

Choosing Optimal STR Markers for Quality Assurance of Distributed Biomaterials in Biobanking

Tae-Hoon Chung[†], Heejung Lee[†], Mihee Lee, Jae-Pil Jeon, Kisang Kim and Bok-Ghee Han*

Korea Biobank, Center for Genome Science, Korean National Institute of Health, Korea Center for Disease Control and Prevention, Seoul 122-701, Korea

Abstract

The quality assurance (QA) is of utmost importance in biobanks when archived biomaterials are distributed to biomedical researchers. For sample authentication and cross-contamination detection, the two fundamental elements of QA, STR genotyping is usually utilized. However, the incorporated number of STR markers is highly redundant for biobanking purposes, resulting in time and cost inefficiency. An index to measure the cross-contamination detection capability of an STR marker, the mixture probability (MP), was developed. MP as well as other forensic parameters for STR markers was validated using STR genotyping data on 2328 normal Koreans with the commercial AmpFISTR kit. For Koreans, 7 STR marker (D2S1338, FGA, D18S51, D8S1179, D13S317, D21S11, vWA) set was sufficient to provide discrimination power of $\sim 10^{-10}$ and cross-contamination detection probability of ~ 1 . Interestingly, similar marker sets were obtained from African Americans, Caucasian Americans, and Hispanic Americans under the same level of discrimination power. Only a small subset of commonly used STR markers is sufficient for QA purposes in biobanks. A procedure for selecting optimal STR markers is outlined using STR genotyping results from normal Korean population.

Keywords: Biobank, cross-contamination detection, quality assurance, sample authentication, short tandem repeat

Introduction

Large scale biomedical researches aiming at identifying diagnostic or therapeutic markers and recent techno-

logical advancements in high-throughput platforms have triggered interests in high quality biomaterials. In particular, as the amount of required biomaterials for basic researches on disease diagnostics and treatment, target identification and subsequent validation is increasing in an unprecedented pace, the need for a stable supply mechanism of reliable biomaterials is urgent and biobanking has attracted a great deal of attention as a viable solution. By collecting and storing biomaterials from large numbers of individuals that can encompass a diverse spectrum of socioeconomic, demographic, and epidemiological characteristics, biobanks are already making great impacts on various epidemiological researches and quite a few biobanks are already up and running world wide (Hakonarson *et al.*, 2003; Triendl, 2003; LifeGene Sweden; Elliott & Peakman, 2008; Bhak *et al.*, 2008; Ahn, 2007).

Biobanks require high standards in every aspect of its usual operation for quality assurance (QA) purposes since poorly managed biomaterials or related biomedical information can do only harm. It is usually required to standardize all necessary steps as much as possible and document them as standard operating protocols (SOPs) in an unmistakable way (Troyer, 2008). Many biobanks are currently in an establishment phase, aiming at collecting and archiving biomaterials and their relevant biomedical information for future use. Owing to this, several working protocols of sample collection, delivery, processing, storage, and retrieval are available for reference purposes (International Society for Biological Environmental Respositories 2008; Elliott and Peakman, 2008). However, the ultimate goal of a biobank is to make the general access of research-oriented people to high quality biomaterials possible with minimal efforts. This requires an optimized procedure for sample distribution and proper measures of QA for archived biomaterials. For instance, when genome-wide association studies which usually require tens of thousands of high quality DNA samples in a short period of time are underway, providing them with a proper level of QA in the required time frame can be a difficult task.

QA of biomaterials in biobanks raises two issues: to authenticate that the specimen in a container matches exactly the participant that can be identified by the label on it and to avoid cross-contamination of biomaterials from different donors. Any of these QA breaches can mislead or nullify the research results in crucial ways. Although efforts to automate all or parts of biobanking

*Corresponding author: E-mail Bokghee@nih.go.kr
Tel +82-2-380-2251, Fax +82-2-354-1078

Accepted 1 March 2009

[†]These authors contributed equally to this work.

processes with the help of high precision machines are going on (Downey and Peakman, 2008; McQuillan and Sales, 2008; Owen, 2008; Fagan and Ball, 2008), human operations still comprise many integral portions, posing the risk of unpredictable quality problems during ordinary operations. Subsequently, there should be a proper safeguard to detect sample authentication problems or cross contamination events prior to sample distribution, preventing the utilization of non-quality-compliant samples and, eventually, erroneous or null-discoveries.

As a biobank providing high quality DNAs for several genome-wide association studies on Koreans, our team has collected DNAs from >123,000 general Koreans (age >40) since year 2001 and has been performing short tandem repeat (STR) genotyping for QA purposes. In particular, we have used the commercially available AmpFISTR kit composed of 15 autosomal and one sex chromosomal STR markers prior to sample distribution. Based on the core STR markers from the combined DNA index system (CODIS), the STR panel works superb for various jobs requiring individual identification such as forensics, mass victim identification, and paternity testing (Jobling and Gill, 2004; Ballantyn, 1997; Alford *et al.*, 1994). For instance, the chance of two random individuals having the same STR typing profile is $\sim 10^{-17}$ in many ethnic populations (Kim *et al.*, 2003). However, we think this level of performance is too redundant for QA in biobanks. As for the Korea biobank, for instance, which is aiming at collecting biomaterials from $\sim 300,000$ participants, the appropriate level of discrimination power can be 10^{-7} when 100-fold redundancy is assumed or 10^{-8} when 1000-fold redundancy is assumed. Under this premise, it is obvious that all 15 autosomal markers are not required and, by incorporating smaller number of highly polymorphic STR markers in an optimized way, cost and time required for QA can be greatly reduced. In this paper, we will introduce a general strategy to select optimal STR markers to be adopted in sample authentication and in detecting sample cross-contamination prior to distribution and illustrate it using our own STR genotyping results.

Methods

Population

Blood samples were collected from Korea university Asan hospital and Ajou University hospital in Seoul, Korea. Informed consents were obtained from all subjects. 2328 Koreans that did not have any apparent disease-related symptoms were recruited in this study.

DNA samples and PCR

Genomic DNA amplification was carried out using about 1.0ng of DNA for PCR amplification. 15 STR loci were co-amplified with the AmpFISTR1 Identifier™ PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. All amplifications were done on a GeneAmp 9700PCR system (Applied Biosystems, Foster City, CA, USA).

Genotyping

Detection and genotyping of PCR products were carried out on the ABI PRISM 3100 DNA Analyzer (Applied Biosystems Division/Perkin-Elmer, Foster City, CA, USA) according to the manufacturer's instructions. Electrophoresis results were analyzed by Genemapper v3.1 software.

Statistical analysis

Allele and genotype frequencies were calculated using the gene counting algorithm. Forensic parameters such as matching probability (MP) (Brenner and Morris, 1990), power of exclusion (PE) (Brenner and Morris, 1990), power of discrimination (PD) (Jones, 1972), probability of identity (PI) (Waits *et al.*, 2001), polymorphic information content (PIC) (Botstein *et al.*, 1980), observed heterozygosity (H), expected heterozygosity (HET) (Nei, 1987), and the deviation from the Hardy-Weinberg equilibrium (HWE) (Guo and Thompson, 1992) were calculated. The HWE was further validated using the Arlequin package (Excoffier *et al.*, 2005).

The STR marker specific probability to identify cross-contamination of biomaterials from two donors is calculated as follows. When a sample is mixed with biomaterials from two donors in approximately equal amounts, >2 peaks will appear in the genotyping result only when the genotypes from two donors are not identical. Let represent the probability that one donor's genotype is given by for a specific marker. Since we are dealing with genotypes whose parental origins are uncertain, we assume without loss of generality. Then, the mixture probability (MP) which is the probability that a marker can detect cross-contamination from two different donors is given by

$$ME = \sum_i \sum_{j \geq i} p_i (1 - p_j) \quad (1)$$

If the marker does not violate the HWE, this equation will be

$$MH = \sum_i p_i^2 (1 - p_i^2) + \sum_i \sum_{j \geq i} 2p_i p_j (1 - 2p_i p_j) \quad (2)$$

We also performed simulations to validate the MPs given by Eq. (1) or (2) using our own genotype results. From each marker's pool of genotypes, we randomly selected two genotypes. If the two genotypes are not identical, observed mixture count is increased by one. Otherwise, it is untouched. After performing the random selection process over times, the simulated MP is calculated by

$$MS=C/N \quad (3)$$

Actual simulations were carried out in a batch mode: we calculated the MS with and repeated this procedure

over 20 times. Then, the average and standard deviation of 20 MSs are reported for each marker.

Results

The forensic parameters and various MPs for 15 autosomal STR markers in AmpFISTR kit for 2328 unrelated normal Koreans are provided in Table 1. Also, full results including allele frequencies as well as parameters shown in Table 1 are provided as a supplementary material. When all 15 autosomal markers are utilized, the combined PD, PI, and PE were 1.339×10^{-17} , 1.169

Table 1. Forensic parameters and mixture probabilities for 15 autosomal STR markers

	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
MP ^a	0,0455	0,0714	0,0893	0,1180	0,1308	0,1477	0,0694	0,0809	0,0294	0,0766	0,0724	0,1905	0,0379	0,0842	0,0364
PI ^b	0,0458	0,0712	0,0883	0,1181	0,1319	0,1508	0,0697	0,0799	0,0291	0,0760	0,0726	0,1866	0,0367	0,0831	0,0340
PD ^c	0,9545	0,9286	0,9107	0,8820	0,8692	0,8523	0,9306	0,9191	0,9706	0,9234	0,9276	0,8095	0,9622	0,9158	0,9636
PE ^d	0,6435	0,5740	0,5256	0,4775	0,4370	0,3608	0,5709	0,5686	0,7007	0,5675	0,5812	0,3762	0,6766	0,5692	0,6618
PIC ^e	0,8189	0,7606	0,7330	0,6839	0,6627	0,6310	0,7696	0,7508	0,8603	0,7541	0,7643	0,5844	0,8393	0,7437	0,8463
H ^f	0,8236	0,7864	0,7592	0,7308	0,7056	0,6540	0,7847	0,7834	0,8530	0,7828	0,7903	0,6650	0,8407	0,7837	0,8331
HET ^g	0,8389	0,7863	0,7674	0,7285	0,7117	0,6790	0,7991	0,7843	0,8736	0,7835	0,7951	0,6458	0,8551	0,7774	0,8611
HWE ^h	0,0168	0,0518	0,0124	0,0004	0,9247	8,0E-05	0,0730	0	0	0,0009	0,1253	0,3816	0	0,6047	0
ME ⁱ	0,9545	0,9286	0,9107	0,8820	0,8692	0,8523	0,9306	0,9191	0,9706	0,9234	0,9276	0,8095	0,9622	0,9158	0,9636
MH ^j	0,9494	0,9226	0,9093	0,8787	0,8662	0,8446	0,9281	0,9189	0,9660	0,9175	0,9241	0,8122	0,9453	0,9157	0,9010
MS ^k	0,9484	0,9287	0,8885	0,8636	0,8605	0,8474	0,9125	0,9149	0,9419	0,9209	0,9102	0,8005	0,9381	0,9101	0,9358
STD ^l	0,0006	0,0004	0,0006	0,0006	0,0007	0,0009	0,0005	0,0007	0,0005	0,0007	0,0005	0,0009	0,0005	0,0005	0,0006

^aMatching probability; ^bProbability of identity; ^cPower of discrimination; ^dPower of exclusion; ^ePolymorphic information content; ^fobserved heterozygosity; ^gexpected heterozygosity; ^hHardy-Weinberg equilibrium test p-value; ⁱexpected mixture probability; ^jmixture probability based on HWE; ^ksimulated mean mixture probability; ^lstandard deviation of simulated mixture probability

Table 2. Cumulative power of discrimination of 15 STR markers for Korean population

Rank ^a	Marker	PD ^b	PI ^c	ME ^d	MH ^d	MS ^d
1	D2S1338	$2,943 \times 10^{-2}$	$2,910 \times 10^{-2}$	0,9705723	0,9660911	0,9419250
2	FGA	$1,071 \times 10^{-3}$	$9,883 \times 10^{-4}$	0,9989294	0,9966438	0,9962725
3	D18S51	$4,052 \times 10^{-5}$	$3,625 \times 10^{-5}$	0,9999595	0,9998163	0,9997692
4	D8S1179	$1,844 \times 10^{-6}$	$1,662 \times 10^{-6}$	0,9999982	0,9999907	0,9999881
5	D13S317	$1,280 \times 10^{-7}$	$1,158 \times 10^{-7}$	0,9999999	0,9999993	0,9999990
6	D21S11	$9,137 \times 10^{-9}$	$8,243 \times 10^{-9}$	1	0,9999999	0,9999999
7	vWA	$6,611 \times 10^{-10}$	$5,983 \times 10^{-10}$	1	1	1
8	D19S433	$5,065 \times 10^{-11}$	$4,548 \times 10^{-11}$	1	1	1
9	D16S539	$4,097 \times 10^{-12}$	$3,635 \times 10^{-12}$	1	1	1
10	D5S818	$3,452 \times 10^{-13}$	$3,020 \times 10^{-13}$	1	1	1
11	D7S820	$3,081 \times 10^{-14}$	$2,668 \times 10^{-14}$	1	1	1
12	CSF1PO	$3,636 \times 10^{-15}$	$3,151 \times 10^{-15}$	1	1	1
13	D3S1358	$4,756 \times 10^{-16}$	$4,155 \times 10^{-16}$	1	1	1
14	TH01	$7,026 \times 10^{-17}$	$6,264 \times 10^{-17}$	1	1	1
15	TPOX	$1,339 \times 10^{-17}$	$1,169 \times 10^{-17}$	1	1	1

^aSTR markers were shown in order according to their capabilities for differentiation and mixture detection for Korean population; ^bpower of discrimination (PD); ^cprobability of identity (PI); ^dmixture probabilities of ME, MH and MS. See also formula (1), (2) and (3) in Methods of the text.

$\times 10^{-17}$, and 0.999996, respectively. Since markers are independent, the collective discrimination powers were calculated by the product rule. The collective discrimination power indexes illustrate the highly polymorphic nature of STR markers in the AmpFISTR panel; these markers provide sufficient differential power for forensic or paternity test purposes. It is interesting to note that HWE was violated in 9 markers of our population (D8S1179, D7S820, CSF1PO, TH01, D16S539, D2S1338, D19S433, D18S51, FGA). This is not due to any computational artifacts of our methodology adopted in this study because similar results were also obtained from an independent population genetics package (Nei, 1987). The exact reason for this exotic behavior, however, is not fully understood yet.

As stated in Introduction, the level of collective discriminative power of the commercial kit is too redundant for QA purposes of usual biobanks where biomaterials from $\sim 10^5$ participants are to be archived for the most ambitious cases. The optimal set of STR markers for QA can, then, be obtained by recruiting markers of the best discrimination power recursively. The number of optimal markers will be determined by the number of markers whose collective discrimination power satisfies the pre-specified target value. In Table 2, we listed the cumulative PD, PI, ME, MH, MS values along STR markers ranked according to their discrimination power. The STR markers ranked according to their discrimination power for Korean population were as follows: D2S1338, FGA, D18S51, D8S1179, D13S317, D21S11, vWA, D19S433, D16S539, D5S818, D7S820, CSF1PO, D3S1358, TH01, TPOX. The discrimination power is determined in principle by the interplay between the number of admissible alleles of a marker and the homogeneity in allele frequency distributions among admissible alleles. Obviously, a marker is not suitable for sample authentication or cross-contamination detection if it has only a few admissible alleles or all of its allele frequencies are condensed in a small number of alleles. For instance, D2S1338 which possessed the best discrimination power for Koreans required 6 alleles of the largest frequency to cover 80% of total number of alleles. To the contrary, TPOX which possessed the worst discrimination power for Koreans required only 2 alleles of the largest frequency to reach that level.

For the purpose of sample authentication or cross-contamination detection, top 5~7 STR markers are enough for the Korea biobank according to our results in Table 2, since the collective individual discrimination power can be $10^{-7} \sim 10^{-10}$ with this number of STR markers. This level of discrimination power is 100~1,000 times redundant for the target number of samples to be recruited for the Korea biobank. For comparison,

Table 3. Optimal marker sets for Koreans, African Americans, Caucasian Americans, and Hispanic Americans

Korean	African American	Caucasian American	Hispanic American
D2S1338	D2S1338	D2S1338	D2S1338
FGA	D18S51	D18S51	FGA
D18S51	FGA	FGA	D18S51
D8S1179	D19S433	D21S11	D21S11
D13S317	D21S11	D8S1179	D13S317
D21S11	vWA	D7S820	D8S1179
vWA	D8S1179	vWA	vWA

In African American and Caucasian American columns, markers different from Koreans are indicated in bold letters. Markers are listed according to their respective discrimination performance in each ethnic population.

the cumulative marker performances of the same STR markers from AmpFISTR kit for African Americans, Caucasian Americans, and Hispanic Americans are detailed in the supplementary materials (Butler, 2003). With discrimination power of 10^{-10} , the optimal marker sets for Koreans, African Americans, Caucasian Americans, and Hispanic Americans are compared in Table 3. Interestingly, optimal marker sets from the 4 ethnic groups were almost similar.

Discussion

Sample authentication and cross-contamination detection are of utmost importance for QA of biomaterials archived in biobanks. Owing partly to the highly polymorphic nature of STR markers and partly to the availability of various commercial kits making the whole process of genotyping handy, STR analysis lends itself as the de facto standard approach for QA. However, the STR markers in most commercial kits are highly redundant for this purpose, with excessive combined discrimination power. When a biobank is designed, its sample collection target volume is first set and, subsequently, the optimal level of collective discrimination power can be specified. For instance, when the sample collection target number is set $\sim 10^5$ level, the optimal level of discrimination power can be $10^{-7} \sim 10^{-8}$, allowing for a level of redundancy 100~1,000 fold. In this paper, we detailed a strategy to select an optimal STR marker set to allow the predetermined target authentication power with sufficient level of cross-contamination detection power at the same time. Although our study was based on genotyping results from a commercial kit, the strategy described in our study is valid in general. After setting target levels of collective sample authentication and cross-contamination detection power, they

can choose an optimal set of STR markers simply by looking at their forensic parameters and mixture probabilities. Of course after an optimal set of STR markers is selected, biobanks should optimize a genotyping protocol so that high-throughput QA can be achieved on a regular basis. Utilizing STR markers from commercial kits can be advantageous because they are already validated for batch operations and many basic statistical characteristics of them are possibly known for the population of interest (Butler, 2006).

The sample collection in biobanks is usually carried out by distributed facilities, in which case the quality control and maintenance of data integrity can be a complicated issue (Troyer, 2008). When the collection of large number of blood samples needs to be completed in a timely manner, for instance, the risk of confusing participants' identifiers and labels on containers or epidemiological questionnaires increases. This risk will be greatly amplified when some samples are obtained from new participants and generating new identifiers will be enough for them while others are obtained from returning participants for follow-ups where existing identifiers endowed to participants should be matched exactly. Furthermore, when a large number of samples are processed and important biomaterials like DNAs are fractionated to multiple small volume tubes for long term storage, the risk of cross-contamination will also increase. In fact, there were several cases of quality breach in the Korea Biobank identified through STR genotypings. Some of the cases were traced back to label mix-ups during blood subsampling while other cases were traced back to cross contaminations during biomaterial fractionations. Although the strict adherence to established SOPs and the utilization of sophisticated machineries and information infrastructure can reduce these risks substantially, any human operations by themselves pose possible risk of errors. As the final safeguard against such unpredictable risks, sample authentication and cross-contamination detection procedures prior to sample distribution will be crucial in practice.

In the present study, various factors that could complicate the cross-contamination detection problem were omitted. For instance, we assumed that only two donors contributed their biomaterials in approximately equal amounts. Apparently, this simplified the whole problem because, in general, contributions are from >2 donors with unequal amounts. Since total amount of biomaterials is small, any transfer of foreign biomaterials during the procedure can hardly be undetected. Utilizing commercial kits can be advantageous in this case also because vendors provide ways to estimate contributions from different donors by comparing allelic peak amplitudes. Supplementing this information may make the

whole discussion more thorough.

When we compared the optimal STR marker sets from 4 ethnic populations (Koreans, African American, Caucasian American, and Hispanic Americans) under the same level of discrimination power (10⁻¹⁰), they were almost identical. In all ethnic populations, the D2S1338 marker provided the most optimal performance and FGA and D18S51 markers provided the next optimal performances. Beyond these markers, the exact ranks of STR markers were slightly different. We are not so sure whether the same universal set of STR markers can be utilized in general biobanks regardless of their ethnic background although our results seem to support the idea. This obviously requires testing our strategy in a lot more ethnic populations than we did in this study.

Acknowledgements

This work was supported by an intramural grant from Korea National Institute of Health, Korea Center for Disease Control and Prevention (4845-301-210-13).

References

- Ahn, C. (2007). Pharmacogenomics in Drug Discovery and Development. *Genomics & Informatics* 5, 41-45.
- Alford, R.L., Hammond, H.A., Coto, I., and Caskey, C.T. (1994). Rapid and efficient resolution of parentage by amplification of short tandem repeats. *Am. J. Hum. Genet.* 55, 190-195.
- Ballantyn, J. (1997). Mass disaster genetics. *Nat. Genet.* 15, 329-331.
- Bhak, J., Ghang, H., Reja, R., and Kim, S. (2008). Personal Genomics, Bioinformatics, and Variomics. *Genomics & Informatics* 6, 161-165.
- Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314-331.
- Brenner, C., and Morris, J. (1990). *Paternity index calculations in single locus hypervariable DNA probes: validation and other studies, in Proceedings for the International Symposium on Human Identification 1989.* Promega Corporation, Madison, WI pp.21-53
- Butler, J.M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *J. Forensic Sci.* 51, 253-265.
- Butler, J.M., Schoske, R., Vallone, P.M., Redman, J.W., and Kline, M.C. (2003). Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *J. Forensic Sci.* 48, 908-911.
- Downey, P., and Peakman, T.C. (2008). Design and implementation of a high throughput biological sample processing facility using modern manufacturing principles. *Int. J. Epidemiol.* 37(Suppl 1), i46-i50.
- Elliott, P., and Peakman, T.C. (2008). The UK Biobank sam-

- ple handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int. J. Epidemiol.* 37, 234-244.
- Elliott, P., and Peakman, T.C. (2008). The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int. J. Epidemiol.* 37, 234-244.
- Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin v3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1, 47-50
- Fagan, M., and Ball, P. (2008). Design and implementation of a large-scale liquid nitrogen archive. *Int. J. Epidemiol.* 37(Suppl 1), i62-i64.
- Guo, S.W., and Thompson, E.A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361-372.
- Hakonarson, H., Gulcher, J.R., and Stefansson, K. (2003). deCODE genetics, Inc. *Pharmacogenomics* 4, 209-215.
- International Society for Biological Environmental Repositories. (2008). Best practices for repositories: collection, storage, retrieval and distribution of human biological materials for research. *Cell. Preserv. Tech.* 6, 3-58.
- Jobling, M.A., and Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* 5, 739-751.
- Jones, D.A. (1972). Blood samples: probability of discrimination. *J. Forensic Sci.* 12, 355-359.
- Kim, Y.L., Hwang, J.Y., Kim, Y.J., et al. (2003). Allele frequencies of 15 STR loci using AmpFISTR identifier kit in a Korean population. *Forensic Sci. Int.* 136, 92-95.
- LifeGene Sweden. http://lifegene.ki.se/research/index_en.html
- McQuillan, A.C., and Sales, S.D. (2008). Designing an automated blood fractionation system. *Int. J. Epidemiol.* 37 (Suppl 1), i51-i55.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.
- Owen, J.M. (2008). Designing and implementing a large scale automated -80°C archive. *Int. J. Epidemiol.* 37 (Suppl 1), i56-i61.
- Triendl, R. (2003). Japan launches controversial Biobank project. *Nat. Med.* 9, 982.
- Troyer, D. (2008). Biorepository standards and protocols for collecting, processing, and storing human tissues. *Methods. Mol. Biol.* 441, 193-220.
- Waits, L.P., Luikart, G., and Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Mol. Ecol.* 10, 249-256.