# Usefulness of archival biobank samples for genetic epidemiologic studies

**Malin I. L. Sjöholm**

Avdelningarna för Klinisk Kemi och Mikrobiologi

Institutionen för Laboratoriemedicin, Malmö

Lunds Universitet

LUND UNIVERSITY
Faculty of Medicine

**Akademisk avhandling**
som med vederbörligt tillstånd från Medicinska Fakulteten vid Lunds Universitet för avläggande av doktorsexamen i Medicinsk Vetenskap kommer att offentligen försvaras i patologiska institutionens föreläsningssal, Universitetssjukhuset MAS, Malmö, onsdagen den 19 mars 2008, kl. 9.00

**Fakultetsopponent**
**Jimmie B. Vaught**
Office of Biorepositories and Biospecimen Research
National Cancer Institute
Bethesda, Maryland, USA

DOKUMENTDATABLAD enl SIS 61 41 21

| Organization | Document name |
|---|---|
| LUND UNIVERSITY<br><br>Departments of Clinical Chemistry and Microbiology<br>Institute of Laboratory Medicine<br>Malmö University Hospital, Malmö | DOCTORAL DISSERTATION |
| | Date of issue<br>March 19, 2008 |
| | Sponsoring organization |
| Author(s)<br>Malin I. L. Sjöholm | |

Title and subtitle

Usefulness of archival biobank samples for genetic epidemiologic studies


Abstract

Sweden has a long history of maintaining population-based registries and biobanks. This has resulted in large sample collections with long follow-up and large numbers of prospectively occurring disease endpoints providing an extensive resource for genetic research. However, the largest biobanks contain sample materials that have previously been considered suboptimal for genetic research.

We have evaluated the usefulness of archival formalin-fixed, paraffin-embedded (FFPE) tissue, plasma, maternity serum and dried blood spots (DBS) for genetic research. We also evaluated the usefulness of archival samples of various biological materials for herpesvirus detection and developed a multiplex matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) method for detection of herpesviruses.

Extraction of DNA from archival FFPE tissue was difficult and time consuming. The DNA was largely degraded and repeat extractions were often necessary. Nonetheless 94% of the FFPE tissue samples could be successfully genotyped.

Archival plasma and maternity serum samples contained small and variable amounts of DNA but 98% of the plasma and over 99% of the serum samples were successful in genetic analyses even after over 20 years of storage. The presence of realistic amounts of foetal DNA of a discordant genotype in the maternity serum will not cause false maternal genotyping results.

DBS samples contained small amounts of extractable DNA but over 97% of these samples were successful in genetic analysis even after 18 years of storage if they had been stored at -20°C to prevent DNA degradation.

DNA from plasma and DBS samples was of sufficient quality for successful multiple displacement amplification (MDA) with maintained bi-allelic representation as long as sufficient amounts of DNA was used as MDA template. DNA from FFPE tissue failed consistently and DNA from serum performed poorly in MDA.

The multiplex MALDI-TOF MS method we developed reliably detected HHVs in a wide variety of archival biological specimens. The concordance rate with reference methods was over 95%.



Key words: Archival samples, biobanks, epidemiology, dried blood spots, serum, plasma, formalin-fixed paraffin-embedded tissue

Classification system and/or index termes (if any):

| Supplementary bibliographical information: | | Language<br>English |
|---|---|---|
| ISSN and key title:<br>1652-8220 | | ISBN<br>978-91-85897-79-7 |
| Recipient's notes | Number of pages<br>95 | Price |
| | Security classification | |


Distribution by (name and address)


I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.


Signature _____ Date 080210

# Usefulness of archival biobank samples for genetic epidemiologic studies

**Malin I. L. Sjöholm**

Departments of Clinical Chemistry and Microbiology
Institute of Laboratory Medicine, Malmö
Lund University

LUND UNIVERSITY
Faculty of Medicine

Malmö 2008

Malin I. L. Sjöholm
Departments of Clinical Chemistry and Microbiology
Institute of Laboratory Medicine
Lund University
Malmö University Hospital
S. Förstadsgatan
205 02 Malmö
Sweden

Printed by Media-Tryck, Lund University, Sweden
© Malin I. L. Sjöholm 2008
ISSN 1652-8220
ISBN978-91-85897-79-7
Lund University, Faculty of Medicine Doctoral Dissertation Series 2008:26

Till minne av mormor och morfar

**CONTENTS**

LIST OF PAPERS 8
ABBREVIATIONS 9
ABSTRACT 10
INTRODUCTION 11
REGISTRIES 13
Population-based registries 13
Regional biobank registries 14
BIOBANKS 15
Healthcare biobanks 15
Research biobanks 17
Biobank networks 19
BIOBANK SAMPLES 21
Plasma and serum 21
Dried blood spots 22
Formalin-fixed paraffin-embedded tissue 22
DNA EXTRACTION 24
WHOLE GENOME AMPLIFICATION 26
HUMAN HERPESVIRUSES 28
PRESENT INVESTIGATIONS 30
Paper I 30
Paper II 32
Paper III 33
Paper IV 35
GENERAL DISCUSSION 37
CONCLUDING REMARKS 40
POPULÄRVETENSKAPLIG SAMMANFATTNING 42
ACKNOWLEDGEMENTS 45
REFERENCES 47
APPENDIX, Papers I-IV

## LIST OF PAPERS

This thesis is based on the following papers, referred to in the text by their roman numerals.

I **Sjöholm, M. I. L.**, Hoffmann, G, Lindgren, S., Dillner, J. and Carlson, J. Comparison of archival plasma and formalin-fixed paraffin-embedded tissue for genotyping in hepatocellular carcinoma. *Cancer Epidemiology, Biomarkers & Prevention* 2005;14(1):251-5

II **Sjöholm, M. I. L.**, Dillner, J. and Carlson, J. Assessing quality and functionality of DNA from fresh and archival dried blood spots and recommendations for quality control guidelines. *Clinical Chemistry* 2007;53(8):1401-7

III **Sjöholm, M. I. L.**, Dillner, J. and Carlson, J. Multiplex detection of human herpesviruses from archival specimens using MALDI-TOF Mass Spectrometry. *Journal of Clinical Microbiology* 2008:46(2). Manuscript accepted

IV **Sjöholm, M. I. L.**, Dillner, J. and Carlson, J. Effect of foetal DNA on genotyping of DNA from maternity serum. Manuscript.

All published articles are reproduced with permission from the publishers.

**ABBREVIATIONS**

| | |
|---|---|
| FFPE | formalin-fixed paraffin-embedded |
| DBS | dried blood spot |
| DNA | deoxyribonucleic acid |
| MDA | multiple displacement amplification |
| ICD | International classification code |
| MDCS | Malmö Diet and Cancer Study |
| MPP | Malmö Preventive Project |
| NSHDC | Northern Sweden Health and Disease Cohort |
| VIP | Västerbotten Intervention Program |
| MONICA | Monitoring Trends and Determinants in Cardiovascular Disease |
| WHO | World Health Organization |
| DiPiS | Diabetes Prediction in Skåne |
| ABIS | All Babies in Southeast Sweden |
| NBSBCCC | Nordic Biological Specimen Banks working group on Cancer Causes and Control |
| CCPRB | Cancer Control using Population-based Registries and Biobanks |
| PCR | polymerase chain reaction |
| RCA | rolling circle amplification |
| HHV | human herpesvirus |
| HSV | herpes simplex virus |
| VZV | Variciella-Zoster virus |
| EBV | Epstein-Barr virus |
| CMV | cytomegalovirus |
| HCC | hepatocellular carcinoma |
| SNP | single nucleotide polymorphism |
| HFE | hemochromatosis |
| AAT | $\alpha_1$-antitrypsin |
| CFTR | cystic fibrosis transmembrane conductance regulator |
| EDTA | ethylene diamine tetra acetate |
| EZNA | Eazy nucleic acid isolation |
| TE | Tris EDTA |
| MALDI-TOF MS | Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry |

**ABSTRACT**

Sweden has a long history of maintaining population-based registries and biobanks. This has resulted in large sample collections with long follow-up and large numbers of prospectively occurring disease endpoints providing an extensive resource for genetic research. However, the largest biobanks contain sample materials that have previously been considered suboptimal for genetic research.

We have evaluated the usefulness of archival formalin-fixed, paraffin-embedded (FFPE) tissue, plasma, maternity serum and dried blood spots (DBS) for genetic research. We also evaluated the usefulness of archival samples of various biological materials for herpesvirus detection and developed a multiplex matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) method for detection of herpesviruses.

Extraction of DNA from archival FFPE tissue was difficult and time consuming. The DNA was largely degraded and repeat extractions were often necessary. Nonetheless 94% of the FFPE tissue samples could be successfully genotyped.

Archival plasma and maternity serum samples contained small and variable amounts of DNA but 98% of the plasma and over 99% of the serum samples were successful in genetic analyses even after over 20 years of storage. The presence of realistic amounts of foetal DNA of a discordant genotype in the maternity serum will not cause false maternal genotyping results.

DBS samples contained small amounts of extractable DNA but over 97% of these samples were successful in genetic analysis even after 18 years of storage if they had been stored at -20°C to prevent DNA degradation.

DNA from plasma and DBS samples was of sufficient quality for successful multiple displacement amplification (MDA) with maintained bi-allelic representation as long as sufficient amounts of DNA was used as MDA template. DNA from FFPE tissue failed consistently and DNA from serum performed poorly in MDA.

The multiplex MALDI-TOF MS method we developed reliably detected HHVs in a wide variety of archival biological specimens. The concordance rate with reference methods was over 95%.

## INTRODUCTION

In 2003 a new act was passed in Sweden that defines a biobank as biological material from one or more humans that is collected and stored for an extended period of time or indefinitely with associated information that makes it possible to trace the material back to the donors. The act also states that in order to ensure the donors integrity, donors must be informed about why the material is stored, what it may be used for, that no biological material is stored or used without the donors consent and that donors can change their consent at any time (1).

Sweden has a long history of maintaining large population-based healthcare and research biobanks and establishing population-based registries. This has resulted in large sample collections with long follow-up and large numbers of prospectively occurring disease endpoints providing an extensive resource for genetic research. All residents in Sweden have a unique personal identification code, obtained at birth or when becoming a Swedish citizen, that is used in all registries and biobanks and enables automatic and rapid identification of biobank samples and registry information from all donors, a prerequisite for efficient large-scale population-based genetic research.

By using population-based registries and biobanks in research study design, much larger numbers of cases and controls, which are highly representative of the population, can be rapidly identified than could be prospectively accumulated in many years, ensuring strong statistical power even for uncommon diseases. The biobank study design also ensures complete attendance of selected subjects and minimises selection biases caused by systematic differences between individuals that do or do not participate in the study (2). In addition, the use of biobank samples collected before disease occurrence enables investigations of causal exposures that may have occurred long before diagnosis and minimises the risk of reverse causality biases.

The rapidly evolving field of genetic research today offers techniques for high throughput genetic analyses that enable rapid analysis of large number of samples, often analysing many genetic markers simultaneously. The alleviation of throughput concerns regarding the collection of eligible study material in sufficiently large numbers for good statistical power and improved performance of the genetic analyses have opened up an increasing interest in the large collections of biobanked material that have previously been considered suboptimal for genetic research. Such collections are comprised of materials that contain small amounts of DNA or that have been stored under such conditions that the quality of DNA may have

been compromised. However, before such samples can be extensively used in genetic research, a thorough and systematic evaluation of the extractability, quality and performance of DNA from these samples must be conducted.

## REGISTRIES

### Population-based registries

By using population-based nationwide registries the full capacity of biobanks can be exploited. By linking registries to biobanks, using the personal identification code, information about sample-availability and location can be retrieved. The following registries are examples of population-based registries in Sweden that can be useful for genetic epidemiologic studies:

The Cancer Registry

The registry was founded in 1958 and contains information about all Swedish residents diagnosed with cancer. All cancers are classified according to the international classification of diseases (ICD). About 50 000 new malignant cancer cases are reported to the register every year (3).

The Cause of Death Registry

The registry contains information about the cause of death of all deceased Swedish residents since 1961 and less complete information since 1952. All diagnoses are classified according to the ICD. About 90 000 – 95 000 deaths have been registered each year for the last few years (4).

The Patient Registry

The registry contains information about Swedish residents that have used the healthcare system. Information including diagnoses, operations and cause of injury is available, but limited, from 1964 and covers all healthcares from 1997 (5).

The Swedish Medical Birth Registry

The registry contains information about pregnancies, labours and deliveries, diagnoses of the mother and child and general health status of newborns in Sweden since 1973. About 100 000 newborns are registered each year (6).

The Foetal Malformation Registry

The registry was founded in 1964 and contains information about foetus disorders and chromosomal abnormalities. About 2000 incidences are reported every year (7).

The Multi-Generation Registry

The registry contains information that link children and their biological or adoptive parents. All individuals who have become Swedish residents after 1961 or were born in 1932 or later and their parents are included. In total the registry contains information about some 11 million unique individuals, with complete coverage from, and including, 1968 (8).

The Pharmacy Registry

The registry contains information about prescription-based pharmaceuticals, consumables and provisions dispatched by pharmacies since 1999. About 90 million prescriptions are dispatched every year (9).

The Swedish Twin Registry

The registry was established in 1961 and contains over 86 000 twin pairs. Information about cancer and death incidences is available for all twin pairs as well as varying additional information depending on the research studies conducted (10).

National Healthcare Quality Registries

Fifty-six registries that contain information about diagnosis, treatment and outcome for, among others, diseases of the respiratory, circulatory, digestive, urinary and genital organs, the endocrine and nervous systems and eye, muscular, skeletal and psychological diseases were included in 2007 (11).

**Regional biobank registries**

The registries were established after the implementation of the new biobank law. The purpose of the registries is to develop and maintain quality control systems for storage and use of biobank materials that ensure integrity protection of the donors. The registries also serve to improve the usefulness of biobanks for research by providing overview, maintaining routines for consent, change of consent, localization and destruction of samples from dissenting donors and maintaining a registry of all sample collections in the region (12).

## BIOBANKS

### Healthcare biobanks

The largest biobanks in Sweden consists of samples collected for healthcare purposes. The joint collection of the biobanks of the Swedish healthcare system has been estimated to contain up to 100 million samples and increasing by over 3 million new samples every year (13).

Biobanks collected for population-based screening

Nationwide serological screening for Rubella immunity and viral infections during pregnancy is scheduled to take place during week 14 of pregnancy. The screening program has almost complete attendance. About 90 000 serum samples are collected each year and stored for 5-10 years, although some regions have stored samples since the 1970s (13, 14). The Northern Sweden Maternity cohort has collected and stored serum samples since 1975 and now consists of almost 120 000 samples from 86 000 women (15). The Southern Sweden Maternity Cohort contains over 100 000 samples from 74 000 women, stored at -20°C (14, 15).

Cervical cytology screening for the detection and prevention of cervical cancer has been conducted in Sweden since the 1960s. About 985 000 samples in the form of cervical cells fixed on glass-slides are collected nationwide each year and stored for at least 10 years although many laboratories have stored samples since the screening was introduced. The combined cervical collections comprise of over 20 million samples (13, 16, 17).

Neonatal screening programs for inborn errors of metabolism, the most common being phenylketonuria originally described by Guthrie in 1963 (18), have been conducted in Sweden since 1974. Samples are stored as dried blood spots (DBS) on filter papers from all newborn infants. This collection comprises DBS samples from about three million subjects which is virtually the entire population of Sweden below 30 years of age (13).

The Swedish Institute for Infectious Disease Control has served as a reference laboratory for many microbiological analyses for decades. The institute has performed several population-based nationwide investigations on the immunity against infections in the Swedish population

and many of the samples that have been analysed have been stored since 1957 as a part of the quality control system. The resulting biobank contains over 900 000 biological samples, most of which are serum samples, stored at -20°C (19).

Microbiology biobanks

About 600 000 serum samples, excluding pregnancy screening samples, are collected nationwide in Sweden each year and stored at clinical microbiology departments (13). The Malmö microbiology biobank contains over 1.3 million samples, stored at -20°C, and increases by about 120 000 samples each year (14, 15). The majority of these are serum samples collected from the serological screening during pregnancy or submitted for diagnosis of viral infections or for research purposes (14).

Cytology and pathology biobanks

About 950 000 new tissue or cellular samples, not including the cervical screening samples, are stored each year at clinical pathology and cytology departments in Sweden (13).

The biobanks at the Departments of Clinical Pathology and Cytology in Malmö contains paraffin block of about 2.4 million surgical tissue samples and about 1.1 million autopsy tissue samples and almost 4 million histological slides. The oldest samples date back to 1944 (16).

The Fresh Tissue Biobank in Uppsala contains about 26 000 fresh-frozen tissue samples, about 5000 fresh-frozen cell pellets, stored at -80°C, and about 20 000 viable cell suspensions, stored in liquid nitrogen, that have been collected since the 1970s. Most of the tissue samples are from solid tumours and most of the cell samples are from leukaemia and lymphoma patients (20).

The tissue biobank at Karolinska Hospital contains fresh-frozen tissue samples obtained during surgery and stored at -70°C. Tissue samples from about 3500 thyroid, parathyroid and abdominal endocrine tumours have been stored since 1986 and samples from breast, urological, gastrointestinal tumours and melanoma have been stored since 2002 (21).

**Research biobanks**

Many population-based biobanks have been created specifically for research purposes. Although these sample collections are generally much smaller than the healthcare biobanks, the samples are usually of larger amounts and often collected for investigations of specific diseases such as cancer, cardiovascular diseases and diabetes. Biological samples and extensive information about the participants are collected using standardised methods and questionnaires within each project. Informed consent for the specific research intended is collected from the participants at enrolment. Listed below are a few of the largest population-based research biobanks in Sweden.

Malmö Diet and Cancer Study (MDCS)

The main objective of the study is to evaluate the effect of diet habits on cancer risk for the most common cancers in Sweden. The goals are also to investigate the risk of cardiovascular diseases, diabetes, hypertension and osteoporosis in relation to dietary habits and to study the evolution of the health status of the participants over a 10 year period.

Over 28 000 men and women, resident in the Malmö area, participated in the study. An extensive survey about the participants' diet habits and general health was conducted at enrolment in 1991-1996. Blood samples were collected from all participants and stored as blood cells, plasma, serum and DNA extracted from buffy coats at -140, -80 or -20°C (22, 23).

Malmö Preventive Project (MPP)

The objective of the MPP project was to identify individuals with high risk of developing cardiovascular diseases or alcohol addiction and offer preventive medical treatment. The study was conducted between 1974 and 1991 with an attendance of about 22 000 men and 10 000 women. Blood samples were collected from all participants and stored as serum and plasma at -20°C. The participants filled in a questionnaire about their medical history and were subjected to physical and laboratory tests (24, 25). About 18 000 DNA samples from the originally attending individuals that participated in a re-examination have also been stored and a new questionnaire was conducted at re-examination. There is an overlap of about 30% between the MPP and MDCS biobanks.

Northern Sweden Health and Disease Cohort (NSHDC)

The cohort contains three sub-cohorts, the Västerbotten Intervention Program (VIP), Monitoring Trends and Determinants in Cardiovascular Disease (MONICA) and the Mammography screening in Västerbotten. About 50% of the MONICA and mammary screening cohorts also participate in the VIP study. In total the NSHDC contains samples from 85 000 individuals, stored at -80°C (15, 26).

The VIP study started in 1985 with the objective to screen for cardiovascular diseases and diabetes and offer prevention treatment. The aim was also to promote a healthy diet and lifestyle in Västerbotten, the northern region of Sweden. Residents of the region are invited to participate in the study as they reach 30, 40, 50 or 60 years of age. Demographic, medical and lifestyle information about the participants was collected through a questionnaire and a medical examination was conducted at recruitment (27). Blood samples from volunteering participants was collected and stored as plasma, buffy coat and erythrocyte fractions. Every 10 years the participants are invited for a health examination, to update their questionnaire information and to donate new blood (27). The biobank contains samples from 70 000 individuals (15).

In 1985 the two most northern counties in Sweden joined the WHO MONICA project. The objective of the project was to increase the understanding of risk factors for cardiovascular diseases. The WHO project terminated in 1994 but the northern Sweden MONICA part continued on with recurring population surveys and registration of cardiovascular disease incidences (28). The MONICA biobank consists of blood samples from 9000 individuals (15).

Blood samples collected in connection with mammary screening in Västerbotten have been stored since 1995. From 1997 the age group 50-69 years is screened and sampling repeated every second year. The biobank contains 48 000 samples from over 27 000 individual women (15, 26).

Diabetes Prediction in Skåne (DiPiS)

The study started in 2000, with he purpose to identify predictive risk factors for diabetes in children, and comprises of over 33 000 participants. The biobank consists of blood samples collected form the umbilical cord and from the mother at delivery as well as blood samples

collected from the children at 2 years of age, once a year until 5 years of age and every other or third year until 15 years of age. Information including familial history of diabetes is collected through questionnaires at enrolment and at samples collections (29).

All Babies in Southeast Sweden (ABIS)

Over 17 000 children that were born in the southeast of Sweden were enrolled in the study between 1997 and 1999. Hair strands and breast milk from the mother and blood samples from the umbilical cord were collected at delivery. Information about diet and infections during the pregnancy as well as the mothers living environment was collected in connection with the birth through a questionnaire. After 12 months information about infections, diet and vaccination of the children as well as blood, urine, faeces and hair strand samples were collected. A follow up of the children started in 2006 and is planned to take place at ages 8, 11 and 14. Whole blood, dried blood spots, urine, faeces, hair and saliva and information about the children's lifestyle will be collected at follow up.

The objectives of the study is to investigate the role of environmental factors during pregnancy and the early years for development of childhood diabetes and to investigate the role of diet habits, environmental factors and heritable genetic traits as risk factors for immunological disorders (30).

**Biobank networks**

National and international collaborations between large population-based biobanks with long follow-up, vast number of disease endpoints and quality assured associated lifestyle information that are linkable to population registries allows for research study designs of extraordinary size and quality. Following is a description of a few biobank network projects, with the objective to study risk factors for cancer and to improve the quality of biobanks and associated register data.

Nordic Biological Specimen Banks working group on Cancer Causes & Control (NBSBCCC)

NBSBCCC is a network consisting of 17 biobank cohorts and 5 cancer registries. The purpose of the network is to provide resources for etiological studies of cancer. The research is specifically focused on longitudinal studies. The combined resources of the network include more than 2 million samples donors, 25 million person years and 100 000 prospective cancers. Over 30 joint network articles have been published (15).

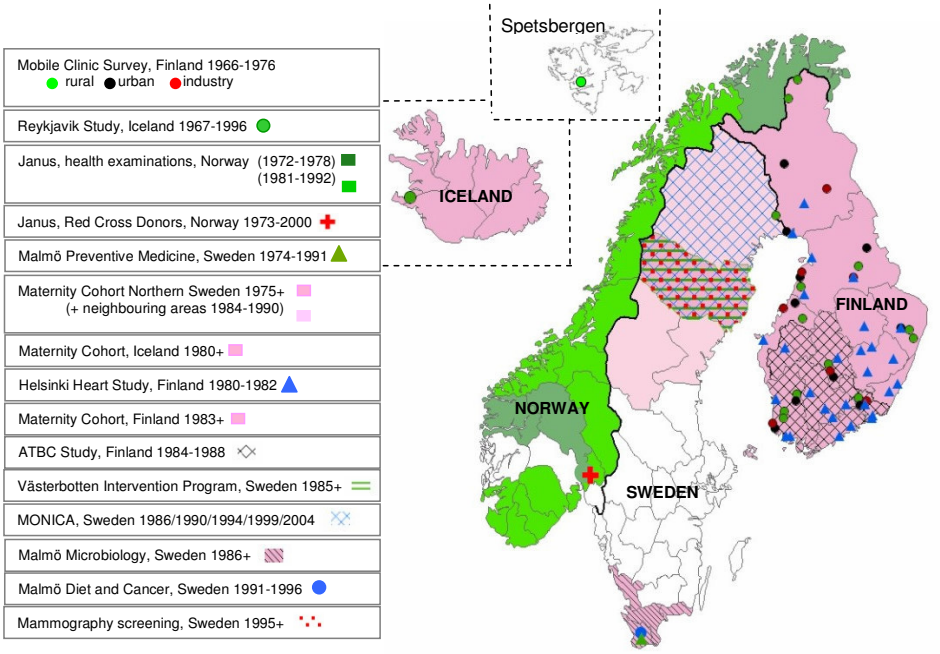


**Figure 1.** Areas covered by the biobanks in the Nordic countries that participate in the NBSBCCC network, adapted from Pukkala *et al,* 2007(15).

Cancer Control using Population-based Registries and Biobanks (CCPRB)

The CCPRB network is a collaboration between 18 partners from 9 European countries, it includes 20 biobanks and 7 registries and has access to a number of platforms for advanced technological analyses. The objectives of the network include providing a study base for large population-based research on genetic and infectious causes of cancer and to define and promote the implementation of quality standards for biobanking and of integrity-proof methods for biobank-based research. The biobanks included in the network comprise of over 60 000 prospective cancer cases with up to 30 years of follow-up. Over 180 joint network articles have been published using resources from the network (31).

## BIOBANK SAMPLES

The biobanks of Sweden consist of many different types of biological material. The research biobanks contain sample types specifically selected for the intended research analyses while healthcare biobanks contain sample materials best suited for the clinical investigations conducted. The pathology and cytology departments store materials such as cell suspensions, cells mounted on glass-slides, cytological brush samples and fresh frozen or formalin-fixed paraffin embedded tissue. The majority of samples stored at the microbiology departments are serum but other samples such as cerebrospinal fluid, sore secretion, blister material, bronchoalvelolar lavage, conjunctival fluid and urine are also submitted for diagnosis of viral or bacterial infections. Other sample types such as buccal swabs, saliva, whole blood, buffy coats and blood clots are also collected and stored.

In my work I have focused on evaluating the usefulness of plasma, serum, FFPE tissue and DBS for human genetic analyses. These are all types of biological material that are commonly found in vast numbers particularly in the healthcare biobanks but not originally intended for use in genetic research.

### Plasma and serum

Cell-free circulating DNA is found in serum and plasma of both healthy individuals and patients suffering from various diseases. Much research has been devoted to using cell-free DNA as a biomarker for diseases, such as cancer (32-37), preeclampsia (38, 39) and rheumatoid arthritis (40), with the rationale that more cells undergo cell death in affected than healthy individuals. It has been reported that serum contains significantly higher amounts of cell-free DNA than plasma but the reason for this difference has not been determined. Sample handling techniques, such as inclusion of buffy coat cells when preparing plasma samples or lysis of nucleated cells after clot formation due to prolonged storage of blood samples before harvesting serum, may affect the DNA content (32, 41-44). The amount of DNA in both plasma and serum is usually small and varies greatly between individuals (32, 34, 40, 41, 43-45) which complicates the use of these samples for genetic studies. Although DNA extracted from serum has been reported to be fragmented (45) it has been successfully used in genetic analyses (46-48).

A large portion of the microbiology biobanks consists of serum collected during pregnancy. Maternity serum from healthy women has been reported to contain cell-free foetal DNA in

fractions of up to 0.5% (49) in early, and up to 7% (50) in late pregnancy. The amount of foetal DNA has been shown to be enough for non-invasive prenatal diagnosis (51-54). A concern with using maternity serum for genetic analyses of the mother is therefore that the foetal DNA could affect the maternal genotyping results.

**Dried Blood Spots**

Collecting DBS on filter papers has been widely used for newborn screening for many decades because the samples are easy to prepare, handle and ship (55). The samples are traditionally used for the analysis of metabolites, hormones and proteins but although DBS samples, particularly small discs, contain small volumes of blood, leucocytes are likely to be present in all samples. The possibility to obtain DNA from DBS was reported in 1987 (56) and DBS samples have since then been demonstrated to be useful for genetic analysis and successfully used for genetic research purposes (57-63) even after storage for 25 years (64). The DNA yield from DBS has been found to be largely independent of storage time (64) but performance in PCR found to increase with decreasing amplicon length particularly for samples stored for a long time (64, 65). The absorption properties have been reported to differ between different filter paper brands (66). Filter papers have also been specifically designed for DNA extraction purposes by binding PCR inhibitors such as protein, haemoglobin and iron while preserving DNA in an aqueous extractable form (67).

**Formalin-fixed paraffin-embedded (FFPE) tissue**

Pathological specimens are routinely formalin-fixed and paraffin-embedded in order to preserve the tissue architecture and proteins necessary for histological evaluation (68). DNA extracted from FFPE tissue has been shown to be largely fragmented (69), difficult to dissolve (70) and to perform poorly in PCR (71) with the result of no amplification products or artificial mutations. Formalin has been reported to denature DNA by breaking hydrogen bonds and unstacking bases (70) and to cause cross-linking between DNA strands and between histones and DNA (72). The cross-linking may cause aberrant incorporation of bases during PCR amplification (73). DNA damage has also been reported to induce recombination, insertion at the end of the fragments and jumping of enzyme between templates (74) during PCR amplification.

The length of time between death and autopsy, between surgical excision and fixation, the type of preservative used and the fixation time have been shown to affect the integrity of the

DNA (69-71, 75). Modifications of DNA have also been found to be dependent on the concentration, temperature and pH of formalin (70, 76).

Nonetheless, evaluation of FFPE tissue for genetic studies has found few or no misclassifications of genotypes (47, 77, 78) and artefacts correlated to the amount of DNA used as PCR template (73). It has also been shown that DNA quality and performance in PCR increases with decreasing sample storage time (68, 79) and amplicon length (79, 80).

## DNA EXTRACTION

In order to acquire the best possible quality of DNA from biological samples it is of great importance to choose an appropriate extraction method. The best choice of method depends on the type of biological starting material, the storage conditions and the intended use of the extracted DNA. While DNA extraction from samples that contain large amounts of nucleated cells and have been stored under optimal conditions are relatively uncomplicated more consideration is required for choosing extraction methods for samples known to contain degraded or small amounts DNA and samples stored under suboptimal conditions.

The phenol/chloroform method has been traditionally used for DNA extraction and purification from various types of biological materials (68, 80-83). The method is laborious and necessitates the use of toxic reagents. Many simpler methods such as the use of resin (64, 68), simple proteinase K digestion (84) and simple boiling (45, 57, 68, 80, 82) have therefore been developed for rapid and in some cases automated extraction. In some reports PCR has been successfully performed directly on crude samples (46, 61, 85).

There are also many commercially available kits for both automated and manual DNA extraction with protocols specifically adapted for different sample materials. Many automated methods use magnetic beads to bind and recover DNA during the purification process (36, 86-89) with the advantage of rapid extractions of many samples simultaneously. Automated or semi automated methods are therefore particularly useful for large-scale studies in which the extraction step is one of the bottlenecks.

Other common extraction methods use spin columns with membranes that bind DNA. These can often be obtained in plate formats that enable the simultaneous extraction of 96 samples. The advantage with spin columns is that the eluate is usually pure and free from contaminants such as blood residues or reagent carryovers that may affect downstream applications. Many spin column products allow for elution with water and have the additional quality of removing very small DNA fragments. This may be particularly beneficial when extracting DNA from FFPE tissue where elimination of the smallest fragments may improve performance in subsequent genetic analyses. DNA extraction from FFPE tissue also requires removal of the paraffin. The most common method to achieve this is xylene/ethanol treatment but more rapid and less toxic methods such as microwave and boiling treatments (68) are also used.

Because plasma, serum and small DBS samples can contain very small or variable amounts of DNA the use of extraction methods that concentrate the DNA in small elution volume may be

appropriate for these samples types. It may also be essential to be able to determine the DNA concentration in the extract for the DNA to be used in genetic analyses where the template amount is critical. When the DNA yield is low it is often necessary to use a large proportion of the extract in genetic analyses in order to ensure sufficient amounts of template DNA. It is therefore particularly important to use extraction methods that result in extracts free from contaminants.

## WHOLE GENOME AMPLIFICATION

All biobanks contain limited amounts of sample material. It is therefore very important to ensure that the samples are used with great care, not to deplete these precious sources of research material that may be difficult or impossible to replenish.

One approach to overcome the problem of low and variable amounts of DNA is the use of whole genome amplification techniques such as multiple displacement amplification (MDA). The MDA technique is based on the rolling circle amplification (RCA) (90) mechanism and was originally developed for amplification of circular DNA templates (91) but was later modified for amplification of linear templates (92, 93).

The MDA procedure used throughout my research is based on the bacteriophage Phi 29 enzyme. The reaction is performed under isothermal conditions after initial rapid heat denaturation. Random hexamer primers anneal to a multitude of sites on the DNA template and the enzyme initiates elongation. Due to the strand-displacement capacity of the enzyme downstream elongated primers are displaced by upstream strands, growing in the same direction. Displaced strands are targets for new random priming events elongated in the opposite direction (93).

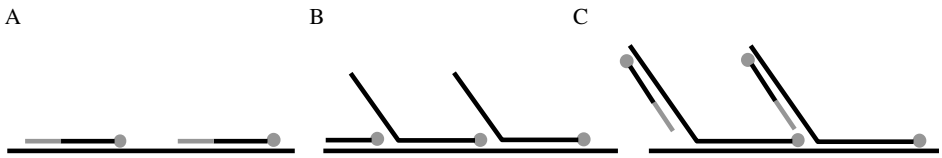


**Figure 2**. Schematic representation of multiple displacement amplification of linear DNA. A) Primers (light grey lines) anneal to the DNA template and Phi 29 (circles) initiates elongation. B) Strand displacement by Phi 29. C) New priming events and elongation takes place on the displaced strands.

By use of MDA microgram quantities of DNA can be produced from a few copies of human genomic DNA (92). However, MDA requires long template strands for successful amplification and is not suitable for sheared DNA, such as DNA extracted from FFPE tissue (93).

The proofreading capacity of the Phi 29 enzyme has an error rate of $3x10^{-6}$ to $5x10^{-6}$ which is about 100-fold lower than that of the *Taq* DNA polymerase (94). High concordance between amplification products and genomic DNA and near-complete coverage of the genome (95-98) as well as allele amplification biases causing genotypic miscalls (99) and chimeric DNA rearrangements (100) have been reported using whole genome amplification methods based on the Phi 29 enzyme. It has also been reported that allelic biases introduced during amplification can be significantly reduced using pooled DNA from several separate amplification reactions (101).

## HUMAN HERPESVIRUSES

The use of large population-based registries and biobanks is not only advantageous for human genetic research but also for large-scale studies evaluating the clinical and epidemiological importance of viral infections. In my work I have focused on herpesvirus screening in sample types commonly submitted for diagnosis of viral infections.

The human herpesviruses (HHV) are large enveloped viruses that contain linear double-stranded DNA. They can cause lytic, persistent, latent/recurrent and in some cases immortalizing infections. Herpesviruses are ubiquitous and infections are common. They usually cause subclinical infections but are also associated with morbidity and mortality, particularly in immune-suppressed patients (102).

There are 10 herpesviruses that infect humans, Herpes Simplex Virus (HSV) 1, Herpes Simplex Virus 2, Varicella-Zoster Virus (VZV), Epstein-Barr virus (EBV) A, Epstein-Barr virus B, Cytomegalovirus (CMV), Human Herpes Virus 6 (HHV6) A, Human Herpes Virus 6B, Human Herpes Virus 7 (HHV7) and Human Herpes Virus 8 (HHV8). An overview of the human herpesviruses and the clinical syndromes they can cause is presented in Table 1 (102-107).

Most herpesviruses are able to cause foetal infections through *in utero* transmission (108). Intrauterine herpesvirus infections may cause birth defects (102), premature delivery (109), foetal varicella syndrome, neurodevelopmental handicap and foetal mortality (108). EBV infection during pregnancy has been associated with childhood acute lymphoblastic leukaemia (110).

The diagnostic techniques that are commonly used for herpesvirus detection include antigen or antibody detection assays (102, 109, 111), PCR (106, 112-115) and dot blot hybridization (106). These techniques usually require separate methods for the detection of each herpesvirus. Evaluation of all herpesviruses in a single sample would thereby be time and sample consuming and cost ineffective. There is therefore an increasing interest in the development of techniques for large-scale, rapid and simultaneous detection of multiple herpesviruses (116-118).

**Table 1.** The human herpesviruses and the clinical syndromes they can cause.

| Subfamily, virus | Abbreviation | Clinical syndromes |
|---|---|---|
| **Alphaherpesviridae** | | |
| Herpes Simplex Virus 1 | HSV1 | Oral lesions, (Genital lesions), Encephalitis |
| Herpes Simplex Virus 2 | HSV2 | Genital lesions, (Oral lesions), Meningitis |
| Varicella-Zoster Virus | VZV | Chicken pox / Shingles, Pneumonia |
| **Gammaherpesviridae** | | |
| Epstein-Barr Virus A/B | EBV | Infectious mononucleosis, Lymphoproliferative diseases |
| Human Herpes Virus 8 | HHV8 | Febrile illness, Kaposi's sarcoma |
| **Betaherpesviridae** | | |
| Cytomegalovirus | CMV | Infectious mononucleosis, Retinitis, Pneumonia |
| Human Herpes Virus 6 A/B | HHV6 | Roseola, Pneumonia, Encephalitis |
| Human Herpes Virus 7 | HHV7 | Rash illnesses, Encephalitis |

## PRESENT INVESTIGATIONS

### Comparison of archival plasma and formalin-fixed paraffin-embedded tissues for genotyping in hepatocellular carcinoma (paper I)

Primary liver cancer is the fifth most common cancer and the third most common cause of cancer mortality worldwide (119-121). Hepatocellular carcinoma (HCC) is the predominant form, accounting for 85-90% of primary liver cancers. In Sweden, liver cancer is predominantly caused by liver cirrhosis due to alcohol abuse and/or chronic hepatitis caused by Hepatitis C Virus (122). Heritable disorders associated with cirrhosis include cystic fibrosis, $\alpha_1$-antitrypsin deficiency and hereditary hemochromatosis (121, 123).

Formalin treatment of tissue has been shown to reduce DNA solubility and induce DNA degradation (70) and the content of genomic DNA in serum or plasma from healthy subjects has been reported to be low and of a wide range (34-37, 40, 47, 124). Biobanks containing these sample types have therefore seldom been used for genetic epidemiologic studies since they have been considered suboptimal for genotyping purposes. We wished to compare the reliability of genetic analyses carried out on these types of archived biological materials and to attempt the use of MDA as a means to overcome the problem of low and variable amounts of DNA in serum and plasma samples.

Using the Swedish Cancer Registry we identified 384 cases of primary HCC, 318 of these were also identified in the autopsy registry at the Department of Pathology at Malmö University Hospital. We retrieved FFPE liver tissue from all 318 autopsied patients and plasma or serum samples, stored for 10-30 years in a separate biobank, from 31 of the same patients.

After extraction the DNA concentration of the FFPE tissue samples was determined using PicoGreen which measures double stranded DNA. The DNA yield from ten 5µm sections (1 x 1 x 0.5cm) was between 0.01 and 1.6µg. Real-time PCR analysis showed that about 40% of the extracted DNA was functional in PCR. The DNA concentration of the plasma and serum samples, extracted from 1mL aliquots, was determined by real-time PCR alone due to the low amounts of DNA present in such sample materials. The DNA yield from plasma and serum was between 0.2 and 72ng/mL. The structural integrity of the DNA was evaluated by agarose gel analysis and found to be largely degraded when extracted from FFPE liver tissue and of insufficient amounts to be visualized on gels when extracted from plasma or serum.

MDA was successful on a titration series from plasma samples and generated 5 000 – 43 000 fold amplification of DNA of large fragments, but repeated attempts of MDA of the FFPE tissue samples failed consistently. Real-time PCR was used to determine the DNA concentration of the MDA product, due to the primer-on-primer amplification that can occur in MDA reactions when no DNA template is present.

The genotype of four proposed genetic risk factors for HCC, three single nucleotide polymorphisms (SNPs) causing hereditary hemochromatosis (HFE C282Y, HFE H63D), $\alpha_1$-antitrypsin deficiency (AAT E342K) and a triplet deletion causing cystic fibrosis (CFTR ΔF508), were determined by two independent assays on neat FFPE tissue, plasma and serum extracts as well as MDA products.

Genotyping was successful in 94% of all tissue samples. Performance in genotype analyses decreased with increasing storage time and PCR amplicon length. This could be due to differences in tissue processing or DNA alterations occurring over time due to formalin treatment. The $\alpha_1$-antitrypsin deficiency SNP had the lowest success rate. The site of this SNP is close to a telomere, it corresponds to a CpG methylation site and the PCR amplicon was relatively long, all of which may explain the low success rate. DNA extraction from and genotype analysis of FFPE tissue samples was challenging and required repeated extractions and PCR analyses. Although the extracted DNA was fragmented and not amplifiable by MDA, sufficient material for repeated extraction and genotype analysis was available.

Genotyping was successful in 98% of the plasma and serum samples. All neat plasma and serum extracts and all MDA products gave genotyping results identical to those of the tissue samples from the same subject as long as ≥0.2ng template DNA was used in the MDA reaction. Using less than 0.2ng DNA as MDA template resulted in allelic dropouts. This suggests that the DNA in plasma samples is structurally intact even after storage for >20 years at suboptimal conditions and that at least 0.2ng DNA should be used as MDA template to ensure bi-allelic representation.

The results of this study indicate that FFPE tissue can be successfully used in genetic epidemiologic studies and that archival plasma is a useful starting material for genetic epidemiologic studies, particularly if the extracted DNA is subjected to MDA prior to genotyping as long as at least 0.2ng DNA is used in the MDA reaction.

**Effect of foetal DNA on genotyping of DNA from maternity serum (paper II)**

During pregnancy foetal DNA circulate in the maternal blood system. Although the amount is small it has been demonstrated to be sufficient for non-invasive prenatal diagnosis (51-54). We therefore wished to investigate whether the content of foetal DNA in maternity serum could be sufficient to alter the maternal genotyping results, which would render maternity serum inappropriate for genetic studies. This was done by evaluating the concordance between genotypes obtained on DNA extracted from fresh whole blood with that extracted from archival maternity serum from the same women.

In order to evaluate the fraction of foetal DNA necessary to affect the maternal genotype DNA extracts from fresh EDTA whole blood samples were mixed to contain 0.5 to 99.5% heterozygous DNA and analysed for a SNP by TaqMan PCR. Less than 10% DNA of a discordant genotype in a sample did not affect the genotype call. Homozygous samples containing between 10 and 50% DNA of a heterozygous genotype gave undetermined results. At least 50% DNA of a discordant genotype was required to alter the genotype result.

Using the sample registry and serum biobank at the Department of Microbiology at Malmö University Hospital serum samples previously obtained during pregnancy from women, with fresh whole blood samples available at the department of clinical chemistry, were identified and retrieved. The serum samples had been stored for up to 21 years.

DNA was extracted from 200µL of the whole blood and archival serum samples. Some of the serum samples were subjected to MDA. The DNA concentrations were determined by real-time PCR and all neat extracts and MDA products were analysed for ten high frequency SNPs. The DNA yield was between 0 and 4800 ng/mL serum and decreased with increasing storage time. About 19% of the samples were unusable in PCR due to low DNA yield. There was no difference in DNA yield between samples taken during the first and second trimester. The genotyping was successful in over 99% of all whole blood and serum samples, as long as 0.4ng DNA were used in the analyses, and all serum samples gave genotyping results identical to those of the corresponding whole blood samples.

Two serum samples produced indeterminate genotypes, situated between genotype clusters. This could be due to interference by foetal DNA on the maternal genotype. However, as these results were obtained for only one SNP for each sample this seems unlikely. If these two

serum samples contained sufficient foetal DNA to alter the genotype result for one SNP, the genotype alterations should be apparent in several of the ten analysed high frequency SNPs.
DNA extracted from the serum samples performed poorly in MDA. Over 90% of the samples that were analysed after MDA could be automatically or manually genotyped, but the genotyping results of some 10% of these were discordant with the corresponding neat serum samples. This poor performance could be due to inhibitors in the serum extracts or degradation of the DNA during storage. Success in MDA and subsequent genotyping call and concordance rates were lowest for the MDA products of the serum samples with the lowest DNA yield, it may therefore be useful to select more samples than needed for statistical power so that the lowest yield stratum can be excluded at an early stage, particularly if the samples will be subjected to MDA.
These results indicate that archival maternity serum taken in the first or second trimester, even after storage for up to 21 years, is a useful source for genetic epidemiologic studies, but not for MDA. The presence of realistic amounts of foetal DNA in maternity serum may cause failure of genotypic assignment but will not cause false maternal genotyping results, on the TaqMan system.

**Assessing quality and functionality of DNA from fresh and archival dried blood spots and recommendations for quality control guidelines (paper III)**

The use of DBS is a convenient and inexpensive method of biobanking. In Sweden DBS samples from newborn infants have been stored since 1975 (64). We wished to evaluate the effects of storage time and temperature on the quality of DNA in DBS and the performance of the extracted DNA in MDA in order to describe a protocol for optimization of procedures and quality control guidelines for use of DBS in genetic epidemiologic studies.
In a pilot study DBS were created from reference whole blood with known white blood cell count to test four common DNA extraction methods (Qiagen, EZNA, Chelex 100 and alkaline lysis). The extracted DNA was quantified by PicoGreen, OliGreen which measures single stranded DNA, and real-time PCR and the size of the DNA fragments was determined by agarose gel analysis. The DNA yield from a DBS discs, measured as a percentage of the theoretical yield, determined with PicoGreen, were 14% using the Qiagen method and 12% using the EZNA method. Similar results were obtained by real-time PCR. DNA extracted

using the Chelex 100 and alkaline lysis methods performed poorly in real-time PCR. The extracts, using all methods except alkaline lysis, contained large DNA fragments.

A titration series of DNA from each extraction method, except alkaline lysis, was successfully subjected to MDA and the products contained large DNA fragments. DNA extracted from the reference whole blood extracts and MDA products were analysed for three SNPs by TaqMan PCR and 254 SNPs by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). All MDA products gave genotyping results identical to those of the original whole blood samples from the same subject as long as 5 ng template DNA extracted with the EZNA method was used in the MDA reaction. Reproducibility of real-time PCR results improved when the MDA product was dissolved in TE buffer over night. The success rates decreased and allelic dropouts increased when lower amounts of template DNA were used in the MDA reaction. These results indicate that DNA of sufficient amount and quality for successful MDA performance can be obtained from small DBS.

Based on the results from the pilot study we chose the EZNA method to extract DNA from archival DBS that had been stored for ~26 years at room temperature or 3 month or ~22 years at -20°C. In order to improve the DNA yield, the proteolysis time was increased. After extraction, all samples were subjected to MDA using 5ng DNA as template. The DNA in the neat DBS extracts and MDA products were evaluated for quantity and size and genotyped for the three TaqMan and 101 of the MALDI-TOF MS SNPs. The degree of fragmentation of DNA in the neat extracts increased with increasing storage time but the size of DNA fragments appeared to be similar after MDA regardless of storage time or temperature. Some of the 26-year old samples showed allelic dropouts, when comparing the genotype results of the MDA products to the neat DBS extract from the same samples, and low concordance rates between duplicates. This poor performance could be due to the samples being largely fragmented, probably caused by degradation of the DNA during storage at room temperature. Archival samples stored at −20°C showed no allelic dropouts, regardless of storage time, and only a small percentage of discrepancies between duplicates.

Based on these results we recommend storage of DBS at −20°C, optimization of extraction methods, the use of ≥5ng DNA as MDA template and solubilization of MDA products prior to genotyping. Our recommendations for quality control guidelines include quality-assured storage and documentation, tracking of sample identity, DNA extraction by a method validated by fragment size evaluation, quantification by PicoGreen and real-time PCR and

performance in MDA of the extracted DNA. Evaluation of MDA product by real-time PCR and performance in genetic analyses is also recommended.
Using these recommendations archival DBS can provide DNA of sufficient amounts and quality to be successfully used in MDA and subsequent genetic epidemiological studies.

**Multiplex detection of human herpesviruses from archival specimens by using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (paper IV)**

The human herpesviruses (HHVs) are involved in a variety of diseases (102-105, 107-110). Common laboratory techniques in HHV detection usually require separate methods for each HHV and clinical information for selection of virus assay (106, 111-115). We wished to develop an efficient screening method for qualitative multiplex detection of all HHVs by MALDI-TOF MS and investigate the usefulness of a wide variety of archival sample types for HHV detection.
Two multiplex systems were designed to detect the human herpesviruses by MALDI-TOF MS. The detection limit was between 5 and 100 copies of HHV plasmid controls and the sensitivity for detecting multiple infections was 78% at 100 copies. The specificity of the method was between 92 and 100% using separate samples, containing 100 copies, for each virus control. The detection limit in plasma spiked with HHV controls was 2500 copies/mL. These results indicate that the multiplex methods we developed are reliable for HHV detection.
Total nucleic acid was extracted from a variety of archival biological material including bronchoalveolar lavage, conjunctival fluid, sore secretion, blister material, plasma, serum and urine. The extracts were analysed for presence of viral DNA using MALDI-TOF MS and, for some of the samples, real-time PCR. The viral yield of control samples containing known amounts of virus was 24.3%. The detection of HHVs fluctuated between runs in some samples. Real-time PCR results indicated that these inconsistent findings were related to low viral loads. The results of the MALDI-TOF methods were compared to results of reference PCR methods. The concordance rates were between 86.4 and 97.9 percent, depending on HHV type, with an overall concordance of 95.6% ($\kappa$ 0.90). In some cases the requested diagnostic testing for VZV was negative but the MALDI-TOF MS analysis detected HSV and vice versa. In other cases multiple infections and unsuspected viruses were detected.

These results indicate that our multiplex MALDI-TOF MS methods will allow large-scale research studies on archival samples of various biological materials. The results also indicate the difficulty in selecting the correct test based on clinical symptoms and suggests that a broad multiplex HHV analysis may be useful also in clinical diagnostic testing.

## GENERAL DISCUSSION

Conducting genetic research using study designs based on case identification in validated population-based registries and linkage to biological material in our extensive population-based biobanks for retrieval of case samples and selection and retrieval of matched control samples ensures studies with large numbers of disease endpoints, minimal selection bias and excellent statistical power.

With the technology that is available for human genetic epidemiologic research today high throughput performance of large-scale studies is no longer a practical or economic impossibility. It is therefore of increasing interest to evaluate the usefulness of typical sample materials, primarily found in archival healthcare biobanks, which have not been extensively exploited for genetic research such as plasma, serum, DBS and FFPE tissue.

When evaluating the quality and potential usefulness of DNA extracted from several sample types, we have developed some basic general guidelines. Regardless of available technical platforms, it is always wise to perform a pilot study using multiple representative aliquots of fresh samples of the intended type of material in the intended analyses. The quantity of DNA produced by candidate DNA extraction protocols can be evaluated by measuring absorbance at 260/280 nm, or by picogreen fluorescence specific for double stranded DNA. Quantitation by real-time PCR demonstrates the functional yield, and dilution series can reveal the presence of inhibitors to PCR. Agarose gel electrophoresis can be used to evaluate the fragment size. The performance of DNA extracts in MDA reactions not only evaluates the intactness of the DNA but also, when successful, produces large amounts of material from small amounts of starting product which allows precious archival samples to be used for large numbers of genetic analyses without depleting the biobanks.

The method of choice for quality control throughout a project depends not only on available technology, cost, accuracy and precision of the method, but also on the amount and rarity of the sample type and the need for exact measures. Although precision of repeat measurements of picogreen fluorescence may be superior, real-time PCR can be performed on very small template amounts and is recommended for determining the DNA concentration in extracts from serum, plasma and DBS. Similarly, as the entire extract frequently contains insufficient DNA for visualisation in agarose gels, the performance in MDA reactions and subsequent genotyping is a better indicator of adequate fragment size. Evaluation of performance in

genetic analyses ultimately determines the usefulness of the sample materials for the specific analyses.

The correctness of genotyping performed on minimal archival materials or MDA products should preferably be evaluated by comparison with a more "reliable" template from the same individual. We have therefore compared genotyping results of neat archival plasma and maternity serum extracts and MDA products of DBS, plasma and maternity serum with those obtained from DNA extracts from fresh EDTA whole blood or FFPE tissue samples from the same individuals.

Among the studies reported here, a large clinical pathology biobank allowed retrieval of all 318 autopsied cases of the total of 384 incident cases of hepatocellular carcinoma within a local geographical area. In addition to being used for quality control of DNA extracted from FFPE tissue, plasma and serum genotyping of HFE and AAT (but not CFTR) confirmed an increased risk for HCC in patients that were homozygote mutant or heterozygote for AAT E342K or homozygote mutant for HFE C282Y. A separate study is evaluating the specific clinical features of these patients.

Population-based biobanks and registries are equally useful for studies evaluating the clinical and epidemiological importance of viral infections and high through-put simultaneous detection of viruses are therefore of great interest. The multiplex herpesvirus detection methods we have developed can be useful for large-scale herpesvirus screening for research purposes as well as improving the diagnostic accuracy, speed and economy of clinical herpesvirus testing. In combination with the methods that have been developed for multiplex screening of other viral groups such as hepatitis (125-128) and human papilloma viruses (129, 130) and panels that are being developed for human genetic disorders our method can also become useful in smaller clinical laboratories.

In the rapidly evolving field of genetic research we can never know what techniques will be available, what biobank samples will be used for or what quality demands there will be on DNA to be useful in genetic research in the future. During the past five years alone our quality control evaluations have evolved from using PCR analysis with visualisation on agarose gel to real-time PCR and genotype analysis of a single SNP per sample by RFLP to multiplex analyses simultaneously detecting over 30 SNPs per samples and whole genome amplification generating enough DNA for a multitude of analyses to be performed on small and precious samples.

We can therefore not ensure that current methods for sample collection and storage, DNA extraction and quality control will be useful and sufficient in the future. Nonetheless we can encourage documented collection, storage and handling of samples under the best available conditions, and evolution of stringent quality control criteria following basic guidelines as new technologies evolve.

## CONCLUDING REMARKS

Although FFPE tissue samples usually yield relatively large amounts of DNA, the extraction of DNA from this type of material is difficult and time consuming and the resulting DNA is often largely degraded, necessitating repeat extractions. Nonetheless, FFPE tissue can be successfully used in genetic epidemiologic studies.

Archival plasma and maternity serum are useful starting materials for genetic epidemiologic studies. DNA from plasma can be successfully used in MDA, if 0.2ng DNA is used in the reaction to ensure bi-allelic representation. Although the presence of realistic amounts of foetal DNA of a discordant genotype in the maternity serum may cause failure of genotypic assignment it will not cause false maternal genotyping results.

As many plasma and serum samples give very low yields, and as MDA and genotyping success is related to yield, projects should be planned with more samples than needed for statistical power so that the lowest yield stratum can be excluded at an early stage of the project.

Archival DBS samples can provide DNA of sufficient quality for successful MDA and subsequent large-scale genetic epidemiologic studies if 5ng DNA is used in the MDA reactions in order to ensure bi-allelic representation. DBS should be stored at -20°C to prevent DNA degradation.

In order to establish quality control criteria it is important to conduct pilot studies to evaluate candidate extraction methods and DNA quality of candidate sample types on the intended analysis platforms. It is also of great importance to continuously re-evaluate the quality control requirements in light of the fast evolving techniques for genetic research.

The multiplex MALDI-TOF MS method we developed reliably detects HHVs in a wide variety of archival biological specimens, allowing for large-scale research studies. It may also be highly useful for multiplex clinical diagnostic testing.

The usefulness of biobanks for genetic epidemiologic research of complex or polygenic diseases has been dramatically demonstrated during the past years in several large collaborative genome wide association studies on diabetes (131), cancer (132-135), cardiovascular disease (136) and hyperlipidemia (137). Studies designed to use high quality samples with associated phenotypic information from research biobanks for genome wide association analysis and smaller amounts of DNA derived from archival biological materials such as serum for replication of findings may become highly important in the next few years. Studies such as those reported here are necessary to guide the selection of useful sample materials and extraction methods and to determine the necessary number of samples for such important research.

## POPULÄRVETENSKAPLIG SAMMANFATTNING

Insamling och förvaring av biologiskt material i biobanker har länge förekommit både inom vården och för forskningsändamål. I en del biobanker har biologiskt material kontinuerligt sparats ända sedan 1940-talet. Detta har medfört att det i Sverige finns ett stort antal biobanker som tillsammans innehåller ett mycket stort antal prov. Enbart provsamlingarna från vården uppskattas innehålla uppemot 100 miljoner prov och växa med över 3 miljoner prov per år.

Tack vare dessa stora biobanker går det relativt snabbt att välja ut tillräckligt många prov för att genomföra storskaliga forskningsstudier. Genom att använda biobanksprov är det möjligt att skapa urvalsgrupper som representerar befolkningen. Detta minskar risken för vilseledande forskningsresultat som ofta kan uppstå på grund av att det kan finnas systematiska skillnader mellan personer som väljer att delta i forskningsstudier och befolkningen som helhet.

De största biobankerna innehåller många prov i form av plasma, serum, blod intorkat på filterpapper samt vävnad som fixerats med formalin och paraffin. Dessa prov har inte använts till genetisk epidemiologisk forskning i så stor utsträckning eftersom de inte har ansetts optimala för genetiska analyser på grund av att de innehåller små mängder eller fragmenterat DNA. Många prov som undersökts för påvisning av virusinfektioner har också sparats i stora biobanker.

Målen med denna avhandling är att undersöka hur användbara olika sorters biologiskt material som lagrats i biobanker är för att undersöka genetiska mutationer hos människor och för virusidentifiering samt att utveckla effektiva metoder som kan identifiera flera olika herpesvirus samtidigt i ett och samma prov.

I delarbete I utvärderade vi kvaliteten på DNA extraherat från fixerad vävnad och serum eller plasma från samma individer. DNAt undersöktes för fyra olika DNA mutationer. Vi undersökte även möjligheterna att kringgå problemet med att plasma- och serumprov innehåller små och varierande mängder DNA med hjälp av helgenomisk amplifiering. Genom helgenomisk amplifiering kan små mängder DNA kopieras tiotusentals gånger så att många analyser kan utföras på prov som innehåller litet DNA.

Fixerad vävnad kunde inte användas till helgenomisk amplifiering eftersom DNA-kvaliteten var för dålig. Det krävdes upprepade extraheringar och analyser för att få bra resultat i mutationsanalyserna av dessa prov. Svårigheterna med vävnadsproven beror förmodligen på att formalinet brutit ner och fragmenterat DNAt. Trots detta gick det att få resultat från 94 %

av proven. DNA-kvaliteten var sämre i prov som lagrats länge än de som lagrats en kortare tid. Mutationsanalyserna gav lyckade resultat för 98 % av plasmaproven. Kvaliteten på DNA extraherat från plasmaproven var tillräckligt bra för lyckad helgenomisk amplifiering. Vävnadsprov, plasmaprov och helgenomiskt amplifierad plasma från samma individer gav identiska resultat i mutationsanalyserna.

I delarbete II undersökte vi om serum taget under graviditet kan användas till genetiska analyser eller om DNA från fostret påverkar analysresultaten. Vi extraherade DNA från färskt blod från ett hundratal kvinnor och från serumprov som tagits under graviditet från samma kvinnor och lagrats i en biobank. Ju längre tid proven hade lagrats desto mindre mängd DNA gick det att utvinna från dem och från en del prov gick det inte att få ut något DNA. Vi undersökte DNAt för tio vanliga genetiska mutationer. Två av serumproven gav misslyckade resultat i en mutationsanalys vardera. Detta skulle kunna ha orsakats av DNA från fostret i serumet men det är inte så troligt eftersom dessa prov då borde ge misslyckade resultat för fler än en av de tio analyserna. Resultaten från resten av serumproven stämde överens med resultaten av blodprovet från samma kvinnor. Genom att blanda två prov som gett olika resultat i en vanlig mutationsanalys kunde vi se att serumprov kan ge misslyckade resultat om det innehåller mellan 10 och 50 % foster-DNA men så länge de innehåller mindre än 50 % foster-DNA finns det ingen risk för resultat som inte stämmer överens med mammans DNA. DNA från serum fungerade dåligt i helgenomisk amplifiering vilket kan bero på att det brutits ner under lagringen.

I delarbete III undersökte vi om DNA extraherat från blod intorkat på filterpapper är av tillräckligt hög kvalitet för att kunna användas till helgenomisk amplifiering och genetiska analyser. Vi applicerade färskt blod på filterpapper och lät det torka. DNA extraherades sedan från små cirklar som stansats ut från blodfläcken. Efter helgenomisk amplifiering av DNA från filterpappren analyserades det för över 250 mutationer. DNA extraherades även från det färska blodprovet och analyserades för samma mutationer. Resultaten av mutationsanalyserna var identiska för alla prov som härstammade från samma individ. Alla helgenomiskt amplifierad prov gav resultat som stämde överens med blodprovet från samma individer i mutationsanalyserna.

Vi undersökte även DNA extraherat från filterpapperprov som sparats i 3 månader i minus 20°C, cirka 22 år i minus 20°C och cirka 26 år i rumstemperatur. DNAt amplifierades helgenomiskt och analyserades sedan för över hundra DNA-mutationer. Resultaten för tre av proven som lagrats i rumstemperatur stämde inte överens före och efter helgenomisk

amplifiering. Mer än tio procent av analysresultaten på duplikat av dessa prov skiljde sig också åt. De dåliga resultaten för dessa prov beror förmodligen på att DNAt brutits ner och fragmenterats vid rumstemperatur. De prov som förvarats i minus 20°C gav identiska resultat före och efter helgenomisk amplifiering för alla individer och endast en liten del av analysresultaten skiljde sig åt mellan provduplikat.

I delarbete IV ville vi utveckla en metod som kan identifiera alla olika herpesvirus på samma gång i ett och samma prov samt undersöka hur användbara arkiverade prov av varierande biologiska material är för herpesvirusidentifiering.

Vi utvecklade två analyser som tillsammans kan identifiera alla olika herpesvirus som infekterar människor. DNA extraherades från patientprov, av varierande biologiska material, och analyserades med vår metod. Alla prov hade tidigare analyserats för herpesvirus med referensmetoder. Resultaten från vår metod stämde överens med referensmetoderna till 95,6 %. Några av de prov där herpesvirus identifierats med referensmetoden men inte med vår metod undersöktes med en kvantitativ metod för att bestämma antalet viruskopior i proven. Mängden virus i alla dessa prov var under detektionsgränsen för vår metod. Några av proven som var negativa i den diagnostiska referensanalysen för ett herpesvirus visade sig vara positiva för ett annat herpesvirus när de analyserades med vår metod. Detta tyder på att det kan vara svårt att välja rätt virustest baserat på kliniska symptom och att vår metod inte bara är användbar för storskalig herpesvirusforskning utan även skulle kunna vara användbar vid klinisk diagnostik.

Sammanfattningsvis visar dessa resultat att arkiverat biologiskt material i form av fixerad vävnad, plasma, serum taget under graviditet och filterpapperprov kan användas för genetiska epidemiologiska studier. Helgenomisk amplifiering av DNA från plasma och filterpapperprov gör det möjligt att utföra ett stort antal analyser på dessa prov trots att de innehåller små mängder DNA. Resultaten visar även att metoden vi utvecklat för herpesviruspåvisning med tillförlitlighet kan identifiera flera herpesvirus samtidigt i ett och samma prov och att arkiverade biologiska prov av varierande material kan användas för herpesvirusanalys.

**ACKNOWLEDGEMENTS**

I would like to express my sincere gratitude to all the people who help me complete this thesis.

I would especially like to thank:

**Joyce Carlson,** my supervisor, for all your guidance and support throughout these years, for being enthusiastic, encouraging and always accessible. I consider my self very fortunate to have had you as my supervisor.

**Joakim Dillner**, my co-supervisor, for welcoming me into your group and for your guidance.

I would also like to thank:

all my co-workers at DNA-lab for creating a joyful working place and for always being glad to lend a helping hand;

**Maria Sterner and Liselotte Hall**, for your constantly good spirits and all your help, particularly with the masspek.

**Rebecca Rappner, Anna Letelier, Eva Lavant and Anna Hedelius**, for creating a warm and happy office space.

**Agneta Hekton Sterner**, for putting up with teaching me the ropes when I first started.

**Agneta Östensson**, for being my support in the freeze room.

**Gun-Britt Lindahl, Zahara Koochekpour, Stefan Strömberg, Daniel Ottwall, Camilla Valtonen André and Christer Haldén.**

**Per Simonsson and Johan Malm** for providing me with a workplace at the department of Clinical Chemistry.

all the members of the Dillner-group: **Kia Sjölin, Christina Gerouda, Aline Marshall, Anna Söderlund Strand, Zoltan Korodi, Janka Ryding, Kristina Hazard, Annika Lundstig, Maria Anderberg, Kristin Andersson, Carina Eklund, Helena Faust, Johanna Kullander, Natasha Vasiljevic, Ola Forslund, Helena Persson and Anna Olofsson Franzoia**, for making me feel very welcome in the group and for fun dinner and games evenings.

**Sophia Harlid,** my soon-to-be co-author, for all your verbal and practical support both at the lab and in our spare time, for putting up with all my whining and for being a good friend.

**Anna Johansson Nilgran,** for all the crazy and fun times we've had during our 20 years of friendship and for still being my friend.

**Monica Bertram**, for the friendly (and sometimes extremely annoying) competitiveness during most of our school years that pushed me to study just a little bit harder and for still being my friend.

**Eva Ytterberg**, for introducing me to the fascinating world of biochemistry and for your enthusiasm for the subject that started this whole thing.

**Lars-Anton Ivarsson**, for your pep-talks, endless love and support and rock steady belief in me.

My parents, **Anna-Lena and Håkan Sjöholm,** and siblings, **Magnus and Marie,** for your encouragement, unconditional love and for always believing in me.

I am also grateful for financial support from the Swedish National Biobanking Program which is financed by the Knut and Alice Wallenberg Foundation and by the EU 6th framework grant CCPRB (Cancer Control using Population-based Registries and Biobanks).

## REFERENCES


1. Sveriges Riksdag. Lag (2002:297) om biobanker i hälso-och sjukvården m.m. www.riksdagen.se/webbnav/index.aspx?nid=3911&bet=2002:297 (Accessed January 22, 2008).
2. Dillner J. [How to handle biological material? New legislation concerning biobanks in Finland]. Lakartidningen 2001;98:4478-9.
3. Socialstyrelsen. Cancerregistret. www.socialstyrelsen.se/Statistik/statistik_amne/cancer/cancerregistret.htm (Accessed January 22, 2008).
4. Socialstyrelsen. Dödsorsaksregistret. www.socialstyrelsen.se/Statistik/statistik_amne/dodsorsaker/Dodsorsaksregistret.htm (Accessed January 22, 2008).
5. Socialstyrelsen. Patientregistret. www.socialstyrelsen.se/Statistik/statistik_amne/sluten_vard/Patientregistret.htm (Accessed January 22, 2008).
6. Socialstyrelsen. Medicinska födelseregistret. www.socialstyrelsen.se/Statistik/statistik_amne/graviditet/MFR.htm (Accessed January 22, 2008).
7. Socialstyrelsen. Verksamheten för övervakning av fosterskador. www.socialstyrelsen.se/Statistik/statistik_amne/Missbildningar/missbildning.htm (Accessed January 22, 2008).
8. Statistiskacentralbyrån. Flergenerationsregistret. www.scb.se/templates/Standard____22842.asp (Accessed January 22, 2008).
9. Socialstyrelsen. Läkemedelsregistret. www.socialstyrelsen.se/Statistik/statistik_amne/lakemedel/Lakemedelsregistret.htm (Accessed January 22, 2008).
10. Karolinska Institutet. Svenska Tvillingregistret. http://ki.se/ki/jsp/polopoly.jsp?d=13013&l=sv (Accessed January 22, 2008).
11. Sveriges Kommuner och Landsting. Nationella Kvalitetsregister inom hälso- och sjukvården. www.skl.se/lopsedel.asp?C=3441 (Accessed January 25, 2008).
12. Regionaltbiobanksregister. Södra sjukvårdsregionen. www.biobanksregistersyd.se/page.asp?page=om (Accessed January 22, 2008).
13. The National Biobank Program. Swedish Biobanks. www.biobanks.se/swedish.htm (Accessed January 24, 2008).
14. The National Biobank Program. Malmö Microbiology Biobank. www.biobanks.se/swedish.htm (Accessed January 24, 2008).
15. Pukkala E, Andersen A, Berglund G, Gislefoss R, Gudnason V, Hallmans G, et al. Nordic biological specimen banks as basis for studies of cancer causes and control--more than 2 million sample donors, 25 million person years and 100,000 prospective cancers. Acta Oncol 2007;46:286-307.

16. The National Biobank Program. Biobanks of the Department of Clinical Pathology and Cytology. www.biobanks.se (Accessed January 24, 2008).

17. Gunnell AS, Ylitalo N, Sandin S, Sparen P, Adami HO, Ripatti S. A longitudinal Swedish study on screening for squamous cell carcinoma and adenocarcinoma: evidence of effectiveness and overtreatment. Cancer Epidemiol Biomarkers Prev 2007;16:2641-8.

18. Guthrie R, Susi A. A Simple Phenylalanine Method for Detecting Phenylketonuria in Large Populations of Newborn Infants. Pediatrics 1963;32:338-43.

19. The Nationa lBiobank Program. Swedish Institute for Infectious Disease Control Biobank. www.biobanks.se/swedishinstitute.htm (Accessed January 25, 2008).

20. The National Biobank Program. Fresh Tissue Biobank at Clinical Pathology Uppsala. www.biobanks.se (Accessed January 24, 2008).

21. The National Biobank Program. Tissue Biobank at Karolinska Hospital. www.biobanks.se/swedish.htm (Accessed January 25, 2008).

22. The National Biobank Program. Malmö Diet and Cancer. www.biobanks.se (Accessed February 5, 2008).

23. Berglund G. Malmö Diet and Cancer Study. www.mdcs.mas.lu.se (Accessed January 24, 2008).

24. Nilsson P, Berglund G. Prevention of cardiovascular disease and diabetes: lessons from the Malmo Preventive Project. J Intern Med 2000;248:455-62.

25. The National Biobank Program. Malmö Preventive Medicine. www.biobanks.se (Accessed February 5, 2008).

26. The National Biobank Program. Medical Biobank. www.biobanks.se (Accessed January 25, 2008).

27. Lukanova A, Bjor O, Kaaks R, Lenner P, Lindahl B, Hallmans G, Stattin P. Body mass index and cancer: results from the Northern Sweden Health and Disease Cohort. Int J Cancer 2006;118:458-66.

28. MONICAregistret. WHO monica center northern Sweden. www.umu.se/phmed/medicin/monica (Accessed January 25, 2008).

29. Diabetesprediktion i Skåne. DiPiS. www.endo.mas.lu.se/dipis (Accessed January 24, 2008).

30. ABIS. Alla Barn i Sydöstra Sverige. www.abis-studien.se (Accessed January 30, 2008).

31. CCPRB. Cancer Control using Population-based Registryes and Biobanks. www.cancerbiobank.org/project_description.htm (Accessed January 25, 2008).

32. Gautschi O, Bigosch C, Huegli B, Jermann M, Marx A, Chasse E, et al. Circulating deoxyribonucleic Acid as prognostic marker in non-small-cell lung cancer patients undergoing chemotherapy. J Clin Oncol 2004;22:4157-64.

33. Flamini E, Mercatali L, Nanni O, Calistri D, Nunziatini R, Zoli W, et al. Free DNA and carcinoembryonic antigen serum levels: an important combination for diagnosis of colorectal cancer. Clin Cancer Res 2006;12:6985-8.

34. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 1977;37:646-50.

35. Sozzi G, Conte D, Mariani L, Lo Vullo S, Roz L, Lombardo C, et al. Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients. Cancer Res 2001;61:4675-8.

36. Stemmer C, Beau-Faller M, Pencreac'h E, Guerin E, Schneider A, Jaqmin D, et al. Use of magnetic beads for plasma cell-free DNA extraction: toward automation of plasma DNA analysis for molecular diagnostics. Clin Chem 2003;49:1953-5.

37. Wu TL, Zhang D, Chia JH, Tsao KH, Sun CF, Wu JT. Cell-free DNA: measurement in various carcinomas and establishment of normal reference range. Clin Chim Acta 2002;321:77-87.

38. Lau TW, Leung TN, Chan LY, Lau TK, Chan KC, Tam WH, Lo YM. Fetal DNA clearance from maternal plasma is impaired in preeclampsia. Clin Chem 2002;48:2141-6.

39. Lo YM, Leung TN, Tein MS, Sargent IL, Zhang J, Lau TK, et al. Quantitative abnormalities of fetal DNA in maternal serum in preeclampsia. Clin Chem 1999;45:184-8.

40. Leon SA, Ehrlich GE, Shapiro B, Labbate VA. Free DNA in the serum of rheumatoid arthritis patients. J Rheumatol 1977;4:139-43.

41. Chan KC, Yeung SW, Lui WB, Rainer TH, Lo YM. Effects of preanalytical factors on the molecular size of cell-free DNA in blood. Clin Chem 2005;51:781-4.

42. Holdenrieder S, Stieber P, Chan LY, Geiger S, Kremer A, Nagel D, Lo YM. Cell-free DNA in serum and plasma: comparison of ELISA and quantitative PCR. Clin Chem 2005;51:1544-6.

43. Taback B, O'Day SJ, Hoon DS. Quantification of circulating DNA in the plasma and serum of cancer patients. Ann N Y Acad Sci 2004;1022:17-24.

44. Umetani N, Hiramatsu S, Hoon DS. Higher amount of free circulating DNA in serum than in plasma is not mainly caused by contaminated extraneous DNA during separation. Ann N Y Acad Sci 2006;1075:299-307.

45. Ekstrom PO, Bjorge T, Dorum A, Longva AS, Heintz KM, Warren DJ, et al. Determination of hereditary mutations in the BRCA1 gene using archived serum samples and capillary electrophoresis. Anal Chem 2004;76:4406-9.

46. Ulvik A, Ueland PM. Single nucleotide polymorphism (SNP) genotyping in unprocessed whole blood and serum by real-time PCR: application to SNPs affecting homocysteine and folate metabolism. Clin Chem 2001;47:2050-3.

47. Blomeke B, Bennett WP, Harris CC, Shields PG. Serum, plasma and paraffin-embedded tissues as sources of DNA for studying cancer susceptibility genes. Carcinogenesis 1997;18:1271-5.

48. Andolfatto S, Namour F, Garnier AL, Chabot F, Gueant JL, Aimone-Gastin I. Genomic DNA extraction from small amounts of serum to be used for alpha1-antitrypsin genotype analysis. Eur Respir J 2003;21:215-9.

49. Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet 1998;62:768-75.

50. Swinkels DW, de Kok JB, Hendriks JC, Wiegerinck E, Zusterzeel PL, Steegers EA. Hemolysis, elevated liver enzymes, and low platelet count (HELLP) syndrome as a complication of preeclampsia in pregnant women increases the amount of cell-free fetal and maternal DNA in maternal plasma and serum. Clin Chem 2002;48:650-3.

51. Bischoff FZ, Nguyen DD, Marquez-Do D, Moise KJ, Jr., Simpson JL, Elias S. Noninvasive determination of fetal RhD status using fetal DNA in maternal serum and PCR. J Soc Gynecol Investig 1999;6:64-9.

52. Dhallan R, Guo X, Emche S, Damewood M, Bayliss P, Cronin M, et al. A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study. Lancet 2007;369:474-81.

53. Honda H, Miharu N, Ohashi Y, Ohama K. Successful diagnosis of fetal gender using conventional PCR analysis of maternal serum. Clin Chem 2001;47:41-6.

54. Gonzalez-Gonzalez MC, Garcia-Hoyos M, Trujillo MJ, Rodriguez de Alba M, Lorda-Sanchez I, Diaz-Recasens J, et al. Prenatal detection of a cystic fibrosis mutation in fetal DNA from maternal plasma. Prenat Diagn 2002;22:946-8.

55. McCabe ER. Utility of PCR for DNA analysis from dried blood spots on filter paper blotters. PCR Methods Appl 1991;1:99-106.

56. McCabe ER, Huang SZ, Seltzer WK, Law ML. DNA microextraction from dried blood spots on filter paper blotters: potential applications to newborn screening. Hum Genet 1987;75:213-6.

57. Yang M, Hendrie HC, Hall KS, Oluwole OS, Hodes ME, Sahota A. Improved procedure for eluting DNA from dried blood spots. Clin Chem 1996;42:1115-6.

58. Zhong XB, Leng L, Beitin A, Chen R, McDonald C, Hsiao B, et al. Simultaneous detection of microsatellite repeats and SNPs in the macrophage migration inhibitory factor (MIF) gene by thin-film biosensor chips and application to rural field studies. Nucleic Acids Res 2005;33:e121.

59. Lin Z, Suzow JG, Fontaine JM, Naylor EW. A high throughput beta-globin genotyping method by multiplexed melting temperature analysis. Mol Genet Metab 2004;81:237-43.

60. Fitness J, Dixit N, Webster D, Torresani T, Pergolizzi R, Speiser PW, Day DJ. Genotyping of CYP21, linked chromosome 6p markers, and a sex-specific gene in neonatal screening for congenital adrenal hyperplasia. J Clin Endocrinol Metab 1999;84:960-6.

61. Borras E, Coutelle C, Rosell A, Fernandez-Muixi F, Broch M, Crosas B, et al. Genetic polymorphism of alcohol dehydrogenase in europeans: the ADH2*2 allele decreases the risk for alcoholism and is associated with ADH3*1. Hepatology 2000;31:984-9.

62. Mitterer G, Bodamer O, Harwanegg C, Maurer W, Mueller MW, Schmidt WM. Microarray-based detection of mannose-binding lectin 2 (MBL2) polymorphisms in a routine clinical setting. Genet Test 2005;9:6-13.

63. Paynter RA, Skibola DR, Skibola CF, Buffler PA, Wiemels JL, Smith MT. Accuracy of multiplexed Illumina platform-based single-nucleotide polymorphism genotyping

compared between genomic and whole genome amplified DNA collected from multiple sources. Cancer Epidemiol Biomarkers Prev 2006;15:2533-6.

64. Hannelius U, Lindgren CM, Melén E, Malmberg A, von Dobeln U, Kere J. Phenylketonuria screening registry as a resource for population genetic studies. J Med Genet 2005;42:e60.

65. Chaisomchit S, Wichajarn R, Janejai N, Chareonsiriwatana W. Stability of genomic DNA in dried blood spots stored on filter paper. Southeast Asian J Trop Med Public Health 2005;36:270-3.

66. Adam BW, Alexander JR, Smith SJ, Chace DH, Loeber JG, Elvers LH, Hannon WH. Recoveries of phenylalanine from two sets of dried-blood-spot reference materials: prediction from hematocrit, spot volume, and paper matrix. Clin Chem 2000;46:126-8.

67. Makowski GS, Davis EL, Hopfer SM. Amplification of Guthrie card DNA: effect of guanidine thiocyanate on binding of natural whole blood PCR inhibitors. J Clin Lab Anal 1997;11:87-93.

68. Coombs NJ, Gough AC, Primrose JN. Optimisation of DNA and RNA extraction from archival formalin-fixed tissue. Nucleic Acids Res 1999;27:e12.

69. Jackson DP, Lewis FA, Taylor GR, Boylston AW, Quirke P. Tissue extraction of DNA and RNA and analysis by the polymerase chain reaction. J Clin Pathol 1990;43:499-504.

70. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol 2002;161:1961-71.

71. Ben-Ezra J, Johnson DA, Rossi J, Cook N, Wu A. Effect of fixation on the amplification of nucleic acids from paraffin-embedded material by the polymerase chain reaction. J Histochem Cytochem 1991;39:351-4.

72. Chalkley R, Hunter C. Histone-histone propinquity by aldehyde fixation of chromatin. Proc Natl Acad Sci U S A 1975;72:1304-8.

73. Williams C, Ponten F, Moberg C, Soderkvist P, Uhlen M, Ponten J, et al. A high frequency of sequence alterations is due to formalin fixation of archival specimens. Am J Pathol 1999;155:1467-71.

74. Paabo S, Irwin DM, Wilson AC. DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem 1990;265:4718-21.

75. Greer CE, Lund JK, Manos MM. PCR amplification from paraffin-embedded tissues: recommendations on fixatives for long-term storage and prospective studies. PCR Methods Appl 1991;1:46-50.

76. Douglas MP, Rogers SO. DNA damage caused by common cytological fixatives. Mutat Res 1998;401:77-88.

77. Ortiz-Pallardo ME, Ko Y, Sachinidis A, Vetter H, Fischer HP, Zhou H. Detection of alpha-1-antitrypsin PiZ individuals by SSCP and DNA sequencing in formalin-fixed and paraffin-embedded tissue: a comparison with immunohistochemical analysis. J Hepatol 2000;32:406-11.

78. Lips EH, Dierssen JW, van Eijk R, Oosting J, Eilers PH, Tollenaar RA, et al. Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed,

paraffin-embedded tumors using single nucleotide polymorphism arrays. Cancer Res 2005;65:10188-91.

79. Bernstein JL, Thompson WD, Casey G, DiCioccio RA, Whittemore AS, Diep AT, et al. Comparison of techniques for the successful detection of BRCA1 mutations in fixed paraffin-embedded tissue. Cancer Epidemiol Biomarkers Prev 2002;11:809-14.

80. Chan PK, Chan DP, To KF, Yu MY, Cheung JL, Cheng AF. Evaluation of extraction methods from paraffin wax embedded tissues for PCR amplification of human and viral DNA. J Clin Pathol 2001;54:401-3.

81. Maniatis T, Fritsch, E.F., Sambrook J. Isolation of high-molecular weight, eucaryotic DNA from cells grown i tissue culture. Molecular Cloning, A laboratory Manual, Vol.: Cold Spring Harbor Laboratory, 1982:280-1.

82. Cao W, Hashibe M, Rao JY, Morgenstern H, Zhang ZF. Comparison of methods for DNA extraction from paraffin-embedded tissues and buccal cells. Cancer Detect Prev 2003;27:397-404.

83. Houfflin-Debarge V, O'Donnell H, Overton T, Bennett PR, Fisk NM. High sensitivity of fetal DNA in plasma compared to serum and nucleated cells using unnested PCR in maternal blood. Fetal Diagn Ther 2000;15:102-7.

84. Forslund O, Nordin P, Hansson BG. Mucosal human papillomavirus types in squamous cell carcinomas of the uterine cervix and subsequently on fingers. Br J Dermatol 2000;142:1148-53.

85. Mercier B, Gaucher C, Feugeas O, Mazurier C. Direct PCR from whole blood, without DNA extraction. Nucleic Acids Res 1990;18:5908.

86. Stormer M, Kleesiek K, Dreier J. High-volume extraction of nucleic acids by magnetic bead technology for ultrasensitive detection of bacteria in blood components. Clin Chem 2007;53:104-10.

87. Nagy M, Otremba P, Kruger C, Bergner-Greiner S, Anders P, Henske B, et al. Optimization and validation of a fully automated silica-coated magnetic beads purification technology in forensics. Forensic Sci Int 2005;152:13-22.

88. Akutsu J, Tojo Y, Segawa O, Obata K, Okochi M, Tajima H, Yohda M. Development of an integrated automation system with a magnetic bead-mediated nucleic acid purification device for genetic analysis and gene manipulation. Biotechnol Bioeng 2004;86:667-71.

89. Riemann K, Adamzik M, Frauenrath S, Egensperger R, Schmid KW, Brockmeyer NH, Siffert W. Comparison of manual and automated nucleic acid extraction from whole-blood samples. J Clin Lab Anal 2007;21:244-8.

90. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet 1998;19:225-32.

91. Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 2001;11:1095-9.

92. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 2002;99:5261-6.

93. Lage JM, Leamon JH, Pejovic T, Hamann S, Lacey M, Dillon D, et al. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. Genome Res 2003;13:294-307.

94. Rector A, Tachezy R, Van Ranst M. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. J Virol 2004;78:4993-8.

95. Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen AC. Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. Nucleic Acids Res 2003;31:e129.

96. Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, et al. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. Nucleic Acids Res 2004;32:e71.

97. Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, et al. Unbiased whole-genome amplification directly from clinical samples. Genome Res 2003;13:954-64.

98. Tranah GJ, Lescault PJ, Hunter DJ, De Vivo I. Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies. Biotechnol Lett 2003;25:1031-6.

99. Murthy KK, Mahboubi VS, Santiago A, Barragan MT, Knoll R, Schultheiss HP, et al. Assessment of multiple displacement amplification for polymorphism discovery and haplotype determination at a highly polymorphic locus, MC1R. Hum Mutat 2005;26:145-52.

100. Lasken RS, Stockwell TB. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnol 2007;7:19.

101. Rook MS, Delach SM, Deyneko G, Worlock A, Wolfe JL. Whole genome amplification of DNA from laser capture-microdissected tissue for high-throughput single nucleotide polymorphism and short tandem repeat genotyping. Am J Pathol 2004;164:23-33.

102. Murray PR, Rosenthal KS, Kobayashi GS, Pfaller MA. Human Herpesviruses. In: Brown M, ed. Medical Microbiology, Vol. 3rd ed. St. Louis: Mosby, 1998:419-39.

103. Abdel-Haq NM, Asmar BI. Human herpesvirus 6 (HHV6) infection. Indian J Pediatr 2004;71:89-96.

104. Griffiths PD. Antivirals in the transplant setting. Antiviral Res 2006;71:192-200.

105. Landolfo S, Gariglio M, Gribaudo G, Lembo D. The human cytomegalovirus. Pharmacol Ther 2003;98:269-97.

106. Whitley RJ. Herpesviruses. In: Baron S, ed. Medical Microbiology, Vol. 4th ed. Texas: The University of Texas Medical Branch at Galveston, 1996.

107. Jenson HB. Human herpesvirus 8 infection. Curr Opin Pediatr 2003;15:85-91.

108. Schleiss MR. Vertically transmitted herpesvirus infections. Herpes 2003;10:4-11.

109. Arvin AM. Varicella-zoster virus. Clin Microbiol Rev 1996;9:361-81.

110. Lehtinen M, Koskela P, Ogmundsdottir HM, Bloigu A, Dillner J, Gudnadottir M, et al. Maternal herpesvirus infections and risk of acute lymphoblastic leukemia in the offspring. Am J Epidemiol 2003;158:207-13.

111. Weber B, Brunner M, Preiser W, Doerr HW. Evaluation of 11 enzyme immunoassays for the detection of immunoglobulin M antibodies to Epstein-Barr virus. J Virol Methods 1996;57:87-93.

112. Watzinger F, Suda M, Preuner S, Baumgartinger R, Ebner K, Baskova L, et al. Real-time quantitative PCR assays for detection and monitoring of pathogenic human viruses in immunosuppressed pediatric patients. J Clin Microbiol 2004;42:5189-98.

113. Mengoli C, Cusinato R, Biasolo MA, Cesaro S, Parolin C, Palu G. Assessment of CMV load in solid organ transplant recipients by pp65 antigenemia and real-time quantitative DNA PCR assay: correlation with pp67 RNA detection. J Med Virol 2004;74:78-84.

114. Madhavan HN, Priya K, Anand AR, Therese KL. Detection of herpes simplex virus (HSV) genome using polymerase chain reaction (PCR) in clinical samples comparison of PCR with standard laboratory methods for the detection of HSV. J Clin Virol 1999;14:145-51.

115. Lewensohn-Fuchs I, Osterwall P, Forsgren M, Malm G. Detection of herpes simplex virus DNA in dried blood spots making a retrospective diagnosis possible. J Clin Virol 2003;26:39-48.

116. Striebel HM, Birch-Hirschfeld E, Egerer R, Foldes-Papp Z, Tilz GP, Stelzner A. Enhancing sensitivity of human herpes virus diagnosis with DNA microarrays using dendrimers. Exp Mol Pathol 2004;77:89-97.

117. Jaaskelainen AJ, Piiparinen H, Lappalainen M, Koskiniemi M, Vaheri A. Multiplex-PCR and oligonucleotide microarray for detection of eight different herpesviruses from clinical specimens. J Clin Virol 2006;37:83-90.

118. Hudnall SD, Chen T, Tyring SK. Species identification of all eight human herpesviruses with a single nested PCR assay. J Virol Methods 2004;116:19-26.

119. Schwartz M, Roayaie S, Konstadoulakis M. Strategies for the management of hepatocellular carcinoma. Nat Clin Pract Oncol 2007;4:424-32.

120. Marrero CR, Marrero JA. Viral hepatitis and hepatocellular carcinoma. Arch Med Res 2007;38:612-20.

121. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 2007;132:2557-76.

122. Socialstyrelsen. Cancer i siffror 2005 - Poopulärvetenskapliga fakta om cancer - dess förekomst, bot och dödlighet. www.socialstyrelsen.se/Publicerat/2005/8759/2005-125-4.htm (Accessed February 1, 2008).

123. Farber ERJL. Pathology. 2nd ed. Philadelphia: J. B. Lippincott Company, 1994.

124. Steinman CR. Free DNA in serum and plasma from normal adults. J Clin Invest 1975;56:512-5.

125. Hong SP, Kim NK, Hwang SG, Chung HJ, Kim S, Han JH, et al. Detection of hepatitis B virus YMDD variants using mass spectrometric analysis of oligonucleotide fragments. J Hepatol 2004;40:837-44.

126. Ilina EN, Malakhova MV, Generozov EV, Nikolaev EN, Govorun VM. Matrix-assisted laser desorption ionization-time of flight (mass spectrometry) for hepatitis C virus genotyping. J Clin Microbiol 2005;43:2810-5.

127. Jurinke C, Zollner B, Feucht HH, Jacob A, Kirchhubel J, Luchow A, et al. Detection of hepatitis B virus DNA in serum samples via nested PCR and MALDI-TOF mass spectrometry. Genet Anal 1996;13:67-71.

128. Kim YJ, Kim SO, Chung HJ, Jee MS, Kim BG, Kim KM, et al. Population genotyping of hepatitis C virus by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis of short DNA fragments. Clin Chem 2005;51:1123-31.

129. Soderlund-Strand A, Dillner J, Carlson J. High-throughput genotyping of oncogenic human papilloma viruses with MALDI-TOF mass spectrometry. Clin Chem 2008;54:86-92.

130. Yang H, Yang K, Khafagi A, Tang Y, Carey TE, Opipari AW, et al. Sensitive detection of human papillomavirus in cervical, head/neck, and schistosomiasis-associated bladder malignancies. Proc Natl Acad Sci U S A 2005;102:7683-8.

131. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881-5.

132. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007;39:1315-7.

133. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 2007;39:984-8.

134. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39:865-9.

135. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087-93.

136. Larson MG, Atwood LD, Benjamin EJ, Cupples LA, D'Agostino RB, Sr., Fox CS, et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. BMC Med Genet 2007;8 Suppl 1:S5.

137. Allayee H, de Bruin TW, Michelle Dominguez K, Cheng LS, Ipp E, Cantor RM, et al. Genome scan for blood pressure in Dutch dyslipidemic families reveals linkage to a locus on chromosome 4p. Hypertension 2001;38:773-8.