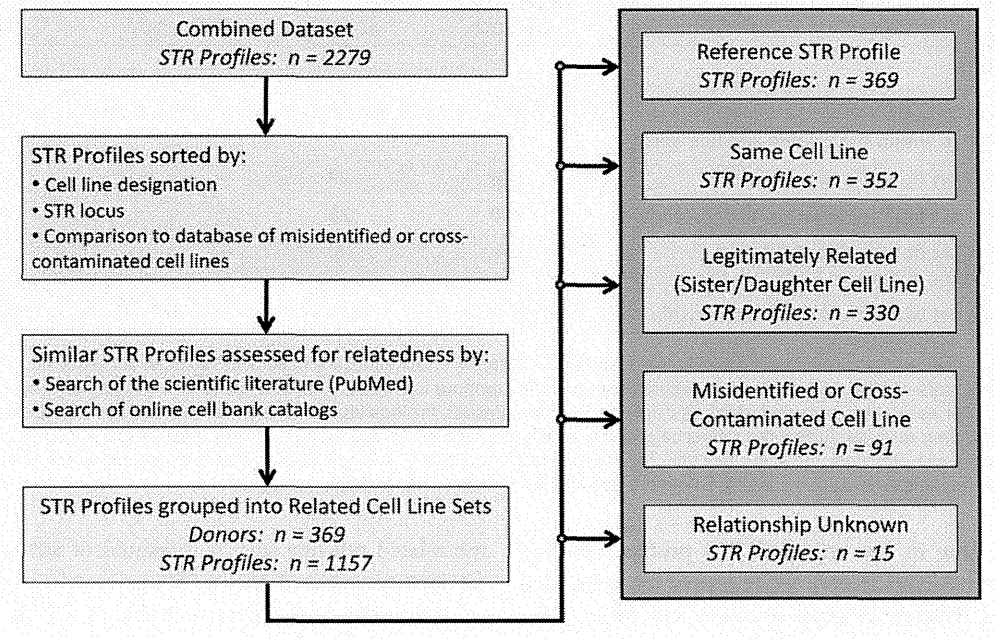

Figure 2. Grouping STR profiles into related cell line sets. The combined dataset of STR profiles was sorted into related cell line sets using the process shown. The process is described further in the Methods section.

clearly labeled as such in the cell bank catalog or, in other cases, were restricted from distribution to the public.

In a small number of cases (15/1,157, 1.3%), a search of the cell bank catalogs and the literature did not bring up evidence that cell lines were legitimately related. Of these STR profiles, 7/15 were identical to another STR profile in the dataset; all fulfilled the match criteria described in the following section. Because these samples made up only 1% of the dataset and otherwise fulfilled the match criteria, they were included in data analysis along with the samples for which relatedness could be clearly documented.

The sets of related cell lines were then examined to see how they differed when compared to the reference profile. A total of 369 sets were examined, locus by locus, looking at the maximal locus differences when each profile was compared to the reference profile. The results are shown in Table 1. More than half of the related cell line sets consisted of identical STR profiles (189/369, 51.2%). Other sets showed

Table 1. Related cell line sets: locus differences

Sample or set under consideration	All loci same (0 different)	8/9 loci same (1 locus different)	7/9 loci same (2 loci different)	≤6/9 loci same (≥3 loci different)	Total
Related cell line sets					
Total	189	93	50	37	369
Individual STR profiles within the related cell line sets					
STR profiles in each set from the same cell line	153	101	52	46	352
STR profiles in each set from legitimately related cell lines	107	90	56	77	330
STR profiles in each set from misidentified or cross-contaminated cell lines	20	13	14	44	91
STR profiles in each set with an unknown relationship	5	2	5	3	15

Number of locus differences seen when STR profiles are grouped into related cell line sets. For all of the related cell line sets (n = 369), profiles were compared to a single reference profile, and the maximum number of locus differences arising from that comparison was recorded. Results for individual STR profiles from the combined dataset (n = 1,157) are shown in the remaining rows, excluding the reference profiles, which are used for comparison (n = 788). Results are sorted by their relationship to the reference profile—samples from the same cell line, legitimately related samples, misidentified or cross-contaminated cell lines, and those with an unknown relationship.

Table 2. Related cell line sets: match criteria

Sample or set under consideration	Lowest match 90–100%	Lowest match 80–89%	Lowest match 70–79%	Lowest match <70%	Total
Related cell line sets					
Total	327	35	4	3	369
Individual STR profiles within the related cell line sets					
STR profiles in each set from the same cell line	324	22	4	2	352
STR profiles in each set from legitimately related cell lines	300	28	1	1	330
STR profiles in each set from misidentified or cross-contaminated cell lines	84	5	1	1	91
STR profiles in each set with an unknown relationship	11	4	0	0	15

Effectiveness of the match criteria recommended within the Standard to determine relatedness for cell line samples. For all of the related cell line sets (n = 369), the Masters algorithm (see Methods section) was used to compare individual sample profiles to a designated reference profile. The lowest percent match for each set is recorded in the first row of the table. Results for individual STR profiles from the combined dataset (n = 1157) are shown in the remaining rows, excluding the reference profiles, which are used for comparison (n = 788). Results are sorted by their relationship to the reference profile—samples from the same cell line, legitimately related samples, misidentified or cross-contaminated cell lines, and those with an unknown relationship. The columns falling below the 80% match threshold are highlighted in grey.

more variability. One in four included STR profiles differing at only one locus (93/369, 25.2%), while one in ten contained STR profiles that differed at three or more loci when compared to the reference profile (37/369, 10.0%). The latter group included a high proportion of misidentified samples (44/91 misidentified profiles), making this an important group of samples to authenticate correctly.

A simple match algorithm with an 80% threshold shows relatedness for 98% of cell lines

A match algorithm is often used to compare STR profiles from cell line samples, allowing for some degree of genetic drift. We assessed the effectiveness of this approach by applying a commonly used and simple match algorithm, the Masters algorithm (see Methods section), to the STR profiles in the related cell line sets. A threshold of 80% match was used for interpretation of results.[7,30]

The results are shown in Table 2. Of all the related cell line sets identified from the combined dataset, 98% demonstrated percent match between the reference profile and all sample profiles at 80% or above (362/369, 98.1%). Four sets included samples displaying percent match in the range 70–79%, and three had samples displaying <70% match (total displaying <80% match in profiles: 7/369, 1.9%). The seven cell line sets that failed to meet the 80% match threshold using the Masters algorithm are listed in Supporting Information Table 1.

Samples within each set were not equally variable with respect to STR profile match. The seven cell line sets failing to meet the 80% match threshold with the Masters algorithm included 35 STR profiles; of those profiles, ten fell below the

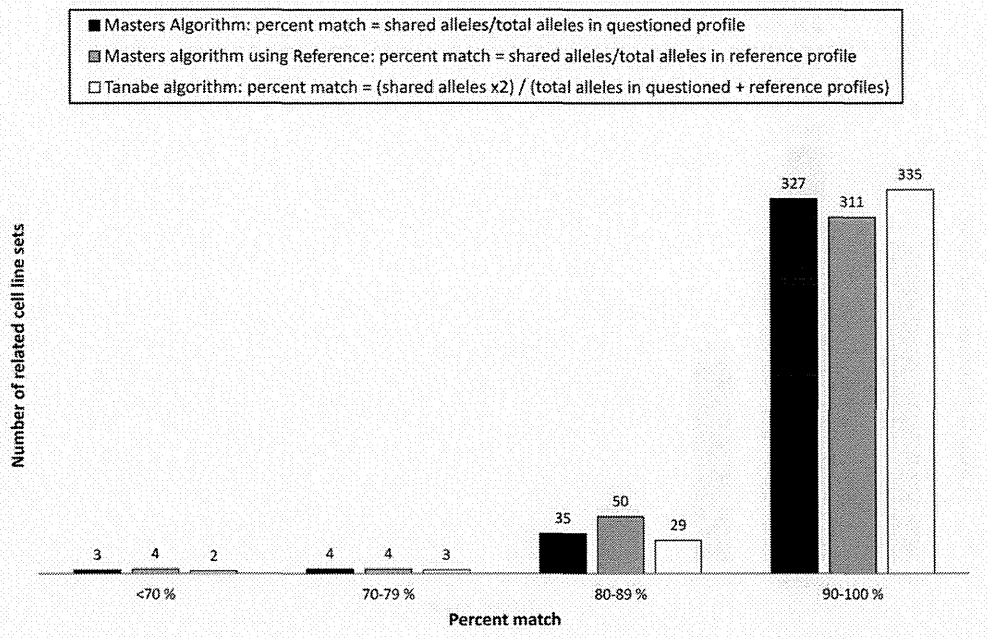

Figure 3. Comparison of match algorithms. Percent match results obtained when comparing related cell line sets using different algorithms. Samples are compared using the Masters algorithm, using the total number of alleles from the questioned profile in the calculation; the Masters algorithm, using the total number of alleles from the reference profile in the calculation; and the Tanabe algorithm. The lowest percent match result for each set is recorded here.

80% match threshold. Those ten samples originated from all of the cell banks contributing data (Supporting Information Table 1). Six samples came from duplicate holdings of the same cell line, two from legitimately related cell lines and two from misidentified cell lines (Table 2), suggesting that a more distant relationship between cultures was not the cause of the variability seen in these STR profiles.

Both algorithms tested are effective to show relatedness

Although the Masters algorithm is simple to apply, it can produce different numerical results depending on whether the calculation is based on the number of total alleles in the questioned or reference profiles. Because of this variation, some laboratories prefer to use the Tanabe algorithm (see Methods section), which uses the numbers of alleles from both questioned and reference profiles in the calculation. To compare the different approaches, percent match was calculated across all related samples using two additional algorithms. The Masters algorithm was re-applied but using the total number of alleles in the reference profile rather than in the questioned profile. The last algorithm tested was the Tanabe algorithm. Percent match results for each of the three algorithms across the related cell line sets are compared in Figure 3.

The Masters algorithm resulted in some subtle changes when the calculations were based on the numbers of alleles in the reference profile instead of in the questioned profile—for example, there were fewer results at $\geq$90% match (Fig. 3). However, almost 98% of sets gave a percent match of 80% or above (361/369, 97.8%). Eight sets failed to meet the 80% match threshold (8/369, 2.2%), four at 70–79% and four at <70% match. In contrast, the Tanabe algorithm gave a larger number of results at $\geq$90% match but its performance relative to the 80% threshold was similar (Fig. 3). Nearly 99% of related cell line sets gave a percent match of 80% or above (364/369, 98.6%) using the Tanabe algorithm; five sets failed to meet the 80% match threshold (5/369, 1.4%), three at 70–79% and two at <70% match.

Whichever algorithm was used, close to 98% of cell line samples were correctly authenticated. However, the profiles for the remaining 2% of samples were more challenging to interpret correctly. Samples from the related cell line sets that failed to meet the 80% match threshold with either algorithm are listed in Supporting Information Table 1. Cell lines within these sets have previously been documented as having highly variable STR profiles,[7,16,28] and many have known microsatellite instability—for example, the leukemia and lymphoma cell lines CCRF-CEM, Jurkat, MOLT-3, MOLT-4, NALM-6[36] and KCL-22[37] and the breast carcinoma cell line MT-3.[38]

Results for all known misidentified cell lines in the dataset (n = 91) are listed in Supporting Information Table 2. All but two misidentified cell lines gave percent match results at $\geq$80% when the contaminant was compared to the reference profile (n = 89), including all HeLa-contaminated samples (n = 32). The two exceptions were RTSG and RMUG-L, both contaminated with SNG-II. Both STR profiles showed more than two allele peaks across at least four loci (Supporting Information Table 2), suggesting that these samples were mixtures at the time of testing.

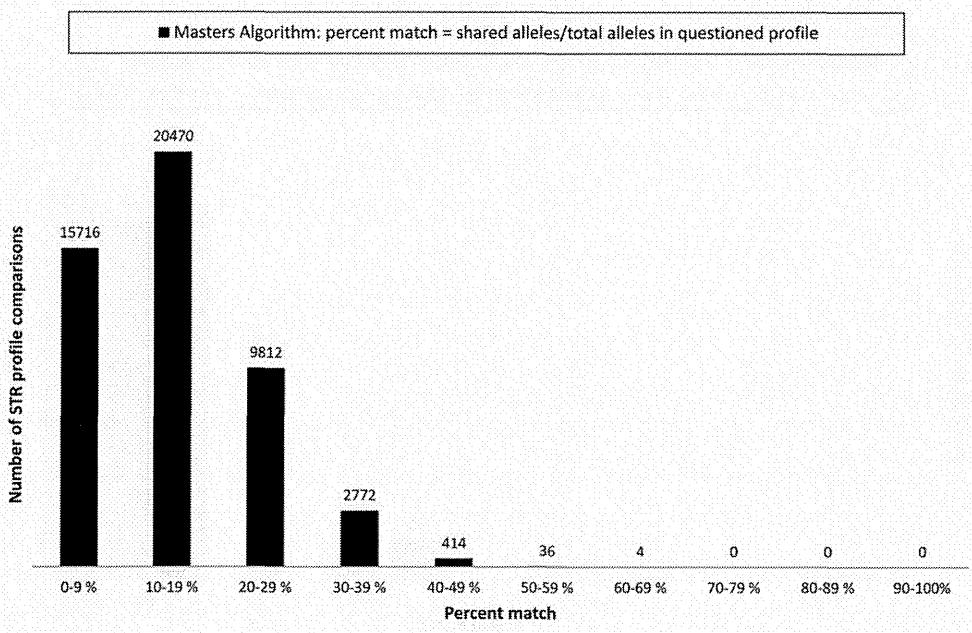

Figure 4. Comparison of unrelated cell lines. Percent match results obtained when comparing unrelated cell line samples. A subset of 223 validated samples, with all related samples removed, was used. Each sample was compared to all of the others in the subset, resulting in a total of 49,224 unrelated results.

Testing on a subset of unrelated STR profiles shows that an 80% threshold discriminates effectively between unrelated cell lines

Having examined the effectiveness of an 80% match threshold when detecting related samples, it was important to assess its ability to discriminate between unrelated samples. To evaluate this, we used a validated subset of the combined dataset containing 223 STR profiles. The combined dataset is likely to contain unrecognized duplicates even after known provenance has been determined, making such a subset necessary for analysis of unrelated samples.

Percent match was calculated for all of the STR profiles in this 223 sample subset, comparing each STR profile to the others in the dataset—a total of 49,506 comparisons. Of these comparisons, 282 were found to involve related STR profiles (to same or sister/daughter cell lines) and were excluded from the analysis, leaving 49,224 unrelated comparisons. The results are summarized in Figure 4.

All percent match results for unrelated cell lines fell below the 80% match threshold. Four were found in the 60–69% match range and 36 in the 50–59% range; the remaining comparisons (n = 49184) were each below 50%. A clear separation can thus be seen between related samples, at $\geq$80% match (Fig. 3), and unrelated samples at <50% match (Fig. 4). Note that an overlap can also be seen in the 50–79% range for a small minority of samples in this dataset.

Discussion

STR profiling is an accepted and reliable method for matching different samples derived from the same human donor.[9] However, a tendency toward genetic drift in malignant samples and cultured cells means that clear guidelines are needed for comparison of STR profiles from cell line samples. Researchers using STR profiling for human cell line authentication need to know the minimum number of STR loci that should be used, and how to interpret results to correctly conclude whether samples come from the same donor. Absence of such guidelines mean that laboratories may easily draw incorrect conclusions from authentication data—for example, saying that differences in STR profiles are due to "genomic instability" when cross-contamination has actually occurred. One such example has occurred recently in relation to the T1 neural stem cell line, showing the need for clear recommendations in this area.[39,40]

STR profiling has been performed on cell lines over more than 10 years and so, inevitably, there are variations in the number of loci and in the specific loci comprising the published profiles. One recent study has questioned the use of only eight STR loci. The authors presented data and performed simulations to show that reliable discrimination between glioma cell lines does not always occur when evaluating eight STR loci plus amelogenin, and that the duplicates present within many databases make interpretation more difficult.[31] In our study, we used the duplicates present within the cell bank holdings and applied known provenance information to generate sets of related cell lines that were then used to analyze these eight STR loci in practice.

Our results are consistent with previous studies showing a clear separation in STR profiles between related and unrelated cell line samples.[7,16] When using the Masters algorithm, more than 98% of related cell line sets (362/369) gave percent

match results at 80% or above; more than 99% of unrelated comparisons (49,184/49,224) gave percent match results below 50%. These results were obtained using eight STR loci plus amelogenin and involved a large dataset comprising many of the common cancer cell lines, including those most likely to cause cross-contamination.[6,13]

Our results also agree with other studies showing that in a small number of cases, determining if two samples are related using STR profiling alone is difficult.[28,31] Using the Masters algorithm, 7/369 of related cell line sets gave at least one result at <80% match; 40/49,224 of unrelated comparisons gave results at 50–69% match. We thus have a documented overlap at 50–79% match between related and unrelated samples. Although discrimination due to insufficient loci may be responsible, a more likely cause is variation in primer choices, test methods and interpretation of results between the cell banks and over time. STR profiles were obtained retrospectively as part of authentication testing of cell bank holdings over a prolonged period of time (>10 years). A previous study performing STR profiling of 253 cell lines within a single centre did not demonstrate a similar overlap,[7] suggesting that laboratory differences are contributory in our study. A prospective study from a single cell bank or testing organization would help to address this possibility, for example, arising from the NIST Identification of Human Cell Lines Project.[20] If such variations are confirmed to contribute to STR profile variability, validation guidelines for all laboratories contributing to the NCBI database of human cell line STR profiles would be essential. Quality criteria for STR profiles have already been developed as part of the human cell line authentication Standard.[12]

It should be emphasized that eight core STR loci plus amelogenin are recommended as a minimum number for effective authentication; adding more loci will further increase the discriminatory power of the technique. The discriminatory power of STR analysis depends on the biological nature of each of the loci included, the number of loci and the population group.[12] For the eight core STR loci recommended here, the probability of a random match has been estimated at 2.71×10^{-8} for African American and 1.14×10^{-8} for US Caucasian populations.[41] The power of discrimination improves by approximately one order of magnitude for each STR locus added[12]; for example, the probability of a random match becomes 5.9×10^{-18} when examining the 15 STR loci of the PowerPlex 16 kit in the US Caucasian population.[42] Increasing the number of STR loci used will not only give greater clarity when comparing cell line samples but also increases the risk that the donor may be unambiguously identified if results are compared to other databases online.

Our analysis of a combined dataset containing the cell lines most frequently associated with cross-contamination[6] shows that in most cases, eight core STR loci plus amelogenin provide sufficient resolving power for effective authentication. A clear separation is seen here between related and unrelated samples even where one or more loci show variation from the reference sample. However, additional loci may be of benefit for a small number of problematic samples. Problematic samples include those related cell line sets with percent match results in the range 50–79% ($n = 7$), or those samples at >80% match with no known relationship ($n = 15$).

Any percent match result that does occur in the range 50–79% is best handled by further testing, using additional loci or an alternate test method.[31] Six of the sets here that failed to meet the 80% match threshold included cell lines with known microsatellite instability.[36–38] STR loci are known to vary in the presence of microsatellite instability, with the degree of variability depending on the locus being assessed.[16,28,43] A greater number of loci may give sufficient data for discrimination, but for cell lines with microsatellite instability, an alternate method such as single nucleotide polymorphism (SNP) analysis should also be considered.[44]

Previous information on the cell line's provenance should be taken into consideration when performing authentication testing. In some cases, what initially appears to be a case of cross-contamination may prove to be quite otherwise when the history of the cell line is known—for example, showing that a cell line exists under two different names, as with Alexander and PLC/PRF/5.[45] Cell lines from different donors but carrying the same name coincidentally may also cause confusion. Where similarly named cell lines exist, it is advisable for the laboratory that established the later cell line to alter its designation so that each cell line is uniquely identified. Where cell lines are held by a repository, supplying the catalog number within the Methods section will allow cross-referencing with provenance information collected from the depositor by the cell bank when that cell line was first deposited.

STR profiling will not discriminate between cell lines established from the same donor. Cell lines from different tissues may be distinguishable using phenotypic markers, although it should be noted that phenotypic markers can be affected by the degree of differentiation shown by each cell line and their length of time in culture.[1,3] To effectively distinguish between cell lines from the same donor, it may be necessary to use microarray analysis or next-generation sequencing to perform in depth analysis of gene expression patterns, underlying mutations and copy number alterations.[46,47]

The ultimate aim of authentication testing is to ensure that cell lines are not misidentified, but rather that they continue to correspond to the individual who first agreed to donate their cells or tissue for research. We have a responsibility to ensure that the donor's gift is used effectively and guarded from misuse. Authentication testing of human cell lines through STR profiling offers an excellent solution, but as with all test methods, results must be interpreted carefully and considered in context. Clear guidelines for authentication testing, documentation of cell line provenance and ongoing validation will help ensure that human cell lines are effective and representative models for biomedical research.

Acknowledgements

All members of the ASN-0002 Standards Workgroup, including Steven R. Bauer, John M. Butler and Zenobia F. Taraporewala, are

acknowledged for their contributions to the Standard and this manuscript. David Duewer (NIST) is gratefully acknowledged for his contributions to the data analysis. The authors are grateful to R. Ian Freshney for his role in developing the list of cross-contaminated and misidentified cell lines. Finally, they thank George Theodosopoulos of CellBank Australia and all the scientists of the contributing cell banks, for their hard work and expertise in performing the STR profiling results comprising the dataset analyzed.

References

1. Masters JR.Human cancer cell lines: fact and fantasy. *Nat Rev Mol Cell Biol* 2000;1:233–6.
2. Reid YA.Characterization and authentication of cancer cell lines: an overview. *Methods Mol Biol* 2011;731:35–43.
3. Hughes P, Marshall D, Reid Y, et al. The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? *Biotechniques* 2007;43:575–84.
4. Stacey GN.Cell culture contamination. *Methods Mol Biol* 2011;731:79–91.
5. MacLeod RA, Dirks WG, Matsuo Y, et al. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer* 1999;83:555–63.
6. Capes-Davis A, Theodosopoulos G, Atkin I, et al. Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int J Cancer* 2010;127: 1–8.
7. Masters JR, Thomson JA, Daly-Burns B, et al. Short tandem repeat profiling provides an international reference standard for human cell lines. *Proc Natl Acad Sci USA* 2001;98:8012–7.
8. Barallon R, Bauer SR, Butler J, et al. Recommendation of short tandem repeat profiling for authenticating human cell lines, stem cells, and tissues. *In Vitro Cell Dev Biol Anim* 2010;46:727–32.
9. Butler JM.Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 2006;51:253–65.
10. Dirks WG, Drexler HG.Online verification of human cell line identity by STR DNA typing. *Methods Mol Biol* 2011;731:45–55.
11. American Type Culture Collection Standards Development Organization Workgroup ASN-0002.Cell line misidentification: the beginning of the end. *Nat Rev Cancer* 2010;10:441–8.
12. ANSI/ATCC ASN-0002–2011. Authentication of Human Cell Lines: Standardization of STR Profiling. ANSI eStandards Store, 2012.
13. Dirks WG, MacLeod RA, Nakamura Y, et al. Cell line cross-contamination initiative: an interactive reference database of STR profiles covering common cancer cell lines. *Int J Cancer* 2010;126: 303–4.
14. Freshney RI.Cell line provenance. *Cytotechnology* 2002;39:55–67.
15. Schweppe RE, Klopper JP, Korch C, et al. Deoxyribonucleic acid profiling analysis of 40 human thyroid cancer cell lines reveals cross-contamination resulting in cell line redundancy and misidentification. *J Clin Endocrinol Metab* 2008;93:4331–41.
16. Lorenzi PL, Reinhold WC, Varma S, et al. DNA fingerprinting of the NCI-60 cell line panel. *Mol Cancer Ther* 2009;8:713–24.
17. Brenner JC, Graham MP, Kumar B, et al. Genotyping of 73 UM-SCC head and neck squamous cell carcinoma cell lines. *Head Neck* 2010;32:417–26.
18. Zhao M, Sano D, Pickering CR, et al. Assembly and initial characterization of a panel of 85 genomically validated cell lines from diverse head and neck tumor sites. *Clin Cancer Res* 2011;17: 7248–64.
19. Korch C, Spillman MA, Jackson TA, et al. DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. *Gynecol Oncol* 2012;127:241–8.
20. National Institute of Standards and Technology.Identification of Human Cell Lines Project. *Fed Regist* 2012;77:5489–91.
21. O'Donoghue LE, Rivest JP, Duval DL.Polymerase chain reaction-based species verification and microsatellite analysis for canine cell line validation. *J Vet Diagn Invest* 2011;23: 780–5.
22. Almeida JL, Hill CR, Cole KD.Authentication of African green monkey cell lines using human short tandem repeat markers. *BMC Biotechnol* 2011;11:102.
23. Poetsch M, Petersmann A, Woenckhaus C, et al. Evaluation of allelic alterations in short tandem repeats in different kinds of solid tumors–possible pitfalls in forensic casework. *Forensic Sci Int* 2004;145:1–6.
24. Vauhkonen H, Hedman M, Vauhkonen M, et al. Evaluation of gastrointestinal cancer tissues as a source of genetic information for forensic investigations by using STRs. *Forensic Sci Int* 2004;139:159–67.
25. MacLeod RA, Dirks WG, Reid YA, et al. Identity of original and late passage Dami megakaryocytes with HEL erythroleukemia cells shown by combined cytogenetics and DNA fingerprinting. *Leukemia* 1997;11:2032–8.
26. Chiong E, Dadbin A, Harris LD, et al. The use of short tandem repeat profiling to characterize human bladder cancer cell lines. *J Urol* 2009;181: 2737–48.
27. Hane B, Tummler M, Jager K, et al. Differences in DNA fingerprints of continuous leukemia-lymphoma cell lines from different sources. *Leukemia* 1992;6:1129–33.
28. Parson W, Kirchebner R, Muhlmann R, et al. Cancer cell line identification by short tandem repeat profiling: power and limitations. *FASEB J* 2005;19:434–6.
29. Drexler HG, Dirks WG, Matsuo Y, et al. False leukemia-lymphoma cell lines: an update on over 500 cell lines. *Leukemia* 2003;17:416–26.
30. Tanabe H, Takada Y, Minegishi D, et al. Cell line individualization by STR multiplex system in the cell bank found cross-contamination between ECV304 and EJ-1/T24. *Tissue Culture Res Commun* 1999;18:329–38.
31. Bady P, Diserens AC, Castella V, et al. DNA fingerprinting of glioma cell lines and considerations on similarity measurements. *Neuro Oncol* 2012;14:701–11.
32. Eltonsy N, Gabisi V, Li X, et al. Detection algorithm for the validation of human cell lines. *Int J Cancer* 2012;131:E1024–E1030.
33. Lynch M.The similarity index and DNA fingerprinting. *Mol Biol Evol* 1990;7:478–84.
34. Sorensen T.A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol Skr* 1948;5:1–34.
35. Kataoka E, Honma M, Ohnishi K, et al. Application of highly polymorphic DNA markers to the identification of HeLa cell sublines. *In Vitro Cell Dev Biol* 1992;28A: 553–6.
36. Inoue K, Kohno T, Takakura S, et al. Frequent microsatellite instability and BAX mutations in T cell acute lymphoblastic leukemia cell lines. *Leuk Res* 2000;24:255–62.
37. Takeuchi S, Takeuchi N, Fermin AC, et al. Frameshift mutations in caspase-5 and other target genes in leukemia and lymphoma cell lines having microsatellite instability. *Leuk Res* 2003;27: 359–61.
38. Seitz S, Wassmuth P, Plaschke J, et al. Identification of microsatellite instability and mismatch repair gene mutations in breast cancer cell lines. *Genes Chromosomes Cancer* 2003;37: 29–35.
39. Wu W, He Q, Li X, et al. Long-term cultured human neural stem cells undergo spontaneous transformation to tumor-initiating cells. *Int J Biol Sci* 2011;7:892–901.
40. Torsvik A, Rosland GV, Bjerkvig R.Spontaneous transformation of stem cells in vitro and the issue of cross-contamination. *Int J Biol Sci* 2012;8: 1051–2.
41. Micka KA, Amiott EA, Hockenberry TL, et al. TWGDAM validation of a nine-locus and a four-locus fluorescent STR multiplex system. *J Forensic Sci* 1999;44:1243–57.
42. Hill CR, Duewer DL, Kline MC, et al. Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex (R) ESX 17 and ESI 17 Systems. *Forensic Sci Int Genet* 2011;5: 269–75.
43. Bacher JW, Flanagan LA, Smalley RL, et al. Development of a fluorescent multiplex assay for detection of MSI-High tumors. *Dis Markers* 2004; 20:237–50.
44. Castro F, Dirks WG, Fahnrich S, et al. High-throughput SNP-based authentication of human cell lines. *Int J Cancer*, Epub ahead of print.
45. Daemer RJ, Feinstone SM, Alexander JJ, et al. PLC/PRF/5 (Alexander) hepatoma cell line: further characterization and studies of infectivity. *Infect Immun* 1980;30:607–11.
46. Kao J, Salari K, Bocanegra M, et al. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* 2009;4: e6146.
47. Nichols AC, Yoo J, Palma DA, et al. Frequent Mutations in TP53 and CDKN2A Found by Next-Generation Sequencing of Head and Neck Cancer Cell Lines. *Arch Otolaryngol Head Neck Surg* 2012;138:732–9.

PLOS | ONE

Protein Kinase C Regulates Human Pluripotent Stem Cell Self-Renewal

Masaki Kinehara[1], Suguru Kawamura[1], Daiki Tateyama[1], Mika Suga[1], Hiroko Matsumura[1], Sumiyo Mimura[1], Noriko Hirayama[2], Mitsuhi Hirata[1], Kozue Uchio-Yamada[3], Arihiro Kohara[2], Kana Yanagihara[1], Miho K. Furue[1]*

1 Laboratory of Stem Cell Cultures, Department of Disease Bioresources Research, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan, **2** Laboratory of Cell Cultures, Department of Disease Bioresources Research, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan, **3** Laboratory of Animal Models for Human Diseases, Department of Disease Bioresources Research, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan

Abstract

Background: The self-renewal of human pluripotent stem (hPS) cells including embryonic stem and induced pluripotent stem cells have been reported to be supported by various signal pathways. Among them, fibroblast growth factor-2 (FGF-2) appears indispensable to maintain self-renewal of hPS cells. However, downstream signaling of FGF-2 has not yet been clearly understood in hPS cells.

Methodology/Principal Findings: In this study, we screened a kinase inhibitor library using a high-throughput alkaline phosphatase (ALP) activity-based assay in a minimal growth factor-defined medium to understand FGF-2-related molecular mechanisms regulating self-renewal of hPS cells. We found that in the presence of FGF-2, an inhibitor of protein kinase C (PKC), GF109203X (GFX), increased ALP activity. GFX inhibited FGF-2-induced phosphorylation of glycogen synthase kinase-3β (GSK-3β), suggesting that FGF-2 induced PKC and then PKC inhibited the activity of GSK-3β. Addition of activin A increased phosphorylation of GSK-3β and extracellular signal-regulated kinase-1/2 (ERK-1/2) synergistically with FGF-2 whereas activin A alone did not. GFX negated differentiation of hPS cells induced by the PKC activator, phorbol 12-myristate 13-acetate whereas Gö6976, a selective inhibitor of PKCα, β, and γ isoforms could not counteract the effect of PMA. Intriguingly, functional gene analysis by RNA interference revealed that the phosphorylation of GSK-3β was reduced by siRNA of PKCδ, PKCε, and ζ, the phosphorylation of ERK-1/2 was reduced by siRNA of PKCε and ζ, and the phosphorylation of AKT was reduced by PKCε in hPS cells.

Conclusions/Significance: Our study suggested complicated cross-talk in hPS cells that FGF-2 induced the phosphorylation of phosphatidylinositol-3 kinase (PI3K)/AKT, mitogen-activated protein kinase/ERK-1/2 kinase (MEK), PKC/ERK-1/2 kinase, and PKC/GSK-3β. Addition of GFX with a MEK inhibitor, U0126, in the presence of FGF-2 and activin A provided a long-term stable undifferentiated state of hPS cells even though hPS cells were dissociated into single cells for passage. This study untangles the cross-talk between molecular mechanisms regulating self-renewal and differentiation of hPS cells.

Citation: Kinehara M, Kawamura S, Tateyama D, Suga M, Matsumura H, et al. (2013) Protein Kinase C Regulates Human Pluripotent Stem Cell Self-Renewal. PLoS ONE 8(1): e54122. doi:10.1371/journal.pone.0054122

Editor: Tadayuki Akagi, Kanazawa University, Japan

Received April 20, 2012; **Accepted** December 10, 2012; **Published** January 21, 2013

Funding: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This study was supported by grants-in-aid from the Ministry of Health, Labor and Welfare of Japan to M.K.F. and A.K., the Ministry of Education, Culture, Sports, Science and Technology of Japan to M.K.F. and M.K. and the Japan Science and Technology Agency to M.K.F.

Competing Interests: The authors have read the journal's policy and have the following conflicts: One of the authors, (MKF) has declared a financial interest in a company, Cell Science & Technology Institute Corporation (Sendai, Japan) whose product, a basal medium ESF was used in this study. However, the licensing fee is less than $10,000 per year. This does not alter the authors adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: mkfurue@nibio.go.jp

Introduction

The self-renewal of human pluripotent stem (hPS) cells including embryonic stem (hES) and induced pluripotent stem (hiPS) cells have been reported to be supported by various signal pathways, including transforming growth factor-β/activin A/Nodal [1–3], sphingosine-1-phosphate/platelet derived growth factor (S1P/PDGF) [4], insulin growth factor (IGF)/insulin [5] and fibroblast growth factor-2 (FGF-2) [6–9]. The process of self-renewal appears to be regulated synergistically through the various pathways via growth factor or cytokine supplementation. Among them, FGF-2 signaling appears indispensable to hPS cells [10–12].

FGF family members including FGF-2, bind to FGF receptors (FGFRs) and induce activation of the mitogen-activated protein kinase/extracellular signal-regulated kinase-1/2 (ERK-1/2) kinase (MEK), phosphatidylinositol-3 kinase (PI3K), and phospholipase C-γ (PLC-γ)/protein kinase C (PKC) pathways [13]. MEK-1/2 activation by FGFR results in ERK-1/2 phosphorylation, which subsequently translocates into the nucleus leading to phosphorylation of transcription factors such as c-Myc, c-Jun, and c-Fos. PI3K, a lipid kinase activates pleckstrin homology (PH) domain containing proteins such as AKT, and 3-phosphoinositide-dependent kinase-1 (PDK1). AKT directly activates murine double minute 2 (MDM2), a negative regulator of p53. p53 is

responsible for DNA damage surveillance and in response initiates cell cycle arrest and DNA repair. Interestingly, AKT also inhibits glycogen synthase kinase-3 (GSK-3), a negative regulator of Wnt signaling by phosphorylation [14]. However, the contributions of FGF-2 downstream pathways in the self-renewal of hPS cells have been controversial [9,14–18]. The ERK pathway has been thought to promote cell proliferation and adhesion but also differentiation in hES cells. The PI3K pathway plays important roles in proliferation, differentiation, survival, and cellular transformation.

Previously, we found that a proteoglycan, heparin promotes FGF-2 activity on the growth of undifferentiated hES cells in a minimal growth factor-defined culture medium, hESF9 [8], in which the effect of exogenous factors can be analyzed without the confounding influences of undefined components [8,19–23] because insulin, transferrin, albumin conjugated with oleic acid, and FGF-2 (10 ng/ml) are the only protein components. Understanding cell signaling in undifferentiated hPS cells has lead to the development of optimal conditions for culturing hPS cells. However, manipulation of hPS cells still remains difficult because hPS cells as a single cell are unstable of self-renewal. Although Rho-associated kinase (ROCK) inhibitor (Y-27632) is quite effective to markedly diminish dissociation-induced apoptosis of single cells of hPS cells [24], the continuous use of the ROCK inhibitor increases differentiated cells [25]. For developing application using hPS cells, such as cell based therapy or toxicity screening tests, handling cell numbers would be beneficial. Even for basic research, handling cell numbers would be useful when the cells are dissociated for passages or differentiation. Presumably, if the culture conditions were able to fully support undifferentiated state, even single cells might maintain undifferentiated state. We suspected that there were unrevealed mechanisms to maintain undifferentiated state of single hPS cells. To further understand FGF-2 related molecular mechanisms regulating self-renewal would enhance understanding unclarified cell signaling in hPS cells. Therefore, we screened a kinase inhibitor library using a high-throughput alkaline phosphatase (ALP) activity-based assay in a minimal growth factor-defined culture medium, hESF9. We found that in the presence of FGF-2, an inhibitor of PKCs, GF109203X (GFX), increased ALP activity, suggesting that PKC reduces self-renewal of hPS cells. GFX inhibited FGF-2-induced GSK-3β phosphorylation. Addition of activin A increased phosphorylation of GSK-3β and ERK-1/2 synergistically with FGF-2 whereas activin A alone did not induce phosphorylation of GSK-3β. GFX negated differentiation of hPS cells induced by a PKC activator, phorbol 12-myristate 13-acetate (PMA) whereas Gö6976, a selective inhibitor of PKCα, β, and γ isoforms did not counteract the effect of PMA. Functional gene analysis by RNA interference revealed that siRNA of PKCδ, ε, and ζ isoforms decreased phosphorylation of GSK-3β and also siRNA of PKCε and ζ isoforms decreased phosphorylation of ERK-1/2 in hPS cells. siRNA of PKCε decreased phosphorylation of AKT. On the basis of these results, we suggest that PKCδ, ε and ζ isoforms are FGF-2 downstream effectors, and they play various roles in regulating hPS cell self-renewal. This study helps to untangle the cross-talk between molecular mechanisms regulating self-renewal and differentiation of hPS cells.

Results

PKC inhibitor increased ALP activity of hiPS cells

Previously, we detected the cell proliferative effect of heparin on hES cells without feeder cells in a minimal growth factor-defined culture medium, hESF9 [8], in which the effect of exogenous factors can be analyzed without the confounding influences of undefined components [8,19–23]. In this culture condition using hESF9 medium (Table S1) on bovine fibronectin (FN), a high-throughput ALP activity-based assay was performed to evaluate a library of chemical kinase inhibitors to understand FGF-2 related molecular mechanisms regulating self-renewal of hPS cells. Nine compounds were found to increase ALP activity of the hiPS cell line 201B7 [26] (Fig. 1): Kenpaullone, which is a substitute for a reprogramming factor KLF-4 in mouse iPS cells [27]; Y-27632, which is a Rho-kinase (ROCK) inhibitor known to enhance hES cells survival [24]; HA-1004, H-89, and HA-1077, which are kinase inhibitors presumed to target ROCK [28]; GF109203X (GFX) [29], which is a inhibitor for PKC isoforms; and H-7, H-8, and H-9, which are also thought to target PKC [30]. These results suggest that FGF-2 induces PKC, and PKC acts downstream of FGF-2 to regulate self-renewal of hPS cells.

Effect of PKC inhibitor on FGF-2 signaling in hPS cells

To examine how GFX influenced FGF-2 signaling in hPS cells, the phosphorylation of AKT, ERK-1/2, and GSK-3β induced by FGF-2 with GFX was confirmed by western blotting analysis (Fig. S1A, S1B, S1C, S1D). Then, the phosphorylation levels were quantified by AlphaScreen® SureFire® assay kit. Human ES cells H9 [31] after starvation of FGF-2 and insulin were treated with FGF-2 with and without GFX. FGF-2 significantly stimulated the phosphorylation of AKT, ERK-1/2, and GSK-3β in H9 cells in 15 minutes (Fig. 2A, 2B, 2C) as described previously [16,32]. Addition of GFX at 5.0 μM in the presence of FGF-2 significantly increased AKT phosphorylation in 15 minutes compared with addition of FGF-2 alone (Fig. 2A, 2B, Fig. S1E). The level of ERK-1/2 phosphorylation induced by FGF-2 with GFX was comparable with that without GFX (Fig. 2A). On the other hand, FGF-2-induced GSK-3β phosphorylation was completely inhibited by GFX (Fig. 2A, 2B) at concentrations higher than 1 μM treatment (Fig. S1E).

Addition of the PI3K inhibitor LY-294002 with FGF-2 completely inhibited AKT phosphorylation and significantly reduced GSK-3β phosphorylation (Fig. 2B, Fig. S1B). Addition of the MEK inhibitor U0126 with FGF-2 reduced ERK-1/2 phosphorylation and had little influence on GSK-3β phosphorylation. Addition of the GSK inhibitor BIO with FGF-2 signifi-

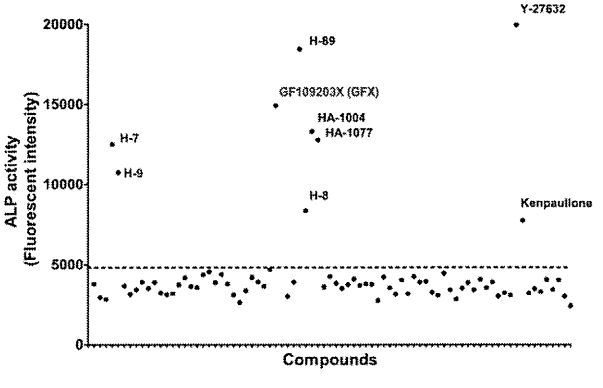

Figure 1. An ALP activity-based high-throughput screening assay of chemical library for PKC inhibitors. The ALP activity using 4-methylumbelliferyl phosphate [59] in 201B7 hiPS cells in a 96-well plate was measured by fluorometry. Each dot on the graph represents the fluorescent intensity for each compound of the kinase inhibitor library. Dotted line indicates the level for DMSO as a control.
doi:10.1371/journal.pone.0054122.g001